Formally Exploring Time-Series Anomaly Detection Evaluation Metrics

Dennis Wagner RPTU Kaiserslautern
Arjun Nair RPTU Kaiserslautern
Billy Joe Franks RPTU Kaiserslautern
Justus Arweiler RPTU Kaiserslautern
Aparna Muraleedharan TUM Straubing
Indra Jungjohann RPTU Kaiserslautern
Fabian Hartung BASF SE
Andriy Balinskyy RPTU Kaiserslautern
Saurabh Varshneya RPTU Kaiserslautern
Mayank C. Ahuja RPTU Kaiserslautern
Nabeel Hussain Syed RPTU Kaiserslautern
Mayank Nagda RPTU Kaiserslautern
Philipp Liznerski RPTU Kaiserslautern
Steffen Reithermann RPTU Kaiserslautern
Maja Waldron UW Madison
Sebastian Vollmer DFKI
Ralf Schulz RPTU Landau
Torsten Katz BASF SE
Stephan Mandt UC Irvine
Michael Bortz Fraunhofer ITWM
Heike Leitte RPTU Kaiserslautern
Daniel Neider TU Dortmund
Jakob Burger TUM Straubing
Fabian Jirasek RPTU Kaiserslautern
Hans Hasse RPTU Kaiserslautern
Sophie Fellenz RPTU Kaiserslautern
Marius Kloft RPTU Kaiserslautern

Abstract

Detecting anomalies in time series is vital to ensure safety and reliability in many real-world applications. Despite the staggering number of anomaly detection methods, it remains unclear which methods perform best, largely due to flawed evaluation practices. Without rigorous analysis, evaluations yield unintuitive or misleading comparisons. Existing evaluation metrics often focus on specifics and, therefore, fail to capture essential aspects of the anomaly detection task. In this work, we formalize the problem by introducing verifiable properties of evaluation metrics that individually reflect important aspects of anomaly detection in time series. By formalizing requirements and analyzing them systematically, we outline a theoretical framework for evaluating time-series anomaly detection that can support principled evaluations and reliable comparisons. We analyze 37 known metrics and prove that most satisfy only few and none satisfy all properties, explaining many observed inconsistencies in evaluations. To address this gap, we introduce a new flexible evaluation metric LARM that provably satisfies all properties. We illustrate the adaptability of this approach by refining the properties to satisfy stricter requirements and adapting LARM to these advanced properties yielding ALARM.[1]

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300. Copyright 2026 by the author(s).

[1] Code available at `https://github.com/wagner-d/tsadm`

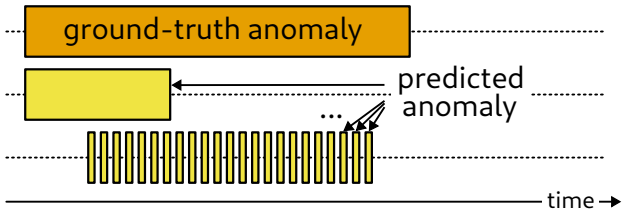


Figure 1: Illustration of a ground-truth anomaly (orange) and two prediction patterns (yellow). The top prediction represents a timely, concise alarm; the bottom a delayed yet persistent alarm pattern with excessive alarms. Although both overlap equally with the true anomaly, they differ substantially in latency and redundancy. Many evaluation metrics assign them the same score, failing to reflect their clear differences and real-world implications.

1 INTRODUCTION

Many modern applications require the processing of data while it is generated. This includes monitoring patients, servers, or industrial production processes (Darban et al., 2024). We call data collected at regular intervals from one or more sensors *time series*. A fundamental task on time-series data is the *detection of anomalies (TSAD)* – rare yet significant, temporally bounded deviations from the norm (Ruff et al., 2021). TSAD plays a critical role in preventing catastrophic failures, increasing productivity, reducing losses, or ensuring safety. Therefore, the performance of different TSAD methods needs to be thoroughly evaluated.

However, TSAD methods are often evaluated with inconsistent and incompatible metrics (Lai et al., 2024; Wang et al., 2024; Kim et al., 2024). Multiple works have shown that widely used metrics suffer from critical flaws, for example, significantly overestimating the performance of random models (Doshi et al., 2022; Gim and Min, 2023; Kim et al., 2022), and several alternatives have been introduced to address such flaws (Wagner et al., 2023; Bhattacharya et al., 2025; Si et al., 2024). Yet, most metrics are tailored to specific applications and lack a formal foundation, making it difficult to compare or generalize their behavior across different settings (Garg et al., 2021; Abdulaal et al., 2021; Gim and Min, 2023; Sørbø and Ruocco, 2024).

To enable trustworthy evaluations, metrics should behave in predictable ways that align with real-world needs. For example, a model that detects an anomaly is generally preferred over ones that fail to do so. Surprisingly, some commonly used metrics do not meet such requirements (Section 5, Table 1). While it is generally difficult to rank two methods definitively, we can often compare them when they differ in a single meaningful aspect, such as response delay or alarm redundancy. From such pairwise comparisons, we derive general properties that we can verify for each metric to characterize their capabilities. Such a characterization provides a useful tool for understanding the behavior of TSAD evaluation metrics while allowing room for task-specific restrictions and refinements.

This paper analyzes and characterizes TSAD evaluation metrics, making the following contributions:

- We introduce a formal framework for evaluating TSAD metrics, defining two sets of nine verifiable properties that capture relevant aspects of evaluation metrics such as detection accuracy, alarm redundancy, and timing.
- We conduct a comprehensive theoretical analysis of 37 existing TSAD metrics, revealing that no standard metric satisfies all properties, and explaining inconsistencies in TSAD evaluations.
- We develop two new evaluation metrics—LARM and ALARM—that provably satisfy all proposed properties, offer flexible control over application-specific trade-offs, and yield intuitive and consistent model rankings in practice.

2 WHAT MAKES GOOD METRICS?

The value of an anomaly detection method lies not only in its predictions, but in how they are judged. Metrics shape our perception of performance—and flawed metrics can distort it. This raises a fundamental question: what makes a good metric for time-series anomaly detection (TSAD)?

Choosing the right method for a TSAD application can be overwhelming. The number of available methods is large, and evaluations often vary considerably. Suppose we assemble a list of candidate methods, select a benchmark dataset, and train the methods. Each produces predictions to be compared with the ground truth. At this point, the central difficulty emerges: which metric should we use?

A popular choice in the literature is point-wise Precision (the fraction of correctly predicted anomalous time steps among all predicted anomalous time steps) and point-wise Recall (the fraction of correctly identified anomalous time steps among all time steps labeled anomalous). By their point-wise nature, these metrics are unable to distinguish any patterns in the predictions such as depicted in Fig. 1. Consider Fig. 2(a). Precision does not distinguish the two methods, while Recall favors the second. In practice, however, ground truth is not available, and every alarm raised by the second method would require investigation. As illustrated in Fig. 2(b), such alarms may stem from distinct

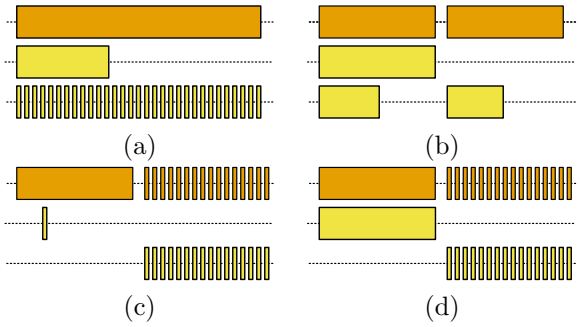


Figure 2: Challenging examples of prediction patterns (yellow) for different ground truths (orange) that common evaluation metrics struggle to order.

anomalies, while Precision and Recall fail to capture the difference.

A common adjustment treats any prediction within an anomaly window as correct if at least one time step is predicted correctly (Xu et al., 2018). Under this definition, Recall now favors the second method. Yet this creates further issues. In Fig. 2(c–d), enlarging the anomaly window makes the first method appear better according to Recall, while Precision remains unchanged. This effect may arise from a single lucky guess, which becomes increasingly likely as the size of the first anomaly window increases (Doshi et al., 2022; Kim et al., 2022; Liu and Paparrizos, 2024).

The rankings of various metrics can differ greatly. In Fig. 3 we compare the rankings of 16 predictions with 17 metrics. While the ranking on the left appropriately places perfect predictions first and effectively random predictions last, all metrics disagree considerably.

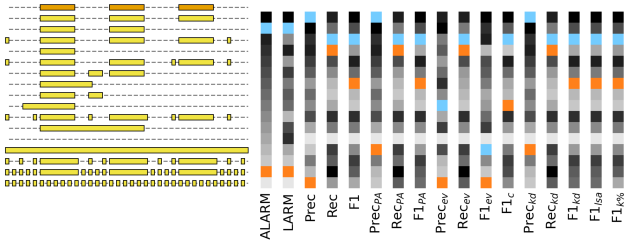


Figure 3: The rankings of 16 synthetic methods assigned an individual color each from best (top) to worst (bottom) by the ALARM score (Thm. 6.2) (left) compared to rankings of 15 other metrics (right), reveal that no two metrics agree on a ranking.

These observations lead to our main hypothesis: Many TSAD evaluation metrics have flaws that are easily overlooked and counterintuitive. This highlights the need for a formal characterization of TSAD evaluation metrics and new evaluation setups that increase the transparency of evaluations.

3 RELATED WORK

Point-adjusted metrics To remedy some of the unwanted effects of point-wise metrics (Paparrizos et al., 2022; Wagner et al., 2023; Bhattacharya et al., 2025), Xu et al. (2018) treat all missed anomalous time points as correct if at least one anomalous time point was predicted correctly in the same window. Several works have investigated these *point-adjusted* methods. Some conclude that point-adjustment introduces even worse unwanted effects (Doshi et al., 2022; Gim and Min, 2023; Kim et al., 2022) such as a preference for random predictions. Others have iterated on the idea and introduced their own variations of point-adjusted metrics (Hundman et al., 2018; Ren et al., 2019; Abdulaal et al., 2021; Kim et al., 2022; Gim and Min, 2023; Si et al., 2024; Bhattacharya et al., 2025).

Range-based metrics Point-adjusted metrics tend to share one major drawback with point-wise metrics: the inability to distinguish patterns of true positives. To address this flaw, Tatbul et al. (2018) extend the classical notion of Precision and Recall to ranges by considering continuous intervals of anomalies and predictions and Wagner et al. (2023) propose extended versions addressing some of their flaws. These metrics retain almost unparalleled degrees of freedom, making them highly flexible but difficult to apply in practice.

Fuzzy metrics Some metrics consider false positives approximately correct, as long as they are close enough to a ground-truth anomaly. Some only include predictions after ground-truth anomalies (Hwang et al., 2019; Lavin and Ahmad, 2015), while others include predictions on both sides of anomalies (Scharwächter and Müller, 2020; Huet et al., 2022). A significant drawback of these metrics is the dependence of their parameters on the dataset. For example, without the expectation of noise in the ground truth, it should be impossible to detect anomalies early. Without extensive domain knowledge, the existence and severity of such noise are impossible to discern based on the data alone.

Detection-delay metrics Most evaluation metrics struggle to consider the delay until an anomaly is detected. Xu et al. (2018) specifically account for this delay by calculating the average delay until a ground-truth anomaly is detected. Similarly, Kovács et al. (2019) consider the temporal distance to the next prediction after the start of a ground-truth anomaly. Both studies propose to use these metrics in addition to other evaluation metrics, which introduces another trade-off to consider when comparing predictions.

Score-based metrics Anomaly detection methods usually compute a scalar score before applying a thresh-

old to obtain binary predictions. Thus, each method can potentially produce a wide range of predictions depending on the threshold. When benchmarking, we only have a finite number of possible scores. Thus, we can consider all evaluations simultaneously by computing the area under some curve that is parameterized by the threshold (Xu et al., 2018; Paparrizos et al., 2022; Doshi et al., 2022; Ghorbani et al., 2024; Si et al., 2024). They are much more complex, but much less prevalent in TSAD evaluations (Xu et al., 2018). In this study, we introduce a formalism based on the comparison of binary predictions and labels. Therefore, score-based metrics are not covered by this analysis.

Characterization of evaluation metrics When choosing an evaluation metric, we implicitly specify the desired behavior of a model. Domain experts have the most insight into the desired behavior for their application. Thus, metrics or at least specifications of the desired behavior should be provided alongside datasets (Hundman et al., 2018; Si et al., 2024). Consequently, it is equally important to specify intended and actual behavior of a metric to match the expectations of a given task. Some works explicitly state that metrics should reward predictions when predicting at least one true positive for a ground-truth anomaly (Garg et al., 2021; Jacob et al., 2020; Lavin and Ahmad, 2015; Gim and Min, 2023; Si et al., 2024), predicting true positives earlier in a ground-truth anomaly window (Abdulaal et al., 2021; Doshi et al., 2022; Garg et al., 2021; Jacob et al., 2020; Lavin and Ahmad, 2015; Gim and Min, 2023; Ghorbani et al., 2024; Ren et al., 2019; Si et al., 2024), minimizing false positives (Abdulaal et al., 2021; Garg et al., 2021; Lavin and Ahmad, 2015), in particular false alarms (Gim and Min, 2023; Si et al., 2024), maximizing true positives (Abdulaal et al., 2021; Garg et al., 2021; Jacob et al., 2020; Ghorbani et al., 2024), or minimizing redundant alarms (Doshi et al., 2022; Jacob et al., 2020). While presenting useful intuitions and arguably useful goals, previous specifications lack formal rigor and are therefore not verifiable, making rigorous comparison or analysis impossible.

Evaluating evaluation metrics Sørbø and Ruocco (2024) and Liu and Paparrizos (2024) compare the rankings of two or more predictions with respect to several evaluation metrics on concrete examples that cover timing of predictions, effect of anomaly length, coverage of ground-truth anomalies, or tolerance towards label noise (Sørbø and Ruocco, 2024), or timing of predictions or ranking of random predictions (Liu and Paparrizos, 2024). While such examples can provide intuition and reveal unintuitive behavior (Liu and Paparrizos, 2024), they can only provide definitive statements for the examples evaluated, which is not sufficient to support general implications as small changes to an example can significantly affect the rankings (see Section 2). Therefore, this study introduces the first formal framework to rigorously specify and evaluate such concepts, providing an explanation for such unintuitive behavior.

4 FORMALIZING PROPERTIES

In this section, we define the problem setting, explain and motivate the core assumptions, and finally define and discuss properties of evaluation metrics.

4.1 Formal Problem Setting

An evaluation metric m is a function that compares a ground truth g with predictions p. Ground truth and predictions are binary sequences of equal length. 0 at any time step indicates normal behavior, while 1 indicates an anomaly. We can think of any sequence of length n as a function on the domain $[n] = \{1, \dots, n\}$. Lastly, we define the set of alarms raised in a binary sequence $s\colon [n] \to \{0, 1\}$. An alarm is a maximal continuous subsequence of 1's. The set of all alarms in a sequence s is given by $I_v(s) = \{[l, u] \subset [n] \mid s_{[l,u]} = v \wedge \nexists [l, u] \subsetneq [\hat{l}, \hat{u}] \subset [n] \colon s_{[\hat{l},\hat{u}]} = v\}$, where s_A is the restriction of s to the indices in $A \subset [n]$. Then the number of alarms raised in s is given by $|I_1(s)|$.

4.2 Assumptions

Labeling data requires extensive domain knowledge. Thus, any uncertainties in the labels is best resolved during the creation of a dataset. With this in mind, we assume the following.

Assumption 1 *The labels are correct.*

The assumption implies that alarms should be raised where the ground truth is labeled anomalous, and alarms should never be raised where the ground truth is labeled normal. The vast majority of metrics and evaluations are based on this assumption.

In real applications, we generally do not have access to the ground truth and thus have to rely entirely on the predictions. This means that the system must act on its predictions as they occur, without additional context or corrections. For example, once an alarm is raised, it is treated as a genuine indication of anomalous behavior, prompting intervention. This clearly differentiates Anomaly Detection from Event Detection, where only the detection of anomalous events as a whole is of interest. This leads us to the next assumption.

Assumption 2 *Predictions are taken at face value.*

This assumption is in line with the standard practice

in the literature while emphasizing the need for precise, reliable predictions to minimize costly or impractical interventions. The main implication of this assumption is that all alarms generate costs.

In general, anomalies are rare, temporally bounded deviations from normal behavior. It is impossible to definitively quantify how rare anomalies are, but to simplify the discussions in this paper, we assume that they are at least rare enough not to overlap.

ASSUMPTION 3 *Anomalies are rare and do not overlap.*

In particular, this assumption implies that multiple alarms raised in a single ground-truth anomaly window are always considered redundant.

To simplify the notation, we assume that g is a ground truth of length n, and p and q are predictions of the same length, and are thus comparable by a metric. In general, a metric produces a scalar value that can either increase or decrease with improved performance. Without loss of generality, we assume that a higher value of m corresponds to better performance.

4.3 Formalizing Properties

Based on the specified problem setting and assumptions, we define nine properties that reflect different aspects of TSAD. While each aspect is important in its own right, different settings might consider some aspects more important than others. In this section, we present one adaptation and discuss how each individual aspect and property can be relaxed or tightened depending on the specific requirements. For each property, we explain the motivation, with motivating examples shown in Fig. 4, and a formal definition.

Detecting anomalies Consider almost identical predictions that differ only in the fact that one raises at least one alarm on some anomaly window while the other does not. The method raising the alarm would definitively detect one anomaly the other does not. The following property formalizes the notion that a metric prefers the former.

PROPERTY 1 (DETECTION OF ANOMALIES) *Let $A \in I_1(g)$, $p_{[n]\setminus A} = q_{[n]\setminus A}$, and $|I_1(p_A)| > |I_1(q_A)| = 0$. Then $m(g,p) > m(g,q)$.*

Additional alarms Any excess alarms raised after an anomaly has been detected first result in additional and unnecessary costs, and, potentially, loss of the users trust in the reliability of the predictions.

PROPERTY 2 (REDUNDANT ALARMS) *Let $A \in I_1(g)$, $q(i) = \mathbb{1}_{i\in I} + p(i)\mathbb{1}_{i\notin I}$ for some $I \subset A$ with $\min I > \max\{i \in A\colon p(i) = 1\}$ and $p_I = 0$, and $|I_1(q_A)| = |I_1(p_A)| + 1$. Then $m(g,p) > m(g,q)$.*

Falsely predicting anomalies Making false positive predictions, time steps where a method incorrectly predicts an anomaly where there is none, decreases the quality of the prediction. However, false predictions can appear in different places and can even affect the number of false alarms, introducing a trade-off. Let us first consider the case where the number of alarms remains unchanged.

PROPERTY 3 (MINIMIZING FALSE POSITIVES) *Let $i^* \in N \in I_0(g)$, $p_{[n]\setminus\{i^*\}} = q_{[n]\setminus\{i^*\}}$, $q(i^*) = 1 \neq 0 = p(i^*)$, and $|I_1(p_N)| = |I_1(q_N)|$. Then $m(g,p) > m(g,q)$.*

Generally, more impactful than the total number of false positives is the number of false alarms raised as a result. No matter how many false positives are predicted, a false alarm significantly decreases the reliability of the underlying method. Too many false alarms render a method unusable for many applications.

PROPERTY 4 (MINIMIZING FALSE ALARMS) *Let $N \in I_0(g)$, $p_{[n]\setminus N} = q_{[n]\setminus N}$, and $|I_1(p_N)| < |I_1(q_N)|$. Then $m(g,p) > m(g,q)$.*

Since a false alarm does not correspond to an underlying anomaly (Asm. 1), the exact time at which it is raised is irrelevant. The following property formalizes the indifference of a metric to the timing of false alarms.

PROPERTY 5 (INVARIANCE UNDER PERMUTATION OF FALSE POSITIVES) *Let $N \in I_0(g)$, $p_{[n]\setminus N} = q_{[n]\setminus N}$, $|q^{-1}(1)| = |p^{-1}(1)|$, and $|I_1(p_N)| = |I_1(q_N)|$. Then $m(g,p) = m(g,q)$.*

Each false alarm severely reduces the user confidence in the underlying method. If the trust of the user erodes enough, they will no longer want to use the method. Thus, predictions with more false alarms are often worse, even if they are otherwise comparable.

PROPERTY 6 (KEEPING USER'S TRUST) *Let $A \in I_1(g)$, $N \in I_0(g)$, $p_{[n]\setminus(A\cup N)} = q_{[n]\setminus(A\cup N)}$, $|I_1(p_A)| = |I_1(q_A)|$, $p_N = 0$, and $q_N = \mathbb{1}_{\{i^*\}}$ for some $i^* \in N$. Then $m(g,p) > m(g,q)$.*

Preciseness of predictions Intuitively, the more anomalous time steps a method detects, the better. However, more such true positives can increase the number of alarms, resulting in worse performance according to Prop. 2.

PROPERTY 7 (MAXIMIZING TRUE POSITIVES) *Let $A \in I_1(g)$, $p_{[n]\setminus\{i^*\}} = q_{[n]\setminus\{i^*\}}$ for some $i^* \in A$, $p(i^*) = 1 \neq 0 = q(i^*)$, and $|I_1(p_A)| \leq |I_1(q_A)|$. Then $m(g,p) > m(g,q)$.*

Timing of alarms Predicting more true positives could decrease the time until an anomaly is first detected. Asm. 2 clearly implies a preference for early alarms. Intentionally avoiding the trade-off with the

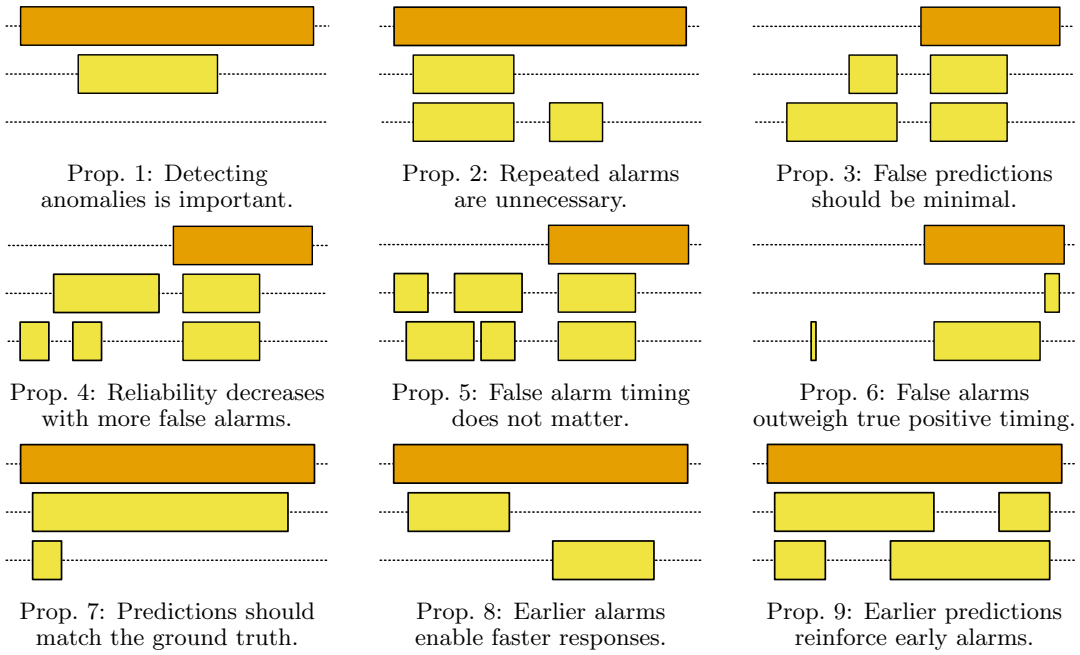


Figure 4: Visual motivation for the formalization of properties of TSAD evaluation metrics. Each panel shows a ground-truth anomaly (orange) and two prediction patterns (yellow), designed to isolate specific evaluation principles. With the exception of Prop. 5, each property formalizes the notion that a metric would prefer the top prediction, when it detects anomalies more reliably (Props. 1, 7), avoids redundancy (Prop. 2), is more precise (Prop. 3), raises fewer false alarms (Props. 4, 6), or responds earlier (Props. 8, 9).

number of alarms and number of true positives, we assume that both are constant for this comparison.

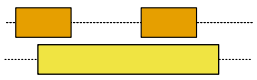
PROPERTY 8 (ALARM TIMING) *Let* $A \in I_1(g)$, $p_{[n]\setminus A} = q_{[n]\setminus A}$, $|I_1(p_A)| = |I_1(q_A)|$, $|p^{-1}(1)| = |q^{-1}(1)|$, *and* $\min\{i \in A \colon p(i) = 1\} < \min\{i \in A \colon q(i) = 1\}$. *Then* $m(g, p) > m(g, q)$.

If the true positives of a prediction are aligned early in a ground-truth anomaly, more time is spent with earlier alarms, leaving more time to address the anomaly and potentially avoid later alarms entirely.

PROPERTY 9 (EARLY BIAS) *Let* $A \in I_1(g)$, $p_{[n]\setminus\{i^*,i^{**}\}} = q_{[n]\setminus\{i^*,i^{**}\}}$ *for some* $i^* < i^{**} \in A$ *with* $p(i^*) = 1 \neq 0 = q(i^*)$ *and* $p(i^{**}) = 0 \neq 1 = q(i^{**})$ *such that* $|I_1(p_A)| \leq |I_1(q_A)|$. *Then* $m(g, p) > m(g, q)$.

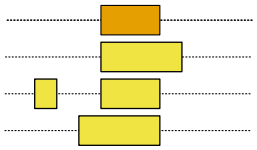
4.4 Are these properties sufficient?

Each property formalizes individual aspects of evaluations, which allows us to rigorously verify which properties each metric satisfies (Section 5) and explicitly design metrics that satisfy a given set of such properties (Section 6). However, the above properties rely on some simplifications. Firstly, we could question whether a ground-truth anomaly is actually detected, if it contains a true positive.



Consider the example where a single predicted window overlaps with two distinct ground-truth anomalies, which do not overlap with other predicted windows. Since only one alarm is raised, we could reasonably argue that only the first anomaly is detected.

Changing this definition changes the notion of false alarms. So far, all windows of false positives are equally important to false alarms. However, predictions that partially overlap with anomalies technically do not produce additional alarms. Therefore, predictions that are not overlapping with any anomaly windows—true false alarms—could be worse than early alarms or late alarms. However, considering Asm. 2 and Asm. 1, a response system or operator might mistakenly dismiss an early alarm as a false alarm, since there are no detectable effects in the data for some time after it was raised. Thus, we could end up with a differentiated ranking such as the one shown on the right. These restrictions on false alarms can have strong implications. For example, a metric that scores a method that produces any false positives lower than the method that produces no predictions could easily satisfy these properties (Thm. 6.1). Additionally, Prop. 6 imposes very strong restrictions



on false alarms which might not apply to all applications equally. However, we can control the influence of false alarms explicitly by introducing a tolerance (Thm. 6.2).

Considering these details introduces additional complexity into the definitions and proofs. We provide a set of advanced properties that incorporate all the points raised in this discussion in Appendix B. There, we illustrate how we can adapt our analysis to such changing requirements and how most results translate to this setting without much effort. For simplicity, we will work with the simple properties unless explicitly stated otherwise. In the following section, we will show that no existing TSAD metric satisfies all simple properties, in fact, most only satisfy few if any.

5 WHERE METRICS FAIL

In this section, we investigate 37 evaluation metrics with respect to the properties defined in Section 4.3. The following theorem summarizes the results of our analysis[2].

Theorem 5.1. *No existing TSAD evaluation metric satisfies all properties Prop. 1 - Prop. 9.*

Table 1 summarizes the results of our formal analysis in detail. While one could argue that individual properties, for example Prop. 6, are too specific to apply in a general setting, our analysis reveals that no metric satisfies a majority of these properties.

By definition, point-wise metrics ignore any temporal information in the data and as a result are unable to satisfy Prop. 8 and Prop. 9 which entirely depend on the temporal information in the data. With few exceptions, point-adjusted metrics generally struggle with the same properties, since most adjustment protocols intentionally remove any temporal information to emphasize detection capabilities. As a consequence, it is no surprise that they also struggle with Prop. 2, Prop. 3, and Prop. 4. On the contrary, range-based metrics are explicitly designed to incorporate temporal information through an explicit weighting scheme inside ground-truth anomaly windows. For this analysis, we focus on parametrizations which are recommended to deal with this temporal information and leave the general case to the interested reader.

Realizing that a single metric is often insufficient to completely explain the performance of different methods, most evaluations rely on a combination of Precision and Recall and their variations. Interestingly, many of these variations do not satisfy Prop. 1 which in some cases can lead to situations where perfect predictions are outperformed by objectively worse methods (for example, see Thm. D.33). Taking the prediction order properties (Prop. 8 and Prop. 9) aside, most metrics only satisfy three properties at most. It goes to show that many metrics often reward predictions counter-intuitively and thus produce unexpected results.

All metrics have their use, but their weaknesses and biases should be actively considered. Our analysis raises the question how applicable many metrics are for the general evaluation of TSAD methods. To solve this question, we introduce new metrics that satisfy all properties in the next section.

6 TOWARDS TRUSTWORTHY EVALUATIONS

The properties defined in Section 4 formalize several relevant aspects of TSAD, leading to the question of whether a single metric can simultaneously satisfy all of them. To answer this question, let us consider the effects of each property on a metric. The properties Prop. 1, Prop. 2, Prop. 7, Prop. 8, and Prop. 9 deal with true positives and are therefore not affected by changes in false positives. Similarly, we can separate terms reflecting the effects of properties that deal with false alarms (Prop. 4, Prop. 5, and Prop. 6), and properties that deal with false positives (Prop. 3). While not represented by one of the simple properties explicitly, we can add another group dealing with detected anomalies to account for the discussion in Section 4.4. We can now define a metric to satisfy all properties simultaneously by separately addressing the effects of each group as long as they do not overlap. The following theorem introduces a variation of this metric that satisfies all simple properties introduced in Section 4.

Theorem 6.1. *The a**L**ignment & **A**ccu**R**acy **M**etric (LARM) defined as*

$$\frac{1}{|I_1(g)|} \sum_{\substack{A \in I_1(g): \\ |I_1(p_A)|>0}} \frac{\alpha(p_A)+1}{2^{|I_1(p_A)|}} - 2 \sum_{A \in I_0(g)} |I_1(p_A)| - \sum_{A \in I_0(g)} \beta(|p_A^{-1}(1)|)$$

satisfies all properties Prop. 1 - Prop. 9, where α is monotonically increasing with $|p_A^{-1}(1)|$, satisfies early bias (in the sense of Prop. 9), and $codom(\alpha) \subset [0,1)$, and β is defined on $\mathbb{N}_{\geq 0}$ and monotonically increasing in $|p_A^{-1}(1)|$ with $codom(\beta) \subset [0,1]$.

For example, $\alpha(p_A) = \sum_{i \in [|A|]} p_A(i) \cdot 2^{-i}$ and $\beta(x) = 1 - x^{-1}$ with $\beta(0) := 0$ satisfy the requirements of Thm. 6.1, which we will use per default. By separating

[2] We provide detailed definitions for all metrics and proofs for all properties in Appendix D.

Table 1: None of the considered metrics fulfill all properties introduced in Section 4. Each entry shows whether a metric satisfies (✓) a property or not (✗). (✗) ((✓))indicates that, while for most parameters there are examples where the property does not hold (hold), the method in question has parameters for which the property holds (does not hold). All proofs can be found in Appendix D.

	Metric↓ \ Property→		1	2	3	4	5	6	7	8	9
point-wise	Precision	(Thm. D.1)	✗	✗	✗	✗	✓	✗	✗	✗	✗
	Recall	(Thm. D.2)	✓	✗	✗	✗	✓	✗	✓	✗	✗
	F1	(Thm. D.3)	✓	✗	✗	✗	✓	✗	✓	✗	✗
point-adjusted	Precision_{PA} (Xu et al., 2018)	(Thm. D.4)	✗	✗	✗	✗	✓	✓	✗	✗	✗
	Recall_{PA} (Xu et al., 2018)	(Thm. D.5)	✓	✗	✗	✗	✓	✗	✗	✗	✗
	F1_{PA} (Xu et al., 2018)	(Thm. D.6)	✓	✗	✗	✗	✓	✓	✗	✗	✗
	Precision_{event} (Hundman et al., 2018)	(Thm. D.7)	✗	✗	✗	✗	✗	✗	✗	✗	✗
	Recall_{event} (Hundman et al., 2018)	(Thm. D.8)	✓	✗	✗	✗	✓	✗	✗	✗	✗
	F1_{event} (Hundman et al., 2018)	(Thm. D.9)	✓	✗	✗	✗	✗	✗	✗	✗	✗
	F_{c1}	(Thm. D.10)	✓	✗	✗	✗	✓	✓	✗	✗	✗
	$\text{Precision}_{k-delay}$ (Ren et al., 2019)	(Thm. D.11)	✗	✗	✗	✗	✓	✗	✗	✗	✗
	$\text{Recall}_{k-delay}$ (Ren et al., 2019)	(Thm. D.12)	✗	✗	✗	✗	✓	✗	✗	✗	✗
	$\text{F1}_{k-delay}$ (Ren et al., 2019)	(Thm. D.13)	✗	✗	✗	✗	✓	✗	✗	✗	✗
	F1_{lsa}(Abdulaal et al., 2021)	(Thm. D.14)	✗	✗	✗	✗	✗	✗	✗	(✗)	✗
	$\text{F1}_{PA\%K}$ (Kim et al., 2022)	(Thm. D.15)	✓	✗	✗	✗	✓	(✗)	(✗)	✗	✗
	$\int_K \text{F1}_{PA\%K}$ (Kim et al., 2022)	(Thm. D.16)	✓	✗	✗	✗	✓	✗	✓	✗	✗
	F1_{PADF} (Gim and Min, 2023)	(Thm. D.17)	✓	✗	✗	✗	✓	(✗)	✗	(✓)	✗
	$\text{F1}_{reduced-length}$ (Si et al., 2024)	(Thm. D.18)	✓	✗	✗	✗	✓	✓	✗	✗	✗
	F1_{BA} (Bhattacharya et al., 2025)	(Thm. D.19)	✗	✗	✗	✗	✗	✗	✗	✗	✗
range-based	Precision_{TS} (Tatbul et al., 2018)	(Thm. D.20)	✗	✗	✗	✗	✗	✗	✗	✗	✗
	Recall_{TS} (Tatbul et al., 2018)	(Thm. D.21)	✓	✗	✗	✗	✓	✗	✓	✗	✓
	F1_{TS} (Tatbul et al., 2018)	(Thm. D.22)	✗	✗	✗	✗	✗	✗	✓	✗	✗
	TPrec (Wagner et al., 2023)	(Thm. D.23)	✗	✗	✗	✗	✗	✗	✗	✗	✗
	TRec (Wagner et al., 2023)	(Thm. D.25)	✓	✗	✗	✗	✓	✗	✓	✗	✓
	TF1 (Wagner et al., 2023)	(Thm. D.27)	✗	✗	✗	✗	✗	✗	✓	✗	✗
fuzzy labels	NAB (Lavin and Ahmad, 2015)	(Thm. D.28)	(✓)	✗	(✓)	✗	(✗)	✗	✗	✓	✗
	TaP (Hwang et al., 2019)	(Thm. D.29)	✗	✗	✗	✗	✗	✗	✗	✗	✗
	TaR (Hwang et al., 2019)	(Thm. D.30)	(✓)	✗	✗	✗	✗	✗	(✓)	✗	✗
	P_δ (Scharwächter and Müller, 2020)	(Thm. D.31)	✗	✗	✗	✗	(✗)	✗	✗	✗	✗
	R_δ (Scharwächter and Müller, 2020)	(Thm. D.32)	(✗)	✗	✗	✗	(✗)	✗	(✗)	✗	✗
	$\text{P}_{Precision}$ (Huet et al., 2022)	(Thm. D.33)	✗	✗	✗	✗	✗	✗	✗	✗	✗
	P_{Recall} (Huet et al., 2022)	(Thm. D.34)	✗	✗	✗	✗	✗	✗	✗	✗	✗
	$\text{F1}_{affiliation}$ (Huet et al., 2022)	(Thm. D.35)	✗	✗	✗	✗	✗	✗	✗	✗	✗
other	AAD (Xu et al., 2018)	(Thm. D.40)	✗	✗	✗	✗	✓	✗	✗	✓	✗
	TD (Kovács et al., 2019)	(Thm. D.38)	✓	✗	✗	✗	✗	✗	✓	✗	✗
	eTaP (Hwang et al., 2022)	(Thm. D.36)	✗	✗	✗	✗	✓	✗	✗	✗	✗
	eTaR (Hwang et al., 2022)	(Thm. D.37)	✗	✗	✗	✗	✓	✗	✗	✗	✗

the metric into individual terms that address specific subsets of properties, we only need to show how the respective terms are affected when comparing methods. The complete proofs can be found in Appendix A.

Using the same methodology, the following theorem introduces a metric that satisfies the advanced properties discussed in Section 4.4 and Appendix B.

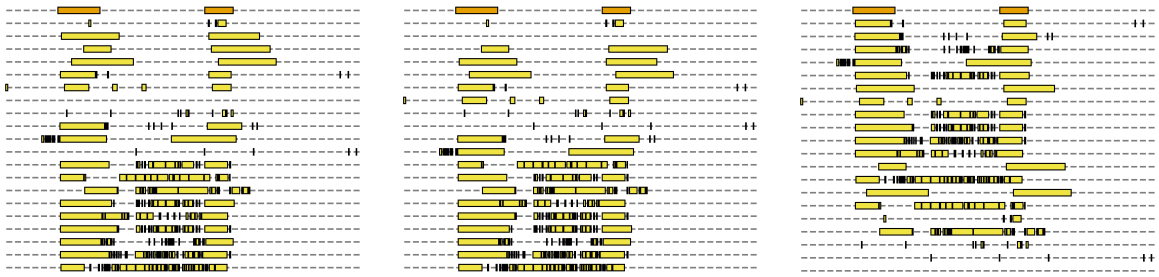
Theorem 6.2. *The **A**dvanced a**L**ignment & **A**ccu**R**acy **M**etric (ALARM)*

$$|DA(p,g)| + \frac{\sum_{A \in DA(p,g)} \frac{\alpha(p_A)+1}{2^{|I_1(p_A)|}}}{|DA(p,g)|} - \beta(|p^{-1}_{g=0}(1)|)$$
$$- t^{-1} \cdot \left(|TA(p,g)| + \frac{3}{2}|EA(p,g)| + \frac{1}{2}|LA(p,g)| \right)$$

satisfies all advanced properties from Section 4.4 and Appendix B, where α and β are defined as in Thm. 6.1, $DA(p,g)$ is the set of detected anomalies, $TA(p,g)$ the set of true false alarms, $LA(p,g)$ the set of late false alarms, $EA(p,g)$ the set of early alarms (formal definitions in Appendix B.), and $t \in \mathbb{N}_{>0}$ the alarm tolerance.

We provide all definitions, a detailed discussion, and complete proofs in Appendix B. The constant prediction $p = 0$ achieves a score of 0. Any predictions with a lower score contain at least one true false alarm and, in general, at least t times as many false alarms as detected anomalies. The alarm tolerance t is the main parameter of the ALARM score. It specifies how many false alarms each detected anomaly compensates, intuitively reflecting the system robustness with respect to false alarms. By default, we set this parameter to two.

To provide intuition for the capabilities of these metrics,



ALARM score LARM score point-wise F1 score

Figure 5: Rankings produced on a subset of SMD8 (Su et al., 2019) of 20 methods (Wagner et al., 2023) by different metrics, illustrate their different priorities. Assuming the labels are correct, most methods produce a staggering number of false alarms almost approaching random predictions, which is severely penalized by both LARM and ALARM. Notably, both group predictions with similar behavior (e.g. rows 2,3, and 4 in the left figure) with similar scores, while classic metrics often fail to do so (see e.g. right figure).

we compare the rankings produced by these metrics with the point-wise F1 score in Fig. 5. The comparison clearly shows the significant impact of false alarms in the LARM metric, which is significantly reduced by the alarm tolerance introduced in Thm. 6.2. We provide a detailed analysis of this comparison and additional evaluations on more datasets in Appendix C.

7 CONCLUSION

In this paper, we study evaluation metrics for TSAD. We illustrate that common metrics produce vastly different rankings of methods and that these can be quite counterintuitive. We formalize several aspects of TSAD through nine simple and nine advanced properties of evaluation metrics and formally prove that no existing metric satisfies all simple properties explaining this counterintuitive behavior. Finally, we introduce two new general metrics that provably satisfy all properties.

ACKNOWLEDGMENTS

The main part of this work was conducted within the DFG Research Unit FOR 5359 (ID 459419731) on Deep Learning on Sparse Chemical Process Data. SF, HH, FJ, and MK further acknowledge support by the Carl-Zeiss Foundation through the initiative Process Engineering 4.0.

References

A. Abdulaal, Z. Liu, and T. Lancewicki. Practical approach to asynchronous multivariate time series anomaly detection and localization. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, KDD '21, pages 2485–2494, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450383325. doi: 10.1145/3447548.3467174. URL `https://doi.org/10.1145/3447548.3467174`.

D. Bhattacharya, S. Mukherjee, C. Kamanchi, V. Ekambaram, A. Jati, and P. Dayama. Towards unbiased evaluation of time-series anomaly detector. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2025.

Z. Z. Darban, G. I. Webb, S. Pan, C. Aggarwal, and M. Salehi. Deep learning for time series anomaly detection: A survey. *ACM Computing Surveys*, 57 (1):1–42, 2024.

K. Doshi, S. Abudalou, and Y. Yilmaz. Tisat: Time series anomaly transformer. *arXiv preprint arXiv:2203.05167*, 2022.

A. Garg, W. Zhang, J. Samaran, R. Savitha, and C.-S. Foo. An evaluation of anomaly detection and diagnosis in multivariate time series. *IEEE Transactions on Neural Networks and Learning Systems*, 33(6): 2508–2517, 2021.

R. Ghorbani, M. J. Reinders, and D. M. Tax. Pate: Proximity-aware time series anomaly evaluation. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 872–883, 2024.

Y. Gim and K. Min. Evaluation strategy of time-series anomaly detection with decay function. *arXiv preprint arXiv:2305.09691*, 2023.

A. Huet, J. M. Navarro, and D. Rossi. Local evaluation of time series anomaly detection algorithms. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 635–645, 2022.

K. Hundman, V. Constantinou, C. Laporte, I. Colwell,

and T. Soderstrom. Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 387–395, 2018.

W.-S. Hwang, J.-H. Yun, J. Kim, and H. C. Kim. Time-series aware precision and recall for anomaly detection: considering variety of detection result and addressing ambiguous labeling. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 2241–2244, 2019.

W.-S. Hwang, J.-H. Yun, J. Kim, and B. G. Min. Do you know existing accuracy metrics overrate time-series anomaly detections? In *Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing*, pages 403–412, 2022.

V. Jacob, F. Song, A. Stiegler, B. Rad, Y. Diao, and N. Tatbul. Exathlon: A benchmark for explainable anomaly detection over time series. *arXiv preprint arXiv:2010.05073*, 2020.

D. Kim, S. Park, and J. Choo. When model meets new normals: Test-time adaptation for unsupervised time-series anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 13113–13121, 2024.

S. Kim, K. Choi, H.-S. Choi, B. Lee, and S. Yoon. Towards a rigorous evaluation of time-series anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 7194–7201, 2022.

G. Kovács, G. Sebestyen, and A. Hangan. Evaluation metrics for anomaly detection algorithms in time-series. *Acta Universitatis Sapientiae, Informatica*, 11:113–130, 12 2019. doi: 10.2478/ausi-2019-0008.

C.-Y. A. Lai, F.-K. Sun, Z. Gao, J. H. Lang, and D. Boning. Nominality score conditioned time series anomaly detection by point/sequential reconstruction. *Advances in Neural Information Processing Systems*, 36, 2024.

A. Lavin and S. Ahmad. Evaluating real-time anomaly detection algorithms–the numenta anomaly benchmark. In *2015 IEEE 14th international conference on machine learning and applications (ICMLA)*, pages 38–44. IEEE, 2015.

Q. Liu and J. Paparrizos. The elephant in the room: Towards a reliable time-series anomaly detection benchmark. *Advances in Neural Information Processing Systems*, 37:108231–108261, 2024.

J. Paparrizos, P. Boniol, T. Palpanas, R. S. Tsay, A. Elmore, and M. J. Franklin. Volume under the surface: a new accuracy evaluation measure for time-series anomaly detection. *Proceedings of the VLDB Endowment*, 15(11):2774–2787, 2022.

H. Ren, B. Xu, Y. Wang, C. Yi, C. Huang, X. Kou, T. Xing, M. Yang, J. Tong, and Q. Zhang. Time-series anomaly detection service at microsoft. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 3009–3017, 2019.

L. Ruff, J. R. Kauffmann, R. A. Vandermeulen, G. Montavon, W. Samek, M. Kloft, T. G. Dietterich, and K.-R. Müller. A unifying review of deep and shallow anomaly detection. *Proceedings of the IEEE*, 109(5): 756–795, 2021.

E. Scharwächter and E. Müller. Statistical evaluation of anomaly detectors for sequences. *arXiv preprint arXiv:2008.05788*, 2020.

H. Si, J. Li, C. Pei, H. Cui, J. Yang, Y. Sun, S. Zhang, J. Li, H. Zhang, J. Han, et al. Timeseriesbench: An industrial-grade benchmark for time series anomaly detection models. In *2024 IEEE 35th International Symposium on Software Reliability Engineering (ISSRE)*, pages 61–72. IEEE, 2024.

S. Sørbø and M. Ruocco. Navigating the metric maze: A taxonomy of evaluation metrics for anomaly detection in time series. *Data Mining and Knowledge Discovery*, 38(3):1027–1068, 2024.

Y. Su, Y. Zhao, C. Niu, R. Liu, W. Sun, and D. Pei. Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 2828–2837, 2019.

N. Tatbul, T. J. Lee, S. Zdonik, M. Alam, and J. Gottschlich. Precision and recall for time series. *Advances in neural information processing systems*, 31, 2018.

D. Wagner, T. Michels, F. C. Schulz, A. Nair, M. Rudolph, and M. Kloft. Timesead: Benchmarking deep multivariate time-series anomaly detection. *Transactions on Machine Learning Research*, 2023.

C. Wang, Z. Zhuang, Q. Qi, J. Wang, X. Wang, H. Sun, and J. Liao. Drift doesn't matter: dynamic decomposition with diffusion reconstruction for unstable multivariate time series anomaly detection. *Advances in Neural Information Processing Systems*, 36, 2024.

H. Xu, W. Chen, N. Zhao, Z. Li, J. Bu, Z. Li, Y. Liu, Y. Zhao, D. Pei, Y. Feng, et al. Unsupervised anomaly detection via variational auto-encoder for seasonal kpis in web applications. In *Proceedings of the 2018 world wide web conference*, pages 187–196, 2018.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes] Section 4.2 discusses the general assumptions and introduces the mathematical setting. Thm. 6.1 and Thm. 6.2 define each component used in the definitions. Section 6 contains details on default parameters.
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Not Applicable]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes]

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes] Section 4.2 discusses the general assumptions. Thm. 6.1 and Thm. 6.2 each specify the assumption of the components.
   (b) Complete proofs of all theoretical results. [Yes] Proof for Thm. 6.1 is provided in Appendix A, proof for Thm. 6.2 is provided in Appendix B, and proof for Thm. 5.1 (Thm. D.1 - Thm. D.37) is provided in Appendix D.
   (c) Clear explanations of any assumptions. [Yes] Section 4.2 discusses the general assumptions and introduces the mathematical setting. Individual assumptions for all properties are discussed in Section 4 and Appendix B. The assumptions for each component of Thm. 6.1 and Thm. 6.2 are provided in their statments.

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Not Applicable]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Not Applicable]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Not Applicable]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Not Applicable]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Yes] Appendix C contains references to used existing assets.
   (b) The license information of the assets, if applicable. [Not Applicable]
   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]
   (d) Information about consent from data providers/curators. [Not Applicable]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

A ALignment & AccuRacy Metric (LARM)

Theorem A.1. *There exist metrics that satisfy all properties Prop. 1 - Prop. 9. In particular, the a**L**ignment & **A**ccu**R**acy **M**etric (LARM) defined as*

$$\frac{1}{|I_1(g)|} \sum_{\substack{A \in I_1(g): \\ |I_1(p_A)|>0}} \frac{\alpha(p_A)+1}{2^{|I_1(p_A)|}} - 2 \sum_{A \in I_0(g)} |I_1(p_A)| - \sum_{A \in I_0(g)} \beta(|p_A^{-1}(1)|),$$

satisfies all properties Prop. 1 - Prop. 9, where α is monotonically increasing with $|p_A^{-1}(1)|$, satisfies early bias (in the sense of Prop. 9) and codom$(\alpha) \subset [0,1)$, and β is monotonically increasing in $|p_A^{-1}(1)|$ with codom$(\beta) \subset [0,1]$.

Proof.

Let $g\colon [n] \to \{0,1\}^n$ be the ground truth and $p, q\colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 Let $W \in I_1(g)$, $q_{[n]\setminus W} = p_{[n]\setminus W}$, $p_W = 0$, and $|q_W^{-1}(1)| > 0$. Then, $m(g,q)$ and $m(g,p)$ differ only in the term

$$\sum_{\substack{W_1 \in I_1(g): \\ |I_1(p_{W_1})|>0}} \frac{\alpha(q_{W_1})+1}{2^{|I_1(q_{W_1})|}} = \sum_{\substack{W_1 \in I_1(g): \\ |I_1(p_{W_1})|>0}} \frac{\alpha(p_{W_1})+1}{2^{|I_1(p_{W_1})|}} + \underbrace{\frac{\alpha(p_W)+1}{2^{|I_1(p_W)|}}}_{>0}.$$

Thus, it holds $m(g,p) < m(g,q)$.

Prop. 2 Let $W \in I_1(g)$, and let $p_W(i) = 0$ for all $i \in I \subset W$, $|I_1(q_W)| = |I_1(p_W)| + 1$, $\min_{i \in I} i > \max_{j \in w:\, p(j)=1} j$, and $q(i) = \mathbb{1}_I(i) + p(i)\mathbb{1}_{[n]\setminus I}(i)$. Then, $m(g,q)$ and $m(g,p)$ differ only in the term

$$\frac{\alpha(q_W)+1}{2^{|I_1(q_W)|}} < \frac{1+1}{2^{|I_1(q_W)|}} = \frac{1}{2^{|I_1(p_W)|}} \leq \frac{\alpha(p_W)+1}{2^{|I_1(p_W)|}}.$$

Thus, it holds $m(g,p) < m(g,q)$

Prop. 3 Let $W \in I_0(g)$, $q_{[n]\setminus\{i^*\}} = p_{[n]\setminus\{i^*\}}$ for some $i^* \in W$, $q(i^*) \neq p(i^*) = 0$, and $|I_1(q_W)| = |I_1(p_W)|$. Then, $m(g,q)$ and $m(g,p)$ differ only in the term

$$\left(2|I_1(q_{W_0})| + \beta(|q_{W_0}^{-1}(1)|)\right) > \left(2|I_1(p_{W_0})| + \beta(|p_{W_0}^{-1}(1)|)\right),$$

since $\beta(|q_{W_0}^{-1}(1)|) > \beta(|p_{W_0}^{-1}(1)|)$. Thus, it holds $m(g,p) > m(g,q)$.

Prop. 4 Let $W \in I_0(g)$, $q_{[n]\setminus W} = p_{[n]\setminus W}$, and $|I_1(q_W)| < |I_1(p_W)|$. Then, $m(g,q)$ and $m(g,p)$ differ only in the term

$$\left(2|I_1(q_{W_0})| + \beta(|q_{W_0}^{-1}(1)|)\right) \leq (2|I_1(q_{W_0})| + 1) < (2(|I_1(q_{W_0})| + 1))) \leq (2|I_1(p_{W_0})|) \leq \left(2|I_1(p_{W_0})| + \beta(|p_{W_0}^{-1}(1)|)\right).$$

Thus, it holds $m(g,p) < m(g,q)$.

Prop. 5 Let $W \in I_0(g)$, $q_{[n]\setminus W} = p_{[n]\setminus W}$, $|q_W^{-1}(1)| = |p_W^{-1}(1)|$, and $|I_1(q_W)| = |I_1(p_W)|$. Then, $m(g,q)$ and $m(g,p)$ can potentially differ only in the term

$$\left(2|I_1(q_W)| + \beta(|q_W^{-1}(1)|)\right)$$

and thus, since $\beta(|q_{W_0}^{-1}(1)|) = \beta(|p_{W_0}^{-1}(1)|)$, it holds $m(g,p) = m(g,q)$.

Prop. 6 Let $W \in I_1(g)$, $\hat{W} \in I_0(g)$, $|I_1(q_W)| = |I_1(p_W)|$, $q_{[n]\setminus(W\cup\hat{W})} = p_{[n]\setminus(W\cup\hat{W})}$, $p_{\hat{W}} = 0$, and $q_{\hat{W}} = \mathbb{1}_{\{i^*\}}$ for some $i^* \in \hat{W}$. It holds

$$0 < \frac{1}{|I_1(g)|} \sum_{\substack{W_1 \in I_1(g):\\ |I_1(p_{W_1})|>0}} \frac{\alpha(p_{W_1})+1}{2^{|I_1(p_{W_1})|}} < \frac{1}{|I_1(g)|} \sum_{\substack{W_1 \in I_1(g):\\ |I_1(p_{W_1})|>0}} \frac{1+1}{2^{|I_1(p_{W_1})|}} \leq \frac{1}{|I_1(g)|} \sum_{\substack{W_1 \in I_1(g):\\ |I_1(p_{W_1})|>0}} 2 \leq 2.$$

This implies

$$0 < \left|\frac{1}{|I_1(g)|} \sum_{\substack{W_1 \in I_1(g):\\ |I_1(p_{W_1})|>0}} \frac{\alpha(p_{W_1})+1}{2^{|I_1(p_{W_1})|}} - \frac{1}{|I_1(g)|} \sum_{\substack{W_1 \in I_1(g):\\ |I_1(q_{W_1})|>0}} \frac{\alpha(q_{W_1})+1}{2^{|I_1(q_{W_1})|}}\right| < 2.$$

Additionally, it holds $|I_1(p_{\hat{W}})| = 0$ and $|I_1(q_{\hat{W}})| = 1$. The second term in the definition only differs in the term

$$\left(2|I_1(q_{\hat{W}})| + \beta(1)\right) = (2+\beta(1)) = 2+\beta(1) > 2+\beta(0) = 2 + \left(2|I_1(q_{\hat{W}})| + \beta(0)\right).$$

This implies

$$-\sum_{\substack{W_0 \in I_1(|g-1|):\\ |I_1(q_{W_0})|>0}} \left(2|I_1(q_{W_0})| + \beta(|q_{W_0}^{-1}(1)|)\right) < -\sum_{\substack{W_0 \in I_1(|g-1|):\\ |I_1(p_{W_0})|>0}} \left(2|I_1(p_{W_0})| + \beta(|p_{W_0}^{-1}(1)|)\right) - 2.$$

Therefore it follows that $m(g,p) > m(g,q)$.

Prop. 7 Let $W \in I_1(g)$, $q_{[n]\setminus\{i^*\}} = p_{[n]\setminus\{i^*\}}$ for some $i^* \in W$, $q(i^*) \neq p(i^*) = 0$, and $|I_1(q_W)| \leq |I_1(p_W)|$. Then, $m(g,q)$ and $m(g,p)$ differ only in the term

$$\frac{1}{|I_1(g)|}\frac{\alpha(q_W)+1}{2^{|I_1(q_W)|}} \geq \frac{1}{|I_1(g)|}\frac{\alpha(q_W)+1}{2^{|I_1(p_W)|}} > \frac{1}{|I_1(g)|}\frac{\alpha(p_W)+1}{2^{|I_1(q_W)|}}$$

Thus, it holds $m(g,p) < m(g,q)$.

Prop. 8 Let $W \in I_1(g)$, $q_{[n]\setminus W} = p_{[n]\setminus W}$, $|I_1(q_W)| = |I_1(p_W)|$, $|q^{-1}(1)| = |p^{-1}(1)|$, and $\min_{i\in W:\, p(i)=1} i < \min_{i\in W:\, q(i)=1} i$. Then $m(g,q)$ and $m(g,p)$ differ only in the term

$$\frac{1}{|I_1(g)|}\frac{\alpha(q_W)+1}{2^{|I_1(q_W)|}} = \frac{1}{|I_1(g)|}\frac{\alpha(q_W)+1}{2^{|I_1(p_W)|}} < \frac{1}{|I_1(g)|}\frac{\alpha(p_W)+1}{2^{|I_1(q_W)|}}$$

Thus, it holds $m(g,p) > m(g,q)$.

Prop. 9 Let $W \in I_1(g)$, $q_{[n]\setminus W} = p_{[n]\setminus W}$, $|I_1(q_W)| \leq |I_1(p_W)|$, let $i^* \in W$ with $p(i^*) = 1$, let $i^{**} < i^* \in W$ with $p(i^{**}) = 0$, and $p'_W(i) = \mathbb{1}_{\{i^{**}\}}(i) + p(i)\mathbb{1}_{W\setminus\{i^*,i^{**}\}}(i)$. Then $m(g,q)$ and $m(g,p)$ differ only in the term

$$\frac{1}{|I_1(g)|}\frac{\alpha(q_W)+1}{2^{|I_1(q_W)|}} \geq \frac{1}{|I_1(g)|}\frac{\alpha(q_W)+1}{2^{|I_1(p_W)|}} > \frac{1}{|I_1(g)|}\frac{\alpha(p_W)+1}{2^{|I_1(p_W)|}}$$

Thus, it holds $m(g,p) > m(g,q)$. □

B Advanced Properties

In this section, we formalize the additional requirements discussed in Section 4.4 and prove Thm. 6.2. Note, that most proofs can easily adapted from Appendix A.

B.1 Definitions

Let $S_V = \{s\colon [n] \to V \mid n \in \mathbb{N}\}$ be the set of all sequences of finite length with values in $V \subset \mathbb{Z}$. We define the set of all maximal intervals in the domain of a sequence $s \in S$ where the sequence takes a value $v \in V$ as the collection $I_v(s)$, or formally $I_v(s) = \{[l,u] \subset s^{-1}(v) \mid s_{[l,u]} = v \wedge \nexists [l,u] \subsetneq [\hat{l},\hat{u}] \subset s^{-1}(v)\colon s_{[\hat{l},\hat{u}]} = v\}$. Furthermore, we define the set of maximal connected intervals where a sequence $s \in S$ takes values $u, v \in V$ respectively as $I_{uv}(s)$, or formally $I_{uv}(s) = \{[l,b] \sqcup [b+1,u] \subset s^{-1}(\{u,v\}) | [l,b] \in I_u(s) \wedge [b+1,u] \in I_v(s)\}$, i.e. the images of elements in $I_{uv}(s)$ of s are of the form $[uuuuuuuvvvvvvvv]$.

We consider two binary sequences of the same length $g, s \in S_{\{0,1\}}$. An element in $I_1(s)$ or $I_1(g)$ is called an **alarm** or **ground-truth alarm**. An **early alarm** is an alarm $A \in I_1(s)$ that is raised before a ground-truth alarm and persists beyond its onset. We Define the set of all early alarms as $EA(s,g) = \{A \in I_1(s_I) | I \in I_{01}(g) \wedge g_A \text{ surjective}\}$. Similarly, a **late alarm** is an alarm that is raised after the ground-truth alarm begins and remains active beyond its end. We define the set of late alarms as $LA(s,g) = \{A \in I_1(s_I) | I \in I_{10}(g) \wedge g_A \text{ surjective}\}$. A **true false alarm** does not overlap with any ground-truth alarm, and can be defined as $TA(s,g) = \{A \in I_1(s) | g_A = 0 \wedge \nexists A' \in LA(s,g) \cup EA(s,g)\colon A \subset A'\}$. Lastly, we define the set of all detected anomalies as $DA(s,g) = \{A \in I_1(g) | \exists A_s \in I_1(s)\colon A_s \cap A \neq \emptyset \wedge (\min A_s \in A \vee g_{[\min A_s, \min A)} = 0)\}$.

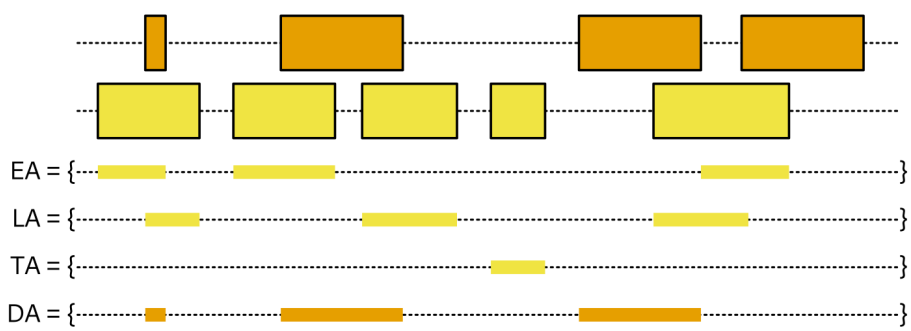


Figure 6: Example of the computation of early alarms (EA), late alarms (LA), true alarms (TA), and detected anomalies (DA). The first prediction completely includes a ground-truth window therefore contributing one interval to EA, LA, and DA. DA contains entire ground-truth anomaly windows, even if the ground-truth is not completely covered, exemplified by the second ground-truth window. The fourth predicted window is the only one not overlapping with any ground-truth anomaly making it the only true false alarm. The last predicted window overlaps with multiple ground-truth anomalies. Only the first anomaly is considered detected.

B.2 Advanced Properties

In this section we revisit the aspects discussed in Section 4.3 using this advanced framework. The differentiation of false alarms implies that changes in predictions within ground-truth anomalies can influence the number of false alarms in the predictions. As a consequence, we have to be particularly careful when specifying properties.

Detecting anomalies The general intuition is the same as in Prop. 1. There, it suffices to consider changes within one anomaly window. In this setting, however, changes in the predictions within one ground-truth anomaly window can result in new early or late false alarms. In such a case, the number of detected anomalies could stay the same, or, if both co-occur, the number of detected anomalies could decrease. Thus, we make sure to exclude these corner cases with appropriate restrictions.

Property 10 (Detection of Anomalies) *Let g be the ground truth of length n and $A \in I_1(g)$. Let p, q be predictions such that $p_{[n]\setminus A} = q_{[n]\setminus A}$, $DA(p,g) = DA(q,g) \sqcup \{A\}$, and for all $A_O \in EA(p,g) \cup LA(p,g)\colon A_O \cap A \neq \emptyset$ it holds $A_O \cap g^{-1}(0) \in TA(q,g)$. Then $m(p,g) > m(q,g)$.*

Additional Alarms In general, the intuition is the same as in Prop. 2. Excess alarms after detecting an anomaly result in additional costs and should therefore result in worse predictions. However, if an anomaly was only detected by a long early alarm, a response system could easily disregard that alarm as a false alarm. In that case, an additional alarm raised inside the anomaly window might be better. This is a delicate trade-off. Therefore, we exclude it explicitly in the following property by requiring at least one predicted window inside a ground-truth window to be present in the reference predictions.

PROPERTY 11 (REDUNDANT ALARMS) *Let g be a ground truth of length n and $A \in DA(p, g)$. Let p, q be predictions with $DA(p, g) = DA(q, g)$, $q(i) = \mathbb{1}_{i \in I} + p(i)\mathbb{1}_{i \notin I}$ for some $I \subset A$ with $\min I > \max\{i \in A \colon p(i) = 1\}$, $p_I = 0$, $|I_1(q_A)| = |I_1(p_A)| + 1$, and $\exists B \in I_1(p)$ with $B \subset A$. Then $m(p, g) > m(q, g)$.*

Falsely Predicting Anomalies Making false positive predictions, time steps where a method incorrectly predicts an anomaly where there is none, should generally decrease a method's ranking as it decreases the quality of the prediction. Similar to Prop. 3, we first consider the setting, where the number of alarms stays the same. The main difference is that we need to consider $I_1(p)$ instead of $I_1(p_N)$, to account for possible early or late alarms.

PROPERTY 12 (MINIMIZING FALSE POSITIVES) *Let g be a ground truth of length n and $N \in I_0(g)$. Let p, q be predictions with $p_{[n]\setminus\{i^*\}} = q_{[n]\setminus\{i^*\}}$, $q(i^*) = 1 \neq 0 = p(i^*)$ for some $i^* \in N$, and $|I_1(p)| \leq |I_1(q)|$. Then $m(p, g) > m(q, g)$.*

Every false alarm incurs an additional cost. Therefore, a method that makes the same number of false predictions as another method but raises fewer false alarms should be ranked higher. In contrast to Prop. 13, we need to consider changes in true false alarms, early alarms, and late alarms, since changing predictions inside a ground-truth normal window could reduce the number of one while increasing the number of another.

PROPERTY 13 (MINIMIZING FALSE ALARMS) *Let g be a ground truth of length n and $N \in I_0(g)$. Let p, q be the predictions such that $p_{[n]\setminus N} = q_{[n]\setminus N}$, $DA(p, g) = DA(q, g)$, $|p^{-1}(1)| = |q^{-1}(1)|$, $|TA(p, g)| \leq |TA(q, g)|$, $|EA(p, g)| \leq |EA(q, g)|$, $|LA(p, g)| \leq |LA(q, g)|$, and $|TA(p, g)| + |EA(p, g)| + |LA(p, g)| < |TA(q, g)| + |EA(q, g)| + |LA(q, g)|$. Then $m(p, g) > m(q, g)$.*

Following the motivation for Prop. 5, timing of true false alarms should not impact the ranking. However, the same reasoning does not apply to early and late alarms, as they are inherently connected to ground-truth anomaly windows.

PROPERTY 14 (INVARIANCE UNDER PERMUTATIONS OF TRUE FALSE ALARMS) *Let g be a ground truth of length n and p, q be the predictions, such that $p_{g=1} = q_{g=1}$, and $|p^{-1}(1)| = |q^{-1}(1)|$, $EA(p, g) = EA(q, g), LA(p, g) = LA(q, g), |TA(p, g)| = |TA(q, g)|$. Then $m(p, g) = m(q, g)$.*

Prop. 6 formalizes a trade-off that imposes strong restrictions on false alarms, which might not apply to all settings equally. Considering the differentiation of alarm types in the advanced setting, we replace this property with another that considers the relative importance of these different types. On the one hand, early alarms can have detrimental effects on the performance, since a response system could disregard a predicted anomaly as a false positive when enough time transpires until the anomaly actually manifests in the data. In this case a technically detected anomaly could be easily missed. Therefore, we argue that an early alarm is worse than a true false alarm. Note, that we consider a greatly restricted setting, where we generally assume no differences outside of the actual false positives, which additionally results in a perfectly predicted anomaly for the prediction with the true false alarm. On the other hand, late alarms are inherently connected to an anomaly that was correctly detected. Therefore, we argue that late alarms are generally preferred over true false alarms, as they raise no additional false alarm that is not related to a ground-truth anomaly. Further cementing this ranking is the fact that it is often much easier to pin-point the exact start of an anomaly window because of a thorough investigation of its root cause or because it was known in advance, but it is often difficult to identify the exact time when the system has fully recovered from all anomalous effects.

PROPERTY 15 (WEIGHING DIFFERENT TYPES OF ALARMS) *Let g be a ground truth of length n and p, q be the predictions, such that $|p^{-1}(1)| = |q^{-1}(1)|$, and $DA(p, g) = DA(q, g)$. Then $m(p, g) > m(q, g)$, if either of the following holds.*

(i) Let $A \in EA(q, g)$, $A_E = A \cap g^{-1}(0)$, $A_T \in TA(p, g)$ such that $A_E \cap A_T = \emptyset$, $p_{[n]\setminus A_E \cup A_T} = q_{[n]\setminus A_E \cup A_T}$, $p_{A_E} = 0$, and $q_{A_T} = 0$.

(ii) Let $A_T \in TA(q,g)$, $A \in LA(p,g)$, $A_L = A \cap g^{-1}(0)$ such that $A_T \cap A_L = \emptyset$, $p_{[n]\setminus A_T \cup A_L} = q_{[n]\setminus A_T \cup A_L}$, $p_{A_T} = 0$, and $q_{A_L} = 0$.

Preciseness of predictions Similar to previous properties, we have to consider the potential trade-off with early alarms when adding more correct predictions when adapting Prop. 16 to the advanced setting. Furthermore, we need to consider how the set of detected anomalies could change by introducing an additional early or late alarm. Excluding both trade-offs results in the following property.

PROPERTY 16 (MAXIMIZING TRUE POSITIVES) *Let $A \in DA(p,g) \cap DA(q,g)$, and $i^* \in A$ such that $p_{[n]\setminus\{i^*\}} = q_{[n]\setminus\{i^*\}}$, $p(i^*) = 1 \neq 0 = q(i^*)$, and $EA(p,g) = EA(p,g)$. Then $m(p,g) > m(q,g)$.*

Timing of alarms Prop. 8 and Prop. 9 formalize the preference for early predictions. In the advanced setting, earlier predictions might result in longer early alarms or even in less detected anomalies. Therefore, we additionally require that early alarms do not change when adapting Prop. 8 and Prop. 9. Another trade-off to consider is the potential loss of late alarms.

PROPERTY 17 (ALARM TIMING) *Let $A \in DA(p,g) \cap DA(q,g)$ such that $p_{[n]\setminus A} = q_{[n]\setminus A}$, $|I_1(p_A)| = |I_1(q_A)|$, $|p^{-1}(1)| = |q^{-1}(1)|$, $EA(p,g) = EA(q,g)$, and $|LA(p,g)| = |LA(q,g)|$. For some $i \in A$ it holds that $\min\{i \colon p(i) = 1\} < \min\{i \colon q(i) = 1\}$. Then $m(p,g) > m(q,g)$.*

PROPERTY 18 (EARLY BIAS) *Let g be a ground truth of length n and p, q be predictions such that $EA(p,g) = EA(q,g)$ and $|LA(p,g)| = |LA(q,g)|$. Let $A \in DA(p,g) \cap DA(q,g)$ such that $|I_1(p_A)| \leq |I_1(q_A)|$ and for some $i^* < i^{**} \in A$ it holds $q(i^{**}) = p(i^*) = 1 \neq 0 = q(i^*) = p(i^{**})$, $p_{[n]\setminus\{i^*,i^{**}\}} = q_{[n]\setminus\{i^*,i^{**}\}}$. Then $m(p,g) > m(q,g)$.*

B.3 Proof of Thm. 6.2

Lemma B.1. *Let $x \in [0,n)$, $y \in [0,1)$, and $n \in \mathbb{N}_{>0}$. Then it holds*

$$\frac{x+y}{n+1} - \frac{x}{n} > -\frac{1}{2}$$

Proof.

$$\frac{x+y}{n+1} - \frac{x}{n} \geq \frac{nx}{n(n+1)} - \frac{x(n+1)}{n(n+1)} = \frac{-x}{n(n+1)} > \frac{-n}{n(n+1)} = \frac{-1}{(n+1)} \geq -\frac{1}{2}$$

□

Theorem B.2. *The ALARM score (Thm. 6.2) satisfies the properties Prop. 10 - Prop. 18.*

Proof.

Prop. 10 By definition it holds $|DA(p,g)| = |DA(q,g)| + 1$. Since $p_{[n]\setminus A} = q_{[n]\setminus A}$, it holds that $\beta(|p^{-1}_{g=0}(1)|) = \beta(|q^{-1}_{g=0}(1)|)$. By the third assumption it follows that the number of late alarms and the number of early alarms can change at most by 1. With every increase, the number of true false alarms decreases. Thus, it follows that

$$\begin{aligned}|TA(q,g)| - 1 + \frac{3}{2}|EA(q,g)| + \frac{1}{2}(|LA(q,g)| + 1) &\leq |TA(p,g)| + \frac{3}{2}|EA(p,g)| + \frac{1}{2}|LA(p,g)| \\ &\leq |TA(q,g)| - 1 + \frac{3}{2}(|EA(q,g)| + 1) + \frac{1}{2}|LA(q,g)|\end{aligned}$$

Thus, using Thm. B.1 to bound the difference in the second term, it holds

$$m(p,g) - m(q,g) > 1 - \frac{1}{2} - \frac{1}{2t} \geq 0$$

Prop. 11 By definition, it holds $|DA(p,g)| = |DA(q,g)|$, $\beta(|p^{-1}_{g=0}(1)|) = \beta(|q^{-1}_{g=0}(1)|)$, and $EA(p,g) = EA(q,g)$. Furthermore, it holds

$$|TA(p,g)| - 1 + \frac{1}{2}(|LA(p,g)| + 1) \leq |TA(q,g)| + \frac{1}{2}|LA(q,g)| \leq |TA(p,g)| + \frac{1}{2}(|LA(p,g)| + 1)$$

and

$$\frac{1+\alpha(q_A)}{2^{|I_1(q_A)|}} < \frac{2}{2^{|I_1(q_A)|}} = \frac{1}{2^{|I_1(p_A)|}} \leq \frac{1+\alpha(p_A)}{2^{|I_1(p_A)|}}$$

and, thus, it holds

$$m(p,g) - m(q,g) > 1 - \frac{1}{2t} \geq 0$$

Prop. 12 By definition, it holds $DA(p,g) = DA(q,g)$, and $\frac{1+\alpha(q_A)}{2^{|I_1(q_A)|}} = \frac{1+\alpha(p_A)}{2^{|I_1(p_A)|}}$ for all $A \in DA(p,g)$. From the second assumption, it follows that $\beta(|q^{-1}_{g=0}(1)|) > \beta(|p^{-1}_{g=0}(1)|)$ and from the last assumption it follows that $|TA(q,g)| \geq |TA(p,g)|$. Furthermore, it holds $EA(q,g) \geq EA(p,g)$ and $LA(q,g) \geq LA(p,g)$. Therefore, it follows that $m(p,g) > m(q,g)$.

Prop. 13 By definition, the scores for p and q only differ in the term m_{FA}. There, it holds by the assumptions on the sum of alarms that this term of the score of q is strictly less than the corresponding term in the score of p due to the last assumption, resulting in the desired ordering.

Prop. 14 By definition, the first, second, and third to fourth terms are not affected by the changes between p and q. Since

$$|p^{-1}(1)| = |p^{-1}(1) \cap g^{-1}(0)| + |p^{-1}(1) \cap g^{-1}(1)| = |p^{-1}(1) \cap g^{-1}(0)| + |q^{-1}(1) \cap g^{-1}(1)| = |q^{-1}(1)|$$

it holds that $|p^{-1}(1) \cap g^{-1}(0)| = |q^{-1}(1) \cap g^{-1}(0)|$ and, therefore, $\beta(|q^{-1}_{g=0}(1)|) = \beta(|p^{-1}_{g=0}(1)|)$

Prop. 15 (i) By definition, the first two terms and the last term are unaffected by the differences between p and q. Furthermore, it holds $|EA(p,g)| + 1 = |EA(q,g)|$, $|TA(p,g)| = |TA(q,g)| + 1$, and $|LA(p,g)| \leq |LA(q,g)|$ Therefore, it holds

$$\begin{aligned} m(p,g) - m(q,g) =& \frac{1}{t}\underbrace{(|TA(q,g)| - |TA(p,g)|)}_{=-1} + \frac{3}{2t}\underbrace{(|EA(q,g)| - |EA(p,g)|)}_{=1} \\ &+ \frac{1}{2t}\underbrace{(|LA(q,g)| - |LA(p,g)|)}_{\geq 0} \geq \frac{1}{2t} > 0 \end{aligned}$$

Prop. 15 (ii) By definition, the first two terms, and the last term are unaffected by the differences between p and q. By the same argument as for (i), it holds

$$\begin{aligned} m(p,g) - m(q,g) =& \frac{1}{t}\underbrace{(|TA(q,g)| - |TA(p,g)|)}_{=1} + \frac{3}{2t}\underbrace{(|EA(q,g)| - |EA(p,g)|)}_{\geq 0} \\ &+ \frac{1}{2t}\underbrace{(|LA(q,g)| - |LA(p,g)|)}_{=-1} \geq \frac{1}{2t} > 0 \end{aligned}$$

Prop. 16 By definition, only one term in the sum of the second term is affected by the differences between p and q. Because of the monotonicity of α, it holds,

$$\frac{1+\alpha(p_A)}{2^{|I_1(p_A)|}} > \frac{1+\alpha(q_A)}{2^{|I_1(p_A)|}} \geq \frac{1+\alpha(q_A)}{2^{|I_1(q_A)|}}$$

and, therefore, it holds $m(p,g) > m(q,g)$.

Prop. 17 By definition, only one term in the sum of the second term is affected by the differences between p and q. By the early bias assumption on α, it holds

$$\frac{1+\alpha(p_A)}{2^{|I_1(p_A)|}} > \frac{1+\alpha(q_A)}{2^{|I_1(q_A)|}}$$

and, therefore, $m(p,g) > m(q,g)$.

Prop. 18 By definition, only one term in the sum of the second term is affected by the differences between p and q. By the early bias assumption on α, it holds

$$\frac{1+\alpha(p_A)}{2^{|I_1(p_A)|}} > \frac{1+\alpha(q_A)}{2^{|I_1(q_A)|}}$$

and, therefore, $m(p,g) > m(q,g)$.

□

C Evaluation

In this section we present additional evaluations of (A)LARM illustrating the differences to other metrics and providing more intuition on the requirements discussed in the main paper. We use the SMD dataset Su et al. (2019) and models trained thereon Wagner et al. (2023) as a basis for our experiments. Due to the differences in the conditions during the collection process, the SMD dataset can be considered a collection of 28 datasets Wagner et al. (2023). Since there is no quantitative measure to quantify the quality of evaluation metrics, we discuss a selection of the full evaluation here in detail to illustrate the differences between evaluation metrics.

C.1 Ranking Variance

To illustrate the high variance in rankings produced by different metrics, we take the predictions of the methods implemented and trained on SMD datasets in Wagner et al. (2023) and evaluate their performance with every metric considered in Table 1 and the (A)LARM scores proposed in Thm. 6.1 and Thm. 6.2. We assign each method a unique color and order them according to their performance for each metric. The results on more of the SMD datasets are shown in Fig. 7. This visualization illustrates the significant differences between the rankings with respect to different metrics. Notably, no existing metric provides consistent rankings with the (A)LARM scores across different datasets.

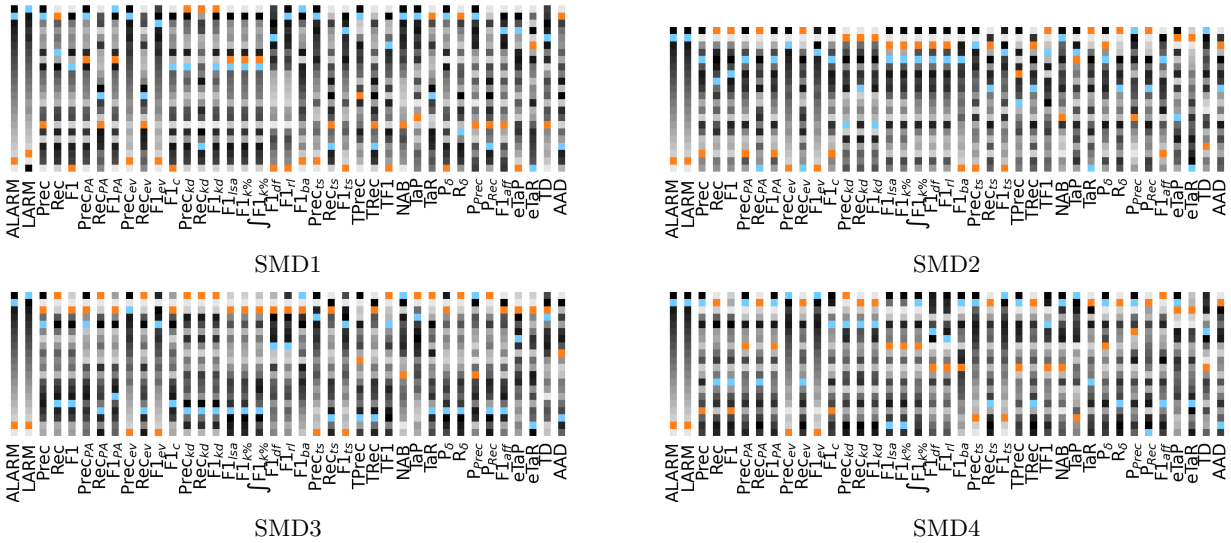


Figure 7: Comparison of the rankings of different evaluation metrics on eight SMD datasets show that (A)LARM provide significantly different rankings compared to other metrics on multiple datasets.

C.2 Qualitative Evaluation of (A)LARM

To illustrate the behavior of the metrics proposed in Thm. 6.1 and Thm. 6.2, we compare the rankings produced by (A)LARM with ranking produced by the point-wise F1 score. In Fig. 8, we can observe the significant impact of false alarms on LARM. Note that there appear to be multiple methods that make no predictions at all and that these predictions seem to appear in different positions in the rankings. Due to the size of the dataset, we have to plot at a reduced resolution where small predictions are not visible. Thus, we visualize a smaller subset of this particular dataset in Fig. 11. To illustrate the differences between rankings, we restrict the metric computation to this region only, which leaves one method with no predictions. We can see that this method is ranked second with LARM and last with the point-wise F1 score. This is a significant difference. Since false positives have an overwhelming negative impact in LARM, it naturally follows that the score decreases with the number of false positives and, in particular, with the number of false alarms. We can see that the three methods ranked just below the method with no predictions have almost the same score, each predicting both anomalies at different times and predicting anomalies for a longer time than indicated in the ground truth. Since these methods exhibit very similar behavior with slight differences in the prediction delay and the number of false positives, they are

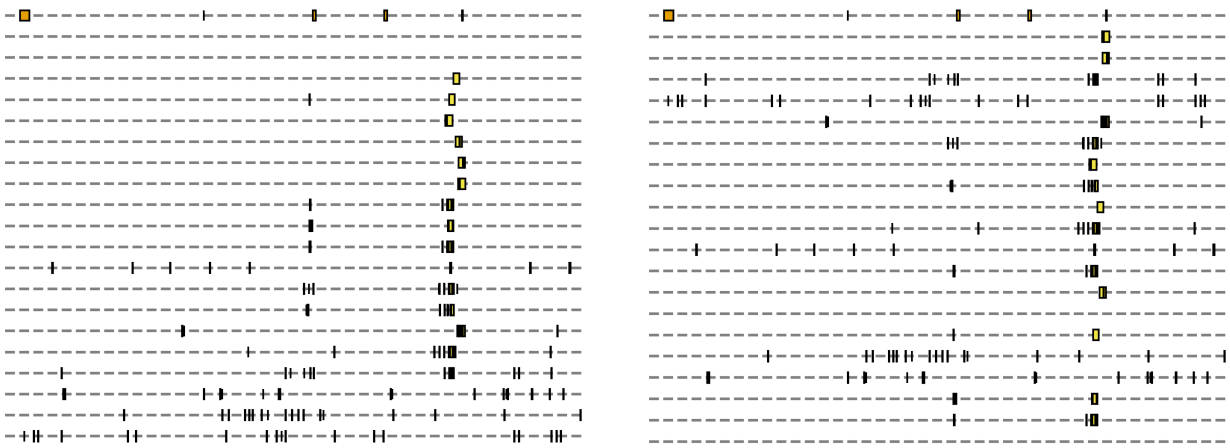


Figure 8: Comparison of LARM (left) and point-wise F1 score (right) on SMD8. The top row represents the ground truth. The methods are presented in descending order, with the top method ranked highest by each metric.

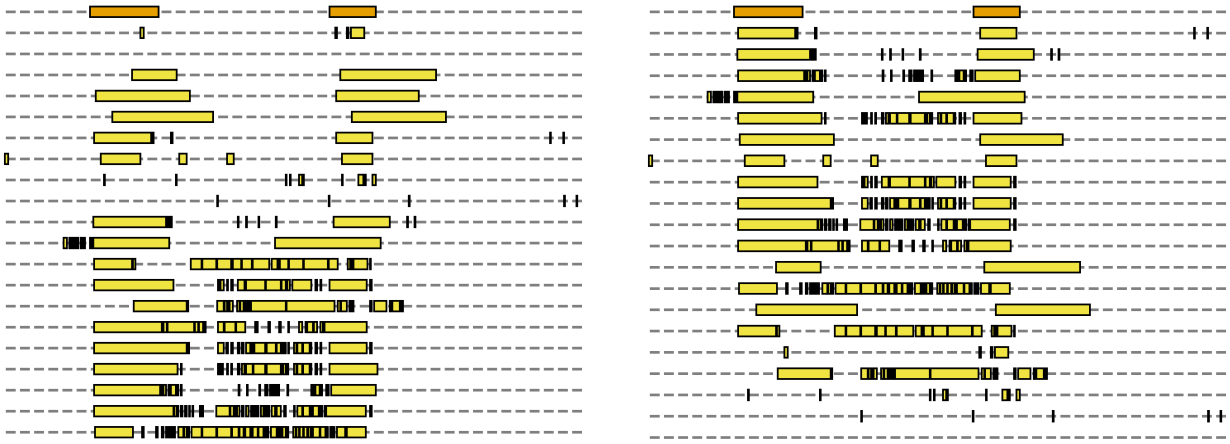


Figure 9: Comparison of LARM (left) and point-wise F1 score (right) on a subset of SMD8. The top row represents the ground truth. The methods are presented in descending order, with the top method ranked highest by each metric.

Figure 10: Three predictions of three different methods on a subset of SMD8 that all predict both anomalies at different times and all exceed the ground-truth labels. Due to these subtle differences on two different anomalies, these methods are difficult to rank.

naturally hard to distinguish. Notably, these methods appear in different positions in the ranking with the F1-score. Since we assume the labels are correct in Asm. 1 and that predictions are to be taken at face value in Asm. 2, a response system or operator might need to investigate the predictions that exceed the labels an additional time. For example, we might know based on the response used to resolve the anomaly when it should end and based on the predictions would need to initiate additional investigations. If these three methods would still represent desired behavior, it could imply that Asm. 1 is violated in this dataset. In this case, we could either introduce more lenient labels, or we could adjust the metric to account for such cases. For example, we could weaken Prop. 4 and Prop. 6 to allow for predictions that exceed the labels by modifying the sum in the

second term, as for example in ALARM (Thm. 6.2). This highlights the open design space for metrics based on (A)LARM and the need for precise specification of the desired properties. Based on our very strict requirements for reliability in LARM, the anomaly detection system might be disabled after the first false alarm, never giving it the chance to detect the second anomaly when a false positive occurs in between. Thus, the ranking produced by the metric in Thm. 6.1 aligns with our assumptions and requirements discussed in Section 4.2 and Section 4. This explains why the following method is clearly ranked lower with LARM, while it is ranked highest with the

F1-score, since the F1-score is clearly not as sensitive to false alarms and values point-wise recall much higher.

Comparing the ranking of ALARM with LARM in the same setting shows that the three methods highlighted above are still ranked closely together. However, due to the tolerance for false alarms, ALARM ranks the method with no predictions much considerably lower. Noticeably, the methods in the bottom half of the rankings of both

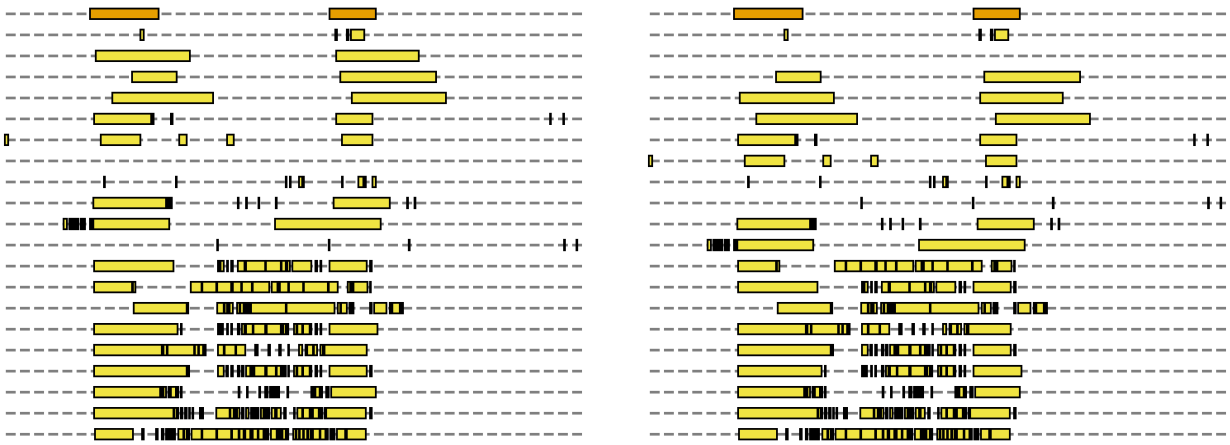


Figure 11: Comparison of ALARM (left) and LARM (right) on a subset of SMD8. The top row represents the ground truth. The methods are presented in descending order, with the top method ranked highest by each metric.

ALARM and LARM produce an increasing number of false alarms approaching almost random predictions under the assumption that the labels are correct (Asm. 1).

D Proofs

Theorem D.1. ***Point-wise Precision*** *defined as*

$$Precision(g, p) = \frac{|p^{-1}(1) \cap g^{-1}(1)|}{|p^{-1}(1)|}$$

satisfies property Prop. 5.

Proof.

Let $: [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p}\colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 Consider the following example:

$$\begin{aligned} g &= 000110011000 \\ p &= 000110000000 \\ \hat{p} &= 000110001000 \end{aligned}$$

Then $Precision(g, p) = 1 = Precision(g, \hat{p})$.

Prop. 2 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000111000000 \\ \hat{p} &= 000111011000 \end{aligned}$$

Then it holds $Precision(g, p) = 1 = Precision(g, \hat{p})$.

Prop. 3 Consider the following example:

$$\begin{aligned} g &= 000000000001 \\ p &= 000000010000 \\ \hat{p} &= 000000110000 \end{aligned}$$

Then it holds $Precision(g, p) = 0 = Precision(g, \hat{p})$.

Prop. 4 Consider the following example:

$$\begin{aligned} g &= 000000111000 \\ p &= 010010010000 \\ \hat{p} &= 011100010000 \end{aligned}$$

Then $Precision(g, p) = \frac{1}{3} > \frac{1}{4} = Precision(g, \hat{p})$.

Prop. 5 Let $W \in A(|g-1|)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $|\hat{p}_W^{-1}(1)| = |p_W^{-1}(1)|$, and $|A(\hat{p}_W)| = |A(p_W)|$. Then $|\hat{p}^{-1}(1) \cap g^{-1}(1)| = |p^{-1}(1) \cap g^{-1}(1)|$ and $|\hat{p}^{-1}(1)| = |p^{-1}(1)|$. Thus, $Precision(g, p) = Precision(g, \hat{p})$

Prop. 6 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000010000010 \\ \hat{p} &= 010111111010 \end{aligned}$$

Then $Precision(g, p) = \frac{1}{2} < \frac{3}{4} = Precision(g, \hat{p})$.

Prop. 7 Consider the following example:

$$g = 000111111000$$
$$p = 000010000000$$
$$\hat{p} = 000111111000$$

Then $Precision(g, p) = 1 = Precision(g, \hat{p})$.

Prop. 8 Consider the following example:

$$g = 000111111000$$
$$p = 000010000000$$
$$\hat{p} = 000000010000$$

Then $Precision(g, p) = 1 = Precision(g, \hat{p})$.

Prop. 9 Consider the following example:

$$g = 000111111000$$
$$p = 000010010000$$
$$\hat{p} = 000110000000$$

Then $Precision(g, p) = 1 = Precision(g, \hat{p})$.

□

Theorem D.2. ***Point-wise Recall*** *defined as*

$$Recall(g, p) = \frac{|p^{-1}(1) \cap g^{-1}(1)|}{|g^{-1}(1)|}$$

satisfies properties Prop. 1, Prop. 5, and Prop. 7.

Proof.

Let $: [n] \to \{0, 1\}^n$ be the ground truth and $p, \hat{p} \colon [n] \to \{0, 1\}^n$ predictions.

Prop. 1 Let $W \in A(g)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $p_W = 0$, and $|\hat{p}_W^{-1}(1)| > 0$. Then $|\hat{p}^{-1}(1) \cap g^{-1}(1)| > |p^{-1}(1) \cap g^{-1}(1)|$. Thus, $Recall(g, p) < Recall(g, \hat{p})$.

Prop. 2 Consider the following example:

$$g = 000111111000$$
$$p = 000111000000$$
$$\hat{p} = 000111011000$$

Then it holds $Recall(g, p) = \frac{1}{2} < \frac{5}{6} = Recall(g, \hat{p})$.

Prop. 3 Consider the following example:

$$g = 000000111000$$
$$p = 000100010000$$
$$\hat{p} = 001100010000$$

Then it holds $Recall(g, p) = \frac{1}{3} = Recall(g, \hat{p})$.

Prop. 4 Consider the following example:

$$g = 000000111000$$
$$p = 000010111000$$
$$\hat{p} = 000000111000$$

Then it holds $Recall(g, p) = 1 = Recall(g, \hat{p})$.

Prop. 5 Let $W \in A(|g-1|)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $|\hat{p}_W^{-1}(1)| = |p_W^{-1}(1)|$, and $|A(\hat{p}_W)| = |A(p_W)|$. Then $|\hat{p}^{-1}(1) \cap g^{-1}(1)| = |p^{-1}(1) \cap g^{-1}(1)|$. Thus, $Recall(g, p) = Recall(g, \hat{p})$.

Prop. 6 Consider the following example:

$$g = 000000111000$$
$$p = 000000111000$$
$$\hat{p} = 000010111000$$

Then it holds $Recall(g, p) = 1 = Recall(g, \hat{p})$.

Prop. 7 Let $W \in A(g)$, $\hat{p}_{[n]\setminus\{i^*\}} = p_{[n]\setminus\{i^*\}}$ for some $i^* \in W$, $\hat{p}(i^*) \neq p(i^*) = 0$, and $|A(\hat{p}_W)| \leq |A(p_W)|$. Then $|\hat{p}^{-1}(1) \cap g^{-1}(1)| > |p^{-1}(1) \cap g^{-1}(1)|$. Thus, $Recall(g, p) < Recall(g, \hat{p})$.

Prop. 8 Consider the following example:

$$g = 000111111000$$
$$p = 000111000000$$
$$\hat{p} = 000000111000$$

Then it holds $Recall(g, p) = \frac{1}{2} = Recall(g, \hat{p})$.

Prop. 9 Consider the following example:

$$g = 000111111000$$
$$p = 000000111000$$
$$\hat{p} = 000001110000$$

Then it holds $Recall(g, p) = \frac{1}{2} = Recall(g, \hat{p})$. □

Theorem D.3. ***Point-wise F1-score*** *defined as*

$$F1(g, p) = \frac{2|p^{-1}(1) \cap g^{-1}(1)|}{2|p^{-1}(1) \cap g^{-1}(1)| + |p^{-1}(1) \cap g^{-1}(0)| + |p^{-1}(0) \cap g^{-1}(1)|} = \frac{2|p^{-1}(1) \cap g^{-1}(1)|}{|p^{-1}(1)| + |g^{-1}(1)|}$$

satisfies properties Prop. 1, Prop. 5, and Prop. 7.

Proof.

Let $: [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p}\colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 Let $W \in A(g)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $p_W = 0$, and $|\hat{p}_W^{-1}(1)| > 0$. Then $|\hat{p}^{-1}(1) \cap g^{-1}(1)| > |p^{-1}(1) \cap g^{-1}(1)|$. Let $\epsilon = |\hat{p}^{-1}(1) \cap g^{-1}(1)| - |p^{-1}(1) \cap g^{-1}(1)|$. Then it holds

$$F1(g, \hat{p}) = \frac{2|\hat{p}^{-1}(1) \cap g^{-1}(1)|}{|\hat{p}^{-1}(1)| + |g^{-1}(1)|} = \frac{2|p^{-1}(1) \cap g^{-1}(1)| + 2\epsilon}{|p^{-1}(1)| + \epsilon + |g^{-1}(1)|} > \frac{2|p^{-1}(1) \cap g^{-1}(1)|}{|p^{-1}(1)| + |g^{-1}(1)|} = F1(g, p)$$

Thus, $F1(g, p) < F1(g, \hat{p})$.

Prop. 2 Consider the following example:

$$g = 000111111000$$
$$p = 000111000000$$
$$\hat{p} = 000111011000$$

Then it holds $F1(g,p) = \frac{2}{3} < \frac{10}{11} = F1(g,\hat{p})$.

Prop. 3 Consider the following example:

$$g = 000000000001$$
$$p = 000000010000$$
$$\hat{p} = 000010010000$$

Then it holds $F1(g,p) = 0 = F1(g,\hat{p})$.

Prop. 4 Consider the following example:

$$g = 000000111000$$
$$p = 010010010000$$
$$\hat{p} = 001100010000$$

Then it holds $F1(g,p) = \frac{1}{3} = F1(g,\hat{p})$.

Prop. 5 Let $W \in A(|g-1|)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $|\hat{p}_W^{-1}(1)| = |p_W^{-1}(1)|$, and $|A(\hat{p}_W)| = |A(p_W)|$. Then $|\hat{p}^{-1}(1) \cap g^{-1}(1)| = |p^{-1}(1) \cap g^{-1}(1)|$ and $|\hat{p}^{-1}(1)| = |p^{-1}(1)|$. Thus, $F1(g,p) = F1(g,\hat{p})$.

Prop. 6 Consider the following example:

$$g = 000111111000$$
$$p = 000000010010$$
$$\hat{p} = 010001111010$$

Then $F1(g,p) = \frac{1}{4} < \frac{2}{3} = F1(g,\hat{p})$.

Prop. 7 Let $W \in A(g)$, $\hat{p}_{[n]\setminus\{i^*\}} = p_{[n]\setminus\{i^*\}}$ for some $i^* \in W$, $\hat{p}(i^*) \neq p(i^*) = 0$, and $|A(\hat{p}_W)| \leq |A(p_W)|$. Then $|\hat{p}^{-1}(1) \cap g^{-1}(1)| = |p^{-1}(1) \cap g^{-1}(1)| + 1$ and $|\hat{p}^{-1}(1)| = |p^{-1}(1)| + 1$. Thus, $F1(g,p) < F1(g,\hat{p})$.

Prop. 8 Consider the following example:

$$g = 000111111000$$
$$p = 000110000000$$
$$\hat{p} = 000000011000$$

Then $F1(g,p) = \frac{1}{2} = F1(g,\hat{p})$.

Prop. 9 Consider the following example:

$$g = 000111111000$$
$$p = 000010010000$$
$$\hat{p} = 000110000000$$

Then $F1(g,p) = \frac{1}{2} = F1(g,\hat{p})$. $\square$

Theorem D.4. ***Point-adjusted Precision***

$$Precision_{PA}(g,p) = \frac{\sum\limits_{\substack{W \in A(g):\\ W \cap p^{-1}(1) \neq \emptyset}} |W|}{|p^{-1}(1) \cap g^{-1}(0)| + \sum\limits_{\substack{W \in A(g):\\ W \cap p^{-1}(1) \neq \emptyset}} |W|}$$

has properties Prop. 5 and Prop. 6.

Proof.

Let $: [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p}\colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 Consider the following example:

$$\begin{aligned} g &= 000110011000 \\ p &= 000110000000 \\ \hat{p} &= 000110011000 \end{aligned}$$

Then $Precision_{PA}(g,p) = 1 = Precision_{PA}(g,\hat{p})$.

Prop. 2 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000111000000 \\ \hat{p} &= 000111011000 \end{aligned}$$

Then it holds $Precision_{PA}(g,p) = 1 = Precision_{PA}(g,\hat{p})$.

Prop. 3 Consider the following example:

$$\begin{aligned} g &= 000000000001 \\ p &= 000000010000 \\ \hat{p} &= 000010010000 \end{aligned}$$

Then it holds $Precision_{PA}(g,p) = 0 = Precision_{PA}(g,\hat{p})$.

Prop. 4 Consider the following example:

$$\begin{aligned} g &= 000000111000 \\ p &= 010010010000 \\ \hat{p} &= 011100010000 \end{aligned}$$

Then $Precision_{PA}(g,p) = \frac{3}{5} > \frac{1}{2} = Precision_{PA}(g,\hat{p})$.

Prop. 5 Let $W \in A(|g-1|)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $|\hat{p}_W^{-1}(1)| = |p_W^{-1}(1)|$, and $|A(\hat{p}_W)| = |A(p_W)|$. Then $|\hat{p}^{-1}(1)| = |p^{-1}(1)|$ and $\{W \in A(g)\colon W \cap \hat{p}^{-1}(1) \neq \emptyset\} = \{W \in A(g)\colon W \cap p^{-1}(1) \neq \emptyset\}$. Thus, $Precision_{PA}(g,p) = Precision_{PA}(g,\hat{p})$.

Prop. 6 Let $W \in A(g)$, $\hat{W} \in A(|g-1|)$, $|A(\hat{p}_W)| = |A(p_W)|$, $\hat{p}_{[n]\setminus(W\cup\hat{W})} = p_{[n]\setminus(W\cup\hat{W})}$, $p_{\hat{W}} = 0$, and $\hat{p}_{\hat{W}} = \mathbb{1}_{\{i^*\}}$ for some $i^* \in \hat{W}$. Then $\{W \in A(g)\colon W \cap \hat{p}^{-1}(1) \neq \emptyset\} = \{W \in A(g)\colon W \cap p^{-1}(1) \neq \emptyset\}$ and $|\hat{p}^{-1}(1) \cap g^{-1}(0)| = |p^{-1}(1) \cap g^{-1}(0)| + 1$. Thus, $Precision_{PA}(g,p) > Precision_{PA}(g,\hat{p})$.

Prop. 7 Consider the following example:

$$g = 000111111000$$
$$p = 000010000000$$
$$\hat{p} = 000110000000$$

Then $Precision_{PA}(g, p) = 1 = Precision_{PA}(g, \hat{p})$.

Prop. 8 Consider the following example:

$$g = 000111111000$$
$$p = 000010000000$$
$$\hat{p} = 000000001000$$

Then $Precision_{PA}(g, p) = 1 = Precision_{PA}(g, \hat{p})$.

Prop. 9 Consider the following example:

$$g = 000111111000$$
$$p = 000010001000$$
$$\hat{p} = 000110000000$$

Then $Precision_{PA}(g, p) = 1 = Precision_{PA}(g, \hat{p})$. □

Theorem D.5. ***Point-adjusted Recall***

$$Recall_{PA}(g, p) = \sum_{\substack{W \in A(g): \\ W \cap p^{-1}(1) \neq \emptyset}} \frac{|W|}{|g^{-1}(1)|}$$

has Prop. 1 and Prop. 5.

Proof.

Let $: [n] \to \{0, 1\}^n$ be the ground truth and $p, \hat{p}\colon [n] \to \{0, 1\}^n$ predictions.

Prop. 1 Let $W \in A(g)$ an interval, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $p_W = 0$, and $|\hat{p}_W^{-1}(1)| > 0$. Then $Recall_{PA}(g, \hat{p}) = Recall_{PA}(g, p) + \frac{|W|}{|g^{-1}(1)|}$. Thus, $Recall_{PA}(g, p) < Recall_{PA}(g, \hat{p})$.

Prop. 2 Consider the following example:

$$g = 000111111000$$
$$p = 000111000000$$
$$\hat{p} = 000111011000$$

Then it holds $Recall_{PA}(g, p) = 1 = Recall_{PA}(g, \hat{p})$.

Prop. 3 Consider the following example:

$$g = 000000111000$$
$$p = 000100010000$$
$$\hat{p} = 011100010000$$

Then it holds $Recall_{PA}(g, p) = 1 = Recall_{PA}(g, \hat{p})$.

Prop. 4 Consider the following example:

$$g = 000000111000$$
$$p = 000010010000$$
$$\hat{p} = 000000010000$$

Then it holds $Recall_{PA}(g, p) = 1 = Recall_{PA}(g, \hat{p})$.

Prop. 5 Let $W \subset W' \in A(|g-1|)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $|\hat{p}_W^{-1}(1)| = |p_W^{-1}(1)|$, and $|A(\hat{p}_W)| = |A(p_W)|$. Then $\{W \in A(g) \colon W \cap \hat{p}^{-1}(1) \neq \emptyset\} = \{W \in A(g) \colon W \cap p^{-1}(1) \neq \emptyset\}$. Thus, $Recall_{PA}(g, p) = Recall_{PA}(g, \hat{p})$.

Prop. 6 Consider the following example:

$$g = 000000111000$$
$$p = 000000010000$$
$$\hat{p} = 000010110000$$

Then it holds $Recall_{PA}(g, p) = 1 = Recall_{PA}(g, \hat{p})$.

Prop. 7 Consider the following example:

$$g = 000111111000$$
$$p = 000010000000$$
$$\hat{p} = 000011000000$$

Then it holds $Recall_{PA}(g, p) = 1 = Recall_{PA}(g, \hat{p})$.

Prop. 8 Consider the following example:

$$g = 000111111000$$
$$p = 000111000000$$
$$\hat{p} = 000000111000$$

Then it holds $Recall_{PA}(g, p) = 1 = Recall_{PA}(g, \hat{p})$.

Prop. 9 Consider the following example:

$$g = 000111111000$$
$$p = 000000111000$$
$$\hat{p} = 000001110000$$

Then it holds $Recall_{PA}(g, p) = 1 = Recall_{PA}(g, \hat{p})$. □

Theorem D.6. ***Point-adjusted F1-score***

$$F1_{PA}(p, g) = \frac{2 \sum\limits_{\substack{W \in A(g) \colon \\ W \cap p^{-1}(1) \neq \emptyset}} |W|}{2 \sum\limits_{\substack{W \in A(g) \colon \\ W \cap p^{-1}(1) \neq \emptyset}} |W| + |p^{-1}(1) \cap g^{-1}(0)| + \sum\limits_{\substack{W \in A(g) \colon \\ W \cap p^{-1}(1) = \emptyset}} |W|}$$
$$= \frac{2 \sum\limits_{\substack{W \in A(g) \colon \\ W \cap p^{-1}(1) \neq \emptyset}} |W|}{\sum\limits_{\substack{W \in A(g) \colon \\ W \cap p^{-1}(1) \neq \emptyset}} |W| + |p^{-1}(1) \cap g^{-1}(0)| + |g^{-1}(1)|}$$

has properties Prop. 1, Prop. 5, and Prop. 6.

Proof.

Let $: [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p} \colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 Let $W \in A(g)$ an interval, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $p_W = 0$, and $|\hat{p}_W^{-1}(1)| > 0$. Then $\{W \in A(g)\colon W \cap \hat{p}^{-1}(1) \neq \emptyset\} = \{W \in A(g)\colon W \cap p^{-1}(1) \neq \emptyset\} \cup \{W\}$ with $|W| > 0$. Thus, $F1_{PA}(g,p) < F1_{PA}(g,\hat{p})$.

Prop. 2 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000111000000 \\ \hat{p} &= 000111011000 \end{aligned}$$

Then it holds $F1_{PA}(g,p) = 1 = F1_{PA}(g,\hat{p})$.

Prop. 3 Consider the following example:

$$\begin{aligned} g &= 000000000001 \\ p &= 000000010000 \\ \hat{p} &= 000010010000 \end{aligned}$$

Then it holds $F1_{PA}(g,p) = 0 = F1_{PA}(g,\hat{p})$.

Prop. 4 Consider the following example:

$$\begin{aligned} g &= 000000111000 \\ p &= 010010010000 \\ \hat{p} &= 001100010000 \end{aligned}$$

Then it holds $F1_{PA}(g,p) = \frac{3}{4} = F1_{PA}(g,\hat{p})$.

Prop. 5 Let $W \subset W' \in A(|g-1|)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $|\hat{p}_W^{-1}(1)| = |p_W^{-1}(1)|$, and $|A(\hat{p}_W)| = |A(p_W)|$. Then $\{W \in A(g)\colon W \cap \hat{p}^{-1}(1) \neq \emptyset\} = \{W \in A(g)\colon W \cap p^{-1}(1) \neq \emptyset\}$ and $|\hat{p}^{-1}(1) \cap g^{-1}(0)| = |p^{-1}(1) \cap g^{-1}(0)|$. Thus, $F1_{PA}(g,p) = F1_{PA}(g,\hat{p})$.

Prop. 6 Let $W \in A(g)$, $\hat{W} \in A(|g-1|)$, $|A(\hat{p}_W)| = |A(p_W)|$, $\hat{p}_{[n]\setminus(W\cup\hat{W})} = p_{[n]\setminus(W\cup\hat{W})}$, $p_{\hat{W}} = 0$, and $\hat{p}_{\hat{W}} = \mathbb{1}_{\{i^*\}}$ for some $i^* \in \hat{W}$. Then $\{W \in A(g)\colon W \cap \hat{p}^{-1}(1) \neq \emptyset\} = \{W \in A(g)\colon W \cap p^{-1}(1) \neq \emptyset\}$ and $|\hat{p}^{-1}(1) \cap g^{-1}(0)| = |p^{-1}(1) \cap g^{-1}(0)| + 1$. Thus, $F1_{PA}(g,p) > F1_{PA}(g,\hat{p})$.

Prop. 7 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000010000000 \\ \hat{p} &= 000011000000 \end{aligned}$$

Then it holds $F1_{PA}(g,p) = 1 = F1_{PA}(g,\hat{p})$.

Prop. 8 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000110000000 \\ \hat{p} &= 000011000000 \end{aligned}$$

Then it holds $F1_{PA}(g,p) = 1 = F1_{PA}(g,\hat{p})$.

Prop. 9 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000010010000 \\ \hat{p} &= 000110000000 \end{aligned}$$

Then it holds $F1_{PA}(g,p) = 1 = F1_{PA}(g,\hat{p})$. $\square$

Theorem D.7. ***Event-wise Precision***

$$Precision_{event}(g,p) = \frac{|\{W \in A(g) \colon W \cap p^{-1}(1) \neq \emptyset\}|}{|\{W \in A(g) \colon W \cap p^{-1}(1) \neq \emptyset\}| + |\{W \in A(p) \colon W \cap g^{-1}(1) = \emptyset\}|}$$

has none of the properties.

Proof.

Let $: [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p} \colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 Consider the following example:

$$\begin{aligned} g &= 000110011000 \\ p &= 000110000000 \\ \hat{p} &= 000110011000 \end{aligned}$$

Then $Precision_{event}(g,p) = 1 = Precision_{event}(g,\hat{p})$.

Prop. 2 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000111000000 \\ \hat{p} &= 000111011000 \end{aligned}$$

Then it holds $Precision_{event}(g,p) = 1 = Precision_{event}(g,\hat{p})$.

Prop. 3 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 001111000000 \\ \hat{p} &= 011111000000 \end{aligned}$$

Then it holds $Precision_{event}(g,p) = 1 = Precision_{event}(g,\hat{p})$.

Prop. 4 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000111000000 \\ \hat{p} &= 001111000000 \end{aligned}$$

Then it holds $Precision_{event}(g,p) = 1 = Precision_{event}(g,\hat{p})$.

Prop. 5 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 110111000000 \\ \hat{p} &= 011111000000 \end{aligned}$$

Then it holds $Precision_{event}(g,p) = \frac{1}{2} < 1 = Precision_{event}(g,\hat{p})$.

Prop. 6 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000111000000 \\ \hat{p} &= 001111000000 \end{aligned}$$

Then it holds $Precision_{event}(g,p) = 1 = Precision_{event}(g,\hat{p})$.

Prop. 7 Consider the following example:

$$g = 000111111000$$
$$p = 000101100000$$
$$\hat{p} = 000111100000$$

Then it holds $Precision_{event}(g, p) = 1 = Precision_{event}(g, \hat{p})$.

Prop. 8 Consider the following example:

$$g = 000111111000$$
$$p = 000111000000$$
$$\hat{p} = 000000111000$$

Then it holds $Precision_{event}(g, p) = 1 = Precision_{event}(g, \hat{p})$.

Prop. 9 Consider the following example:

$$g = 000111111000$$
$$p = 000001001000$$
$$\hat{p} = 000011000000$$

Then it holds $Precision_{event}(g, p) = 1 = Precision_{event}(g, \hat{p})$.

□

Theorem D.8. ***Event-wise Recall***

$$Recall_{event}(g, p) = \frac{|\{W \in A(g) \colon W \cap p^{-1}(1) \neq \emptyset\}|}{|A(g)|}$$

has properties Prop. 1 and Prop. 5.

Proof.

Let $: [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p} \colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 Let $W \in A(g)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $p_W = 0$, and $|\hat{p}_W^{-1}(1)| > 0$. Then $\{W' \in A(g) \colon W' \cap \hat{p}^{-1}(1) \neq \emptyset\} = \{W' \in A(g) \colon W' \cap p^{-1}(1) \neq \emptyset\} \cup \{W\}$. Thus $Recall_{event}(g, p) < Recall_{event}(g, \hat{p})$.

Prop. 2 Consider the following example:

$$g = 000111111000$$
$$p = 000111000000$$
$$\hat{p} = 000111011000$$

Then it holds $Recall_{event}(g, p) = 1 = Recall_{event}(g, \hat{p})$.

Prop. 3 Consider the following example:

$$g = 000111111000$$
$$p = 001111000000$$
$$\hat{p} = 011111000000$$

Then it holds $Recall_{event}(g, p) = 1 = Recall_{event}(g, \hat{p})$.

Prop. 4 Consider the following example:

$$g = 000111111000$$
$$p = 001111000000$$
$$\hat{p} = 000111000000$$

Then it holds $Recall_{event}(g, p) = 1 = Recall_{event}(g, \hat{p})$.

Prop. 5 Let $W \in A(|g-1|)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $|\hat{p}_W^{-1}(1)| = |p_W^{-1}(1)|$, and $|A(\hat{p}_W)| = |A(p_W)|$. Then $\{W' \in A(g) \colon W' \cap \hat{p}^{-1}(1) \neq \emptyset\} = \{W' \in A(g) \colon W' \cap p^{-1}(1) \neq \emptyset\}$. Thus $Recall_{event}(g, p) = Recall_{event}(g, \hat{p})$.

Prop. 6 Consider the following example:

$$g = 000111111000$$
$$p = 000111000000$$
$$\hat{p} = 001111000000$$

Then it holds $Recall_{event}(g, p) = 1 = Recall_{event}(g, \hat{p})$.

Prop. 7 Consider the following example:

$$g = 000111111000$$
$$p = 000111000000$$
$$\hat{p} = 000111100000$$

Then it holds $Recall_{event}(g, p) = 1 = Recall_{event}(g, \hat{p})$.

Prop. 8 Consider the following example:

$$g = 000111111000$$
$$p = 000111000000$$
$$\hat{p} = 000000111000$$

Then it holds $Recall_{event}(g, p) = 1 = Recall_{event}(g, \hat{p})$.

Prop. 9 Consider the following example:

$$g = 000111111000$$
$$p = 000010100000$$
$$\hat{p} = 000110000000$$

Then it holds $Recall_{event}(g, p) = 1 = Recall_{event}(g, \hat{p})$. □

Theorem D.9. ***Event-wise F1-score***

$$F1_{event}(g, p) = \frac{2|\{W \in A(g) \colon W \cap p^{-1}(1) \neq \emptyset\}|}{2|\{W \in A(g) \colon W \cap p^{-1}(1) \neq \emptyset\}| + |\{W \in A(p) \colon W \cap g^{-1}(1) = \emptyset\}| + |\{W \in A(g) \colon W \cap p^{-1}(1) = \emptyset\}|}$$

has property Prop. 1.

Proof.

Let $: [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p} \colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 Let $W \in A(g)$ an interval, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $p_W = 0$, and $|\hat{p}_W^{-1}(1)| > 0$. Then $|\{W \in A(g) \colon W \cap \hat{p}^{-1}(1) \neq \emptyset\}| = |\{W \in A(g) \colon W \cap p^{-1}(1) \neq \emptyset\}| + 1$, $|\{W \in A(\hat{p}) \colon W \cap g^{-1}(1) = \emptyset\}| \leq |\{W \in A(p) \colon W \cap g^{-1}(1) = \emptyset\}|$, and $|\{W \in A(g) \colon W \cap p^{-1}(1) = \emptyset\}| = |\{W \in A(g) \colon W \cap \hat{p}^{-1}(1) = \emptyset\}| + 1$. Thus, $F1_{event}(g, \hat{p}) > F1_{event}(g, p)$.

Prop. 2 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000111000000 \\ \hat{p} &= 000111011000 \end{aligned}$$

Then it holds $F1_{event}(g, p) = 1 = F1_{event}(g, \hat{p})$.

Prop. 3 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 001111000000 \\ \hat{p} &= 011111000000 \end{aligned}$$

Then it holds $F1_{event}(g, p) = 1 = F1_{event}(g, \hat{p})$.

Prop. 4 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 001111000000 \\ \hat{p} &= 000111000000 \end{aligned}$$

Then it holds $F1_{event}(g, p) = 1 = F1_{event}(g, \hat{p})$.

Prop. 5 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 001111000000 \\ \hat{p} &= 010111000000 \end{aligned}$$

Then it holds $F1_{event}(g, p) = 1 > \frac{2}{3} = F1_{event}(g, \hat{p})$.

Prop. 6 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000111000000 \\ \hat{p} &= 001111000000 \end{aligned}$$

Then it holds $F1_{event}(g, p) = 1 = F1_{event}(g, \hat{p})$.

Prop. 7 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000111000000 \\ \hat{p} &= 000111100000 \end{aligned}$$

Then it holds $F1_{event}(g, p) = 1 = F1_{event}(g, \hat{p})$.

Prop. 8 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000111000000 \\ \hat{p} &= 000000111000 \end{aligned}$$

Then it holds $F1_{event}(g, p) = 1 = F1_{event}(g, \hat{p})$.

Prop. 9 Consider the following example:

$$g = 000111111000$$
$$p = 000010100000$$
$$\hat{p} = 000110000000$$

Then it holds $F1_{event}(g,p) = 1 = F1_{event}(g,\hat{p})$. □

Theorem D.10. ***Composite F1-score***

$$F_{c1} = 2\frac{Precision(g,p) \cdot Recall_{event}(g,p)}{Precision(g,p) + Recall_{event}(g,p)}$$

where

$$Precision(g,p) = \frac{|p^{-1}(1) \cap g^{-1}(1)|}{|p^{-1}(1)|} \qquad \textit{and} \qquad Recall_{event}(g,p) = \frac{|\{W \in A(g) \colon W \cap p^{-1}(1) \neq \emptyset\}|}{|A(g)|}$$

has properties Prop. 1, Prop. 5, and Prop. 6.

Proof.

Let $: [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p} \colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 Let $W \in A(g)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $p_W = 0$, and $|\hat{p}_W^{-1}(1)| > 0$. Then $|\{W \in A(g) \colon W \cap \hat{p}^{-1}(1) \neq \emptyset\}| = |\{W \in A(g) \colon W \cap p^{-1}(1) \neq \emptyset\}| + 1$. Thus, $Recall_{event}(g,p) < Recall_{event}(g,\hat{p})$. We also know that $Precision(g,p) < Precision(g,\hat{p})$ (Thm. D.1). Thus, $F_{c1}(g,p) < F_{c1}(g,\hat{p})$.

Prop. 2 Consider the following example:

$$g = 000111111000$$
$$p = 000111000000$$
$$\hat{p} = 000111011000$$

Then it holds $Precision(g,p) = 1 = Precision(g,\hat{p})$, $Recall_{event}(g,p) = 1 = Recall_{event}(g,\hat{p})$, and $F_{c1}(g,p) = 1 = F_{c1}(g,\hat{p})$.

Prop. 3 Consider the following example:

$$g = 000000000001$$
$$p = 000000010000$$
$$\hat{p} = 000010010000$$

Then it holds $F_{c1}(g,p) = 0 = F_{c1}(g,\hat{p})$.

Prop. 4 Consider the following example:

$$g = 000000111000$$
$$p = 010010010000$$
$$\hat{p} = 001100010000$$

Then it holds $Precision(g,p) = 1 = Precision(g,\hat{p})$, $Recall_{event}(g,p) = 1 = Recall_{event}(g,\hat{p})$, and $F_{c1}(g,p) = 1 = F_{c1}(g,\hat{p})$.

Prop. 5 Let $W \in A(|g-1|)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $|\hat{p}_W^{-1}(1)| = |p_W^{-1}(1)|$, and $|A(\hat{p}_W)| = |A(p_W)|$. Then $Precision(g,p) = Precision(g,\hat{p})$ (Thm. D.1), $|\{W \in A(g) \colon W \cap \hat{p}^{-1}(1) \neq \emptyset\}| = |\{W \in A(g) \colon W \cap p^{-1}(1) \neq \emptyset\}|$, and $Recall_{event}(g,p) = Recall_{event}(g,\hat{p})$. Thus, $F_{c1}(g,p) = F_{c1}(g,\hat{p})$.

Prop. 6 Let $W \in A(g)$, $\hat{W} \in A(|g-1|)$, $|A(\hat{p}_W)| = |A(p_W)|$, $\hat{p}_{[n]\setminus(W\cup\hat{W})} = p_{[n]\setminus(W\cup\hat{W})}$, $p_{\hat{W}} = 0$, and $\hat{p}_{\hat{W}} = \mathbb{1}_{\{i^*\}}$ for some $i^* \in \hat{W}$. Then $|\{W \in A(g)\colon W \cap \hat{p}^{-1}(1) \neq \emptyset\}| = |\{W \in A(g)\colon W \cap p^{-1}(1) \neq \emptyset\}|$ and $Recall_{event}(g,p) = Recall_{event}(g,\hat{p})$. $|p^{-1}(1) \cap g^{-1}(1)| = |\hat{p}^{-1}(1) \cap g^{-1}(1)|$ and $|\hat{p}^{-1}(1)| = |p^{-1}(1)| + 1$ thus $Precision(g,p) > Precision(g,\hat{p})$ and finally $F_{c1}(g,p) > F_{c1}(g,\hat{p})$.

Prop. 7 Consider the following example:

$$\begin{aligned} g &= 000000111000 \\ p &= 000000011000 \\ \hat{p} &= 000000111000 \end{aligned}$$

Then it holds $Precision(g,p) = 1 = Precision(g,\hat{p})$, $Recall_{event}(g,p) = 1 = Recall_{event}(g,\hat{p})$, and $F_{c1}(g,p) = 1 = F_{c1}(g,\hat{p})$.

Prop. 8 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000111000000 \\ \hat{p} &= 000000111000 \end{aligned}$$

Then it holds $Precision(g,p) = 1 = Precision(g,\hat{p})$, $Recall_{event}(g,p) = 1 = Recall_{event}(g,\hat{p})$, and $F_{c1}(g,p) = 1 = F_{c1}(g,\hat{p})$.

Prop. 9 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000000111000 \\ \hat{p} &= 000001110000 \end{aligned}$$

Then it holds $Precision(g,p) = 1 = Precision(g,\hat{p})$, $Recall_{event}(g,p) = 1 = Recall_{event}(g,\hat{p})$, and $F_{c1}(g,p) = 1 = F_{c1}(g,\hat{p})$. □

Theorem D.11. ***K-Delay Precision***

$$Prescision_{k-delay}(g,p) = \frac{\sum\limits_{\substack{W \in A(g):\\ \min\limits_{i \in W \cap p^{-1}(1)} i \leq \min\limits_{i \in W} i+k}} |W|}{|p^{-1}(1) \cap g^{-1}(0)| + \sum\limits_{\substack{W \in A(g):\\ \min\limits_{i \in W \cap p^{-1}(1)} i \leq \min\limits_{i \in W} i+k}} |W|}$$

where $k \geq 0$ has property Prop. 5.

Proof.

Let $: [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p}\colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 Consider the following example: $g = \mathbb{1}_{[a,a+k+1]} + \mathbb{1}_{[b,c]}$ with $0 < a < a+k+1 < b < c < n$, $p = \mathbb{1}_{[b,c]}$, and $\hat{p} = \mathbb{1}_{\{a+k+1\}} + \mathbb{1}_{[b,c]}$. Then it holds $\min\limits_{i \in [a,a+k+1] \cap p^{-1}(1)} i = a+k+1 > a+k = \min\limits_{i \in [a,a+k+1]} i+k$. Thus, $Prescision_{k-delay}(g,p) = 1 = Prescision_{k-delay}(g,\hat{p})$.

For example, let $k = 1$:

$$\begin{aligned} g &= 000111011000 \\ p &= 000000011000 \\ \hat{p} &= 000001011000 \end{aligned}$$

Prop. 2 Consider the following example:

$$g = 000111111000$$
$$p = 000111000000$$
$$\hat{p} = 000111011000$$

Then it holds $Prescision_{k-delay}(g, p) = 1 = Prescision_{k-delay}(g, \hat{p})$.

Prop. 3 Consider the following example:

$$g = 000000000001$$
$$p = 000000010000$$
$$\hat{p} = 000010010000$$

Then it holds $Prescision_{k-delay}(g, p) = 0 = Prescision_{k-delay}(g, \hat{p})$.

Prop. 4 Consider the following example:

$$g = 000000111000$$
$$p = 010100111000$$
$$\hat{p} = 011000111000$$

Then it holds $Prescision_{k-delay}(g, p) = \frac{3}{5} = Prescision_{k-delay}(g, \hat{p})$.

Prop. 5 Let $W \in A(|g-1|)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $|\hat{p}_W^{-1}(1)| = |p_W^{-1}(1)|$, and $|A(\hat{p}_W)| = |A(p_W)|$. Then

$$\sum_{\substack{W \in A(g): \\ \min\limits_{i \in W \cap p^{-1}(1)} i \leq \min\limits_{i \in W} i+k}} |W| = \sum_{\substack{W \in A(g): \\ \min\limits_{i \in W \cap \hat{p}^{-1}(1)} i \leq \min\limits_{i \in W} i+k}} |W|$$

and $|\hat{p}^{-1}(1) \cap g^{-1}(0)| = |p^{-1}(1) \cap g^{-1}(0)|$. Thus $Prescision_{k-delay}(g, p) = Prescision_{k-delay}(g, \hat{p})$.

Prop. 6 Consider the following example: $g = \mathbb{1}_{[a,a+k+1]}$ with $a > 1$, $p = \mathbb{1}_{\{a+k+1\}}$, and $\hat{p} = \mathbb{1}_{\{a\}} + \mathbb{1}_{\{0\}}$. Then it holds $Prescision_{k-delay}(g, p) = 0 < \frac{k+2}{k+3} = Prescision_{k-delay}(g, \hat{p})$.
For example, let $k = 0$:

$$g = 000110000000$$
$$p = 000010000000$$
$$\hat{p} = 100100000000$$

Then it holds $Prescision_{k-delay}(g, p) = 0 < \frac{2}{3} = Prescision_{k-delay}(g, \hat{p})$.

Prop. 7 Consider the following example:

$$g = 000111111000$$
$$p = 000110000000$$
$$\hat{p} = 000111000000$$

Then it holds $Prescision_{k-delay}(g, p) = 1 = Prescision_{k-delay}(g, \hat{p})$.

Prop. 8 Consider the following example for $k > 0$:

$$g = 000111111000$$
$$p = 000111000000$$
$$\hat{p} = 000011100000$$

Then it holds $Prescision_{k-delay}(g,p) = 1 = Prescision_{k-delay}(g,\hat{p})$.
Consider the following example for $k = 0$:

$$g = 000111111000$$
$$p = 000011000000$$
$$\hat{p} = 000001100000$$

Then it holds $Prescision_{k-delay}(g,p) = 0 = Prescision_{k-delay}(g,\hat{p})$.

Prop. 9 Consider the following example:

$$g = 000111111000$$
$$p = 000100100000$$
$$\hat{p} = 000110000000$$

Then it holds $Prescision_{k-delay}(g,p) = 1 = Prescision_{k-delay}(g,\hat{p})$. □

Theorem D.12. ***K-Delay Recall***

$$Recall_{k-delay}(g,p) = \frac{\sum\limits_{\substack{W \in A(g):\\ \min\limits_{i \in W \cap p^{-1}(1)} i \leq \min\limits_{i \in W} i+k}} |W|}{\sum\limits_{W \in A(g)} |W|}$$

has property Prop. 5.

Proof.

Let $: [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p} \colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 Consider the following example: $g = \mathbb{1}_{[a,a+k+1]} + \mathbb{1}_{[b,c]}$ with $0 < a < a+k+1 < b < c < n$, $p = \mathbb{1}_{[b,c]}$, and $\hat{p} = \mathbb{1}_{\{a+k+1\}} + \mathbb{1}_{[b,c]}$. Then it holds $\min\limits_{i \in [a,a+k+1] \cap \hat{p}^{-1}(1)} i = a+k+1 > a+k = \min\limits_{i \in [a,a+k+1]} i+k$. Thus, $Recall_{k-delay}(g,p) = \frac{c-b+1}{c-b+k+3} = Recall_{k-delay}(g,\hat{p})$.
For example, let $k = 1$:

$$g = 000111011000$$
$$p = 000000011000$$
$$\hat{p} = 000001011000$$

Prop. 2 Consider the following example:

$$g = 000111111000$$
$$p = 000111000000$$
$$\hat{p} = 000111011000$$

Then it holds $Recall_{k-delay}(g,p) = 1 = Recall_{k-delay}(g,\hat{p})$.

Prop. 3 Consider the following example:

$$g = 000111111000$$
$$p = 010111000000$$
$$\hat{p} = 110111000000$$

Then it holds $Recall_{k-delay}(g,p) = 1 = Recall_{k-delay}(g,\hat{p})$.

Prop. 4 Consider the following example:

$$g = 000111111000$$
$$p = 000111000000$$
$$\hat{p} = 010111000000$$

Then it holds $Recall_{k-delay}(g, p) = 1 = Recall_{k-delay}(g, \hat{p})$.

Prop. 5 Let $W \in A(|g-1|)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $|\hat{p}_W^{-1}(1)| = |p_W^{-1}(1)|$, and $|A(\hat{p}_W)| = |A(p_W)|$. Then

$$\sum_{\substack{W \in A(g): \\ \min\limits_{i \in W \cap \hat{p}^{-1}(1)} i \leq \min\limits_{i \in W} i+k}} |W| = \sum_{\substack{W \in A(g): \\ \min\limits_{i \in W \cap p^{-1}(1)} i \leq \min\limits_{i \in W} i+k}} |W|$$

. Thus, $Recall_{k-delay}(g, p) = Recall_{k-delay}(g, \hat{p})$.

Prop. 6 Consider the following example:

$$g = 000111111000$$
$$p = 000110100000$$
$$\hat{p} = 010110010000$$

Then it holds $Recall_{k-delay}(g, p) = 1 = Recall_{k-delay}(g, \hat{p})$.

Prop. 7 Consider the following example:

$$g = 000111111000$$
$$p = 000100000000$$
$$\hat{p} = 000110000000$$

Then it holds $Recall_{k-delay}(g, p) = 1 = Recall_{k-delay}(g, \hat{p})$.

Prop. 8 Consider the following example for $k > 0$:

$$g = 000111111000$$
$$p = 000111000000$$
$$\hat{p} = 000011100000$$

Then it holds $Recall_{k-delay}(g, p) = 1 = Recall_{k-delay}(g, \hat{p})$.
Consider the following example for $k = 0$:

$$g = 000111111000$$
$$p = 000011000000$$
$$\hat{p} = 000001100000$$

Then it holds $Recall_{k-delay}(g, p) = 0 = Recall_{k-delay}(g, \hat{p})$.

Prop. 9 Consider the following example:

$$g = 000111111000$$
$$p = 000100100000$$
$$\hat{p} = 000110000000$$

Then it holds $Recall_{k-delay}(g, p) = 1 = Recall_{k-delay}(g, \hat{p})$. $\square$

Theorem D.13. ***K-Delay F1-Score***

$$F1_{k-delay}(g,p) = \frac{2 \sum_{\substack{W \in A(g): \\ \min_{i \in W \cap p^{-1}(1)} i \leq \min_{i \in W} i+k}} |W|}{|p^{-1}(1) \cap g^{-1}(0)| + 2 \sum_{\substack{W \in A(g): \\ \min_{i \in W \cap p^{-1}(1)} i \leq \min_{i \in W} i+k}} |W| + \sum_{\substack{W \in A(g): \\ \min_{i \in W \cap p^{-1}(1)} i > \min_{i \in W} i+k}} |W| + \sum_{\substack{W \in A(g): \\ W \cap p^{-1}(1) = \emptyset}} |W|}$$

has properties Prop. 5.

Proof.

Let $: [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p}\colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 Consider the following example: $g = \mathbb{1}_{[a,a+k+1]} + \mathbb{1}_{[b,c]}$ with $0 < a < a+k+1 < b < c < n$, $p = \mathbb{1}_{[b,c]}$, and $\hat{p} = \mathbb{1}_{\{a+k+1\}} + \mathbb{1}_{[b,c]}$. Then it holds $\min_{i \in [a,a+k+1] \cap p^{-1}(1)} i = a+k+1 > a+k = \min_{i \in [a,a+k+1]} i+k$. Thus, $F1_{k-delay}(g,p) = \frac{2*(c-b+1)}{2*(c-b+1)+k+2} = F1_{k-delay}(g,\hat{p})$.

For example, let $k = 1$:

$$\begin{aligned} g &= 000111011000 \\ p &= 000000011000 \\ \hat{p} &= 000001011000 \end{aligned}$$

Prop. 2 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000111000000 \\ \hat{p} &= 000111011000 \end{aligned}$$

Then it holds $F1_{k-delay}(g,p) = 1 = F1_{k-delay}(g,\hat{p})$.

Prop. 3 Consider the following example:

$$\begin{aligned} g &= 000000000001 \\ p &= 000000010000 \\ \hat{p} &= 000010010000 \end{aligned}$$

Then it holds $F1_{k-delay}(g,p) = 0 = F1_{k-delay}(g,\hat{p})$.

Prop. 4 Consider the following example:

$$\begin{aligned} g &= 000000111000 \\ p &= 010100111000 \\ \hat{p} &= 011000111000 \end{aligned}$$

Then it holds $F1_{k-delay}(g,p) = \frac{3}{4} = F1_{k-delay}(g,\hat{p})$.

Prop. 5 Let $W \in A(|g-1|)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $|\hat{p}_W^{-1}(1)| = |p_W^{-1}(1)|$, and $|A(\hat{p}_W)| = |A(p_W)|$. Then

$$\sum_{\substack{W \in A(g): \\ \min_{i \in W \cap p^{-1}(1)} i \leq \min_{i \in W} i+k}} |W| = \sum_{\substack{W \in A(g): \\ \min_{i \in W \cap \hat{p}^{-1}(1)} i \leq \min_{i \in W} i+k}} |W|$$

and $|\hat{p}^{-1}(1) \cap g^{-1}(0)| = |p^{-1}(1) \cap g^{-1}(0)|$. Thus $F1_{k-delay}(g,p) = F1_{k-delay}(g,\hat{p})$.

Prop. 6 Consider the following example: $g = \mathbb{1}_{[a,a+k+1]}$ with $a > 1$, $p = \mathbb{1}_{\{a+k+1\}}$, and $\hat{p} = \mathbb{1}_{\{a\}} + \mathbb{1}_{\{0\}}$. Then it holds $F1_{k-delay}(g,p) = 0 < \frac{2(k+2)}{2(k+2)+1} = F1_{k-delay}(g,\hat{p})$.
For example, let $k = 0$:

$$\begin{aligned} g &= 000110000000 \\ p &= 000010000000 \\ \hat{p} &= 100100000000 \end{aligned}$$

Then it holds $F1_{k-delay}(g,p) = 0 < \frac{4}{5} = F1_{k-delay}(g,\hat{p})$.

Prop. 7 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000110000000 \\ \hat{p} &= 000111000000 \end{aligned}$$

Then it holds $F1_{k-delay}(g,p) = 1 = F1_{k-delay}(g,\hat{p})$.

Prop. 8 Consider the following example for $k > 0$:

$$\begin{aligned} g &= 000111111000 \\ p &= 000111000000 \\ \hat{p} &= 000011100000 \end{aligned}$$

Then it holds $F1_{k-delay}(g,p) = 1 = F1_{k-delay}(g,\hat{p})$.
Consider the following example for $k = 0$:

$$\begin{aligned} g &= 000111111000 \\ p &= 000011000000 \\ \hat{p} &= 000001100000 \end{aligned}$$

Then it holds $F1_{k-delay}(g,p) = 0 = F1_{k-delay}(g,\hat{p})$.

Prop. 9 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000100100000 \\ \hat{p} &= 000110000000 \end{aligned}$$

Then it holds $F1_{k-delay}(g,p) = 1 = F1_{k-delay}(g,\hat{p})$. □

Theorem D.14. ***Latency- and Sparsity-Aware F1-score***

$$F1_{lsa}(g,p) = \frac{2|p_{\delta_{PA}}^{-1}(1) \cap g_{\delta}^{-1}(1)|}{2|p_{\delta_{PA}}^{-1}(1) \cap g_{\delta}^{-1}(1)| + |p_{\delta_{PA}}^{-1}(1) \cap g_{\delta}^{-1}(0)| + |p_{\delta_{PA}}^{-1}(0) \cap g_{\delta}^{-1}(1)|}$$

where:

$$\begin{aligned} b &\in \mathbb{N} \\ g_{\delta}(i) &= \max_{j \in [b \cdot i, b \cdot (i+1)-1]} g(j) \\ g' &= \sum_{i \in [\lceil \frac{n}{b} \rceil]} \mathbb{1}_{[i*b,(i+1)*b-1] \cap [n]} \mathbb{1}_{g(i*b)=1} \\ i^*(g',i) &= \begin{cases} \left\lfloor \frac{\min_{k \in W : W \in A(g'), bi \in W} k}{b} \right\rfloor & ,if\ \exists W \in A(g') : bi \in W \\ bi & ,else \end{cases} \\ p_{\delta_{PA}}(i) &= \max_{j \in [bi^*(g',i), b(i+1)-1]} p(j) \end{aligned}$$

has property Prop. 8 if $b = 1$.

Proof.

Let $: [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p}\colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 Consider the following example: $g = \mathbb{1}_{\{b-1\}}$, $p = \mathbb{1}_{\{0\}}$, and $\hat{p} = \mathbb{1}_{\{0\}} + \mathbb{1}_{\{b-1\}}$. Then it holds $F1_{lsa}(g,p) = 1 = F1_{lsa}(g,\hat{p})$.

For $b = 3$:

$$\begin{aligned} g &= 001000000000 \\ p &= 100000000000 \\ \hat{p} &= 101000000000 \end{aligned}$$

Then it holds $F1_{lsa}(g,p) = 1 = F1_{lsa}(g,\hat{p})$.

Prop. 2 Consider the following example: $g = \mathbb{1}_{[a,c]}$ where $0 < a < c-1$, $p = \mathbb{1}_{\{a\}}$, and $\hat{p} = \mathbb{1}_{\{a\}} + \mathbb{1}_{\{a+2\}}$. Then it holds $F1_{lsa}(g,p) = 1 = F1_{lsa}(g,\hat{p})$.

For $b = 3$:

$$\begin{aligned} g &= 000111111000 \\ p &= 000100000000 \\ \hat{p} &= 000101000000 \end{aligned}$$

Then it holds $F1_{lsa}(g,p) = 1 = F1_{lsa}(g,\hat{p})$.

Prop. 3 Consider the following example: $g = \mathbb{1}_{\{0\}}$, $p = \mathbb{1}_{\{b\}}$, and $\hat{p} = \mathbb{1}_{\{b\}} + \mathbb{1}_{\{2b\}}$. Then it holds $F1_{lsa}(g,p) = 0 = F1_{lsa}(g,\hat{p})$.

For $b = 3$:

$$\begin{aligned} g &= 100000000000 \\ p &= 000100000000 \\ \hat{p} &= 000100100000 \end{aligned}$$

Then it holds $F1_{lsa}(g,p) = 0 = F1_{lsa}(g,\hat{p})$.

Prop. 4 Consider the following example: $g = \mathbb{1}_{\{0\}}$, $p = \mathbb{1}_{\{0\}} + \mathbb{1}_{\{b\}}$, and $\hat{p} = \mathbb{1}_{\{0\}} + \mathbb{1}_{\{b\}} + \mathbb{1}_{\{2b-1\}}$. Then it holds $F1_{lsa}(g,p) = \frac{2}{3} = F1_{lsa}(g,\hat{p})$.

For $b = 3$:

$$\begin{aligned} g &= 100000000000 \\ p &= 100100000000 \\ \hat{p} &= 100101000000 \end{aligned}$$

Then it holds $F1_{lsa}(g,p) = \frac{2}{3} = F1_{lsa}(g,\hat{p})$.

Prop. 5 Consider the following example: $g = \mathbb{1}_{\{0\}}$, $p = \mathbb{1}_{\{0\}} + \mathbb{1}_{\{b-1\}}$, and $\hat{p} = \mathbb{1}_{\{0\}} + \mathbb{1}_{\{b\}}$. Then it holds $F1_{lsa}(g,p) = 1 > \frac{2}{3} = F1_{lsa}(g,\hat{p})$.

For $b = 3$:

$$\begin{aligned} g &= 100000000000 \\ p &= 101000000000 \\ \hat{p} &= 100100000000 \end{aligned}$$

Then it holds $F1_{lsa}(g,p) = 1 > \frac{2}{3} = F1_{lsa}(g,\hat{p})$.

Prop. 6 Consider the following example: $g = \mathbb{1}_{[0,2b-1]}$, $p = \mathbb{1}_{\{b\}}$, and $\hat{p} = \mathbb{1}_{\{0\}} + \mathbb{1}_{\{2b\}}$. Then it holds $F1_{lsa}(g,p) = \frac{2}{3} < \frac{4}{5} = F1_{lsa}(g,\hat{p})$.

For $b = 3$:

$$\begin{aligned} g &= 111111000000 \\ p &= 000100000000 \\ \hat{p} &= 100000100000 \end{aligned}$$

Then it holds $F1_{lsa}(g,p) = \frac{2}{3} < \frac{4}{5} = F1_{lsa}(g,\hat{p})$.

Prop. 7 Consider the following example:

$$\begin{aligned} g &= 110000000000 \\ p &= 100000000000 \\ \hat{p} &= 110000000000 \end{aligned}$$

Then it holds $F1_{lsa}(g,p) = 1 = F1_{lsa}(g,\hat{p})$.

Prop. 8 **Case $b = 1$:** Let $W \in A(g)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $|A(\hat{p}_W)| = |A(p_W)|$, $|\hat{p}^{-1}(1)| = |p^{-1}(1)|$, and $\min_{i \in W\,:\,p(i)=1} i < \min_{i \in W\,:\,\hat{p}(i)=1} i$. Then due to the right-ward point adjusting of $p_{\delta_{PA}}$, $|\hat{p}^{-1}_{\delta_{PA}}(1) \cap g^{-1}_{\delta}(1)| > |p^{-1}_{\delta_{PA}}(1) \cap g^{-1}_{\delta}(1)|$ while $|\hat{p}^{-1}_{\delta_{PA}}(0) \cap g^{-1}_{\delta}(1)| < |p^{-1}_{\delta_{PA}}(0) \cap g^{-1}_{\delta}(1)|$. Thus $F1_{lsa}(g,p) < F1_{lsa}(g,\hat{p})$.

Case $b \geq 2$: Consider the following example:

$$\begin{aligned} g &= 110000000000 \\ p &= 100000000000 \\ \hat{p} &= 010000000000 \end{aligned}$$

Then it holds $F1_{lsa}(g,p) = 1 = F1_{lsa}(g,\hat{p})$.

Prop. 9 Consider the following example: $g = \mathbb{1}_{[0,2b-1]}$, $p = \mathbb{1}_{\{0\}} + \mathbb{1}_{\{2b-1\}}$, and $\hat{p} = \mathbb{1}_{\{0\}} + \mathbb{1}_{\{b\}}$. Then it holds $F1_{lsa}(g,p) = 1 = F1_{lsa}(g,\hat{p})$.

For $b = 3$:

$$\begin{aligned} g &= 111111000000 \\ p &= 100001000000 \\ \hat{p} &= 100100000000 \end{aligned}$$

Then it holds $F1_{lsa}(g,p) = 1 = F1_{lsa}(g,\hat{p})$. □

Theorem D.15. ***Point-adjusted F1-Score at $k\%$***

$$F1_{PA\%k}(g,p) = \frac{2\left(\sum_{\substack{W \in A(g):\\ \frac{|p^{-1}(1)\cap W|}{|W|} > k}} |W| + \sum_{\substack{W \in A(g):\\ \frac{|p^{-1}(1)\cap W|}{|W|} \leq k}} |p^{-1}(1) \cap W|\right)}{2\left(\sum_{\substack{W \in A(g):\\ \frac{|p^{-1}(1)\cap W|}{|W|} > k}} |W| + \sum_{\substack{W \in A(g):\\ \frac{|p^{-1}(1)\cap W|}{|W|} \leq k}} |p^{-1}(1) \cap W|\right) + |p^{-1}(1) \cap g^{-1}(0)| + \sum_{\substack{W \in A(g):\\ \frac{|p^{-1}(1)\cap W|}{|W|} \leq k}} |p^{-1}(0) \cap W|}$$

satisfies properties Prop. 1 and Prop. 5. Furthermore, it satisfies Prop. 6 for $k = 0$.

Proof.

Let $: [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p}\colon [n] \to \{0,1\}^n$ predictions. We define:

$$\#correct(p,k) := \sum_{\substack{W \in A(g):\\ \frac{|p^{-1}(1)\cap W|}{|W|} > k}} |W| + \sum_{\substack{W \in A(g):\\ \frac{|p^{-1}(1)\cap W|}{|W|} \leq k}} |p^{-1}(1) \cap W|$$

Then

$$F1_{PA\%k}(g,p) = \frac{2\#correct(p,k)}{2\#correct(p,k) + |p^{-1}(1) \cap g^{-1}(0)| + \sum_{\substack{W \in A(g):\\ \frac{|p^{-1}(1)\cap W|}{|W|} \leq k}} |p^{-1}(0) \cap W|}$$

Prop. 1 Let $W \in A(g)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $p_W = 0$, and $|\hat{p}_W^{-1}(1)| > 0$. Then $\#correct(\hat{p},k) > \#correct(p,k)$ while $|\hat{p}^{-1}(1) \cap g^{-1}(0)| = |p^{-1}(1) \cap g^{-1}(0)|$ and $\sum_{\substack{W \in A(g):\\ \frac{|\hat{p}^{-1}(1)\cap W|}{|W|} \leq k}} |\hat{p}^{-1}(0) \cap W| < \sum_{\substack{W \in A(g):\\ \frac{|p^{-1}(1)\cap W|}{|W|} \leq k}} |p^{-1}(0) \cap W|$. Thus $F1_{PA\%k}(g,\hat{p}) > F1_{PA\%k}(g,p)$.

Prop. 2 Consider the following example:

$$g = 000111111000$$
$$p = 000111000000$$
$$\hat{p} = 000111011000$$

Then it holds

- For $k < 0.5 : F1_{PA\%k}(g,p) = 1 = F1_{PA\%k}(g,\hat{p})$.
- For $0.5 \leq k < \frac{5}{6} : F1_{PA\%k}(g,p) = \frac{2}{3} < 1 = F1_{PA\%k}(g,\hat{p})$.
- For $\frac{5}{6} \leq k : F1_{PA\%k}(g,p) = \frac{2}{3} < \frac{10}{11} = F1_{PA\%k}(g,\hat{p})$.

Regardless of the choice of k this property does not hold.

Prop. 3 Consider the following example:

$$g = 000000000001$$
$$p = 000000010000$$
$$\hat{p} = 000010010000$$

Then it holds $F1_{PA\%k}(g,p) = 0 = F1_{PA\%k}(g,\hat{p})$.

Prop. 4 Consider the following example:

$$g = 000000111000$$
$$p = 001010111000$$
$$\hat{p} = 000110111000$$

Then it holds $F1_{PA\%k}(g,p) = \frac{3}{4} = F1_{PA\%k}(g,\hat{p})$.

Prop. 5 Let $W \in A(|g-1|)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $|\hat{p}_W^{-1}(1)| = |p_W^{-1}(1)|$, and $|A(\hat{p}_W)| = |A(p_W)|$. Then $\#correct(\hat{p},k) = \#correct(p,k)$, $\sum_{\substack{W \in A(g):\\ \frac{|\hat{p}^{-1}(1)\cap W|}{|W|} \leq k}} |\hat{p}^{-1}(0) \cap W| = \sum_{\substack{W \in A(g):\\ \frac{|p^{-1}(1)\cap W|}{|W|} \leq k}} |p^{-1}(0) \cap W|$, and $|\hat{p}^{-1}(1) \cap g^{-1}(0)| = |p^{-1}(1) \cap g^{-1}(0)|$. Thus $F1_{PA\%k}(g,\hat{p}) = F1_{PA\%k}(g,p)$.

Prop. 6 **Case** $k = 0$: In this case it holds $F1_{PA\%k}(g, p) = F1_{PA}(g, p)$ and thus by **??** this property holds.
Case $k > 0$: Consider the following example: Let $2 \leq k' \in \mathbb{N}$ such that $\frac{1}{k'} \leq k$. Such a k' exists because $k > 0$. $g = \mathbb{1}_{[a,a+k'-1]}$ with $a > 0$, $p = \mathbb{1}_{\{a\}}$, and $\hat{p} = \mathbb{1}_{\{0\}} + \mathbb{1}_{[a,a+k'-1]}$. Then it holds $F1_{PA\%k}(g, p) = \frac{2}{k'+1} \leq \frac{k'}{k'+1} = F1_{PA\%k}(g, \hat{p})$.

For $k \geq \frac{1}{3}$:

$$\begin{aligned} g &= 000000111000 \\ p &= 000000100000 \\ \hat{p} &= 000010111000 \end{aligned}$$

Then it holds $F1_{PA\%k}(g, p) = \frac{1}{2} < \frac{6}{7} = F1_{PA\%k}(g, \hat{p})$.

Prop. 7 **Case** $k = 1$: In this case it holds $F1_{PA\%k}(g, p) = F1(g, p)$ and thus by Thm. D.3 this property holds.

Case $k < 1$: Consider the following example: Let $k' \in \mathbb{N}$ such that $\frac{k'-1}{k'} > k$. Such a k' exists because $k < 0$. $g = \mathbb{1}_{[a,a+k'-1]}$ with $a > 0$, $p = \mathbb{1}_{[a,a+k'-2]}$, and $\hat{p} = \mathbb{1}_{[a,a+k'-1]}$. Then it holds $F1_{PA\%k}(g, p) = 1 = F1_{PA\%k}(g, \hat{p})$.

For $k < \frac{5}{6}$:

$$\begin{aligned} g &= 000111111000 \\ p &= 000111110000 \\ \hat{p} &= 000111111000 \end{aligned}$$

Then it holds $F1_{PA\%k}(g, p) = 1 = F1_{PA\%k}(g, \hat{p})$.

Prop. 8 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000111000000 \\ \hat{p} &= 000000111000 \end{aligned}$$

Then it holds

- For $k < \frac{1}{2}$ $F1_{PA\%k}(g, p) = 1 = F1_{PA\%k}(g, \hat{p})$.
- For $k \geq \frac{1}{2}$ $F1_{PA\%k}(g, p) = \frac{2}{3} = F1_{PA\%k}(g, \hat{p})$.

In either case the property does not hold.

Prop. 9 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000100100000 \\ \hat{p} &= 000110000000 \end{aligned}$$

Then it holds

- For $k < \frac{1}{3}$ $F1_{PA\%k}(g, p) = 1 = F1_{PA\%k}(g, \hat{p})$.
- For $k \geq \frac{1}{3}$ $F1_{PA\%k}(g, p) = \frac{1}{3} = F1_{PA\%k}(g, \hat{p})$.

In either case the property does not hold. □

Theorem D.16. ***Integrated point-adjusted F1-Score at $k\%$***

$$\int F1_{PA\%k}(g,p)$$

$$= \int_0^1 \frac{2\left(\sum_{\substack{W\in A(g):\\ \frac{|p^{-1}(1)\cap W|}{|W|}>k}} |W| + \sum_{\substack{W\in A(g):\\ \frac{|p^{-1}(1)\cap W|}{|W|}\leq k}} |p^{-1}(1)+W|\right)}{2\left(\sum_{\substack{W\in A(g):\\ \frac{|p^{-1}(1)\cap W|}{|W|}>k}} |W| + \sum_{\substack{W\in A(g):\\ \frac{|p^{-1}(1)\cap W|}{|W|}\leq k}} |p^{-1}(1)+W|\right) + |p^1(1)\cap g^{-1}(0)| + \sum_{\substack{W\in A(g):\\ \frac{|p^{-1}(1)\cap W|}{|W|}\leq k}} |p^{-1}(0)+W|} dk$$

has properties Prop. 1, Prop. 5, and Prop. 7.

Proof.

Let $: [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p}\colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 Let $W \in A(g)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $p_W = 0$, and $|\hat{p}_W^{-1}(1)| > 0$. Consider D.15, which implies that $F1_{PA\%k}(g,\hat{p}) > F1_{PA\%k}(g,p)$ for any k. Thus $\int F1_{PA\%k}(g,\hat{p}) > \int F1_{PA\%k}(g,p)$.

Prop. 2 Consider the counter example in D.15, which shows that $F1_{PA\%k}(g,\hat{p}) \geq F1_{PA\%k}(g,p)$ for any k. Thus $\int F1_{PA\%k}(g,\hat{p}) \geq \int F1_{PA\%k}(g,p)$ and this property does not hold.

Prop. 3 Consider the following example:

$$g = 000000000001$$
$$p = 000000010000$$
$$\hat{p} = 000010010000$$

Then it holds $\int F1_{PA\%k}(g,p) = 0 = \int F1_{PA\%k}(g,\hat{p})$.

Prop. 4 Consider the counter example in D.15, which shows that $F1_{PA\%k}(g,\hat{p}) = F1_{PA\%k}(g,p)$ for any k. Thus $\int F1_{PA\%k}(g,\hat{p}) = \int F1_{PA\%k}(g,p)$ and this property does not hold.

Prop. 5 Let $W \in A(|g-1|)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $|\hat{p}_W^{-1}(1)| = |p_W^{-1}(1)|$, and $|A(\hat{p}_W)| = |A(p_W)|$. Consider D.15, which implies that $F1_{PA\%k}(g,\hat{p}) = F1_{PA\%k}(g,p)$ for any k. Thus $\int F1_{PA\%k}(g,\hat{p}) = \int F1_{PA\%k}(g,p)$.

Prop. 6 Consider the following example:

$$g = 000000111000$$
$$p = 000000100000$$
$$\hat{p} = 000010111000$$

Then for $k \geq \frac{1}{3}$ it holds $F1_{PA\%k}(g,p) = \frac{1}{2}$ and $\frac{6}{7} = F1_{PA\%k}(g,\hat{p})$. For $k < \frac{1}{3}$ it holds that $F1_{PA\%k}(g,p) = 1$ and $\frac{6}{7} = F1_{PA\%k}(g,\hat{p})$. From this we can follow that $\int F1_{PA\%k}(g,p) = \frac{1}{3}1 + \frac{2}{3}\frac{1}{2} = \frac{2}{3}$ and $\int F1_{PA\%k}(g,\hat{p}) = \frac{6}{7}$. Thus $\int F1_{PA\%k}(g,p) = \frac{2}{3} < \frac{6}{7} = \int F1_{PA\%k}(g,\hat{p})$.

Prop. 7 Let $W \in A(g)$, $\hat{p}_{[n]\setminus\{i^*\}} = p_{[n]\setminus\{i^*\}}$ for some $i^* \in W$, $\hat{p}(i^*) \neq p(i^*) = 0$, and $|A(\hat{p}_W)| \leq |A(p_W)|$. Let $w_1 = min_{i\in W} i$, $w_2 = max_{i\in W} i$, $c = |p^{-1}(1) \cap W|$, then $c+1 = |\hat{p}^{-1}(1) \cap W|$.

- For $k < \frac{c}{|W|}$ $F1_{PA\%k}(g,p) = F1_{PA\%k}(g,\hat{p})$.
- For $\frac{c}{|W|} \leq k$ $F1_{PA\%k}(g,p) < F1_{PA\%k}(g,\hat{p})$.

From this we can follow that $\int F1_{PA\%k}(g,p) < \int F1_{PA\%k}(g,\hat{p})$.

Prop. 8 Consider the counter example in D.15, which shows that $F1_{PA\%k}(g,\hat{p}) = F1_{PA\%k}(g,p)$ for any k. Thus $\int F1_{PA\%k}(g,\hat{p}) = \int F1_{PA\%k}(g,p)$ and this property does not hold.

Prop. 9 Consider the counter example in D.15, which shows that $F1_{PA\%k}(g,\hat{p}) = F1_{PA\%k}(g,p)$ for any k. Thus $\int F1_{PA\%k}(g,\hat{p}) = \int F1_{PA\%k}(g,p)$ and this property does not hold. □

Theorem D.17. ***Point-Adjusted F1-score with Decay Function***

$$F1_{PAdf}(g,p) = \frac{2 \sum\limits_{\substack{W \in A(g): \\ W \cap p^{-1}(1) \neq \emptyset}} d^{\left(\min\limits_{i \in W \cap p^{-1}(1)} i - \min\limits_{i \in W} i\right)} |W|}{2 \sum\limits_{\substack{W \in A(g): \\ W \cap p^{-1}(1) \neq \emptyset}} |W| + |p^{-1}(1) \cap g^{-1}(0)| + \sum\limits_{\substack{W \in A(g): \\ W \cap p^{-1}(1) = \emptyset}} |W|}$$

where:
$d \in (0,1]$ = *decay rate*
has properties Prop. 1 and Prop. 5. It also has property Prop. 8 for $d < 1$ *or Prop. 6 if* $d = 1$.

Proof.

Let g: $[n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p}$: $[n] \to \{0,1\}^n$ predictions.

Prop. 1 **Case** $d = 1$ Then $F1_{PAdf}(g,p) = F1_{PA}(g,p)$ and this property is fulfilled following Thm. D.6.

Case $d < 1$ Consider the following example: Let $a \in \mathbb{N}$ be large enough such that $\frac{1}{a+1} > d^{a+2}$. $g = \mathbb{1}_{\{0\}} + \mathbb{1}_{[2,a]}$, $p = \mathbb{1}_{\{0\}}$, and $\hat{p} = \mathbb{1}_{\{0\}} + \mathbb{1}_{\{a\}}$. Then it holds $F1_{PAdf}(g,p) = \frac{2}{a+1} > \frac{2+2d^{a-2}(a-1)}{2a} = F1_{PAdf}(g,\hat{p})$.

For example for $d \leq 0.7$ (this example also holds for some more $0.7 < d < 0.8$, however, the exact value is irrational):

$$\begin{aligned} g &= 101111111111 \\ p &= 100000000000 \\ \hat{p} &= 100000000001 \end{aligned}$$

Then it holds $F1_{PAdf}(g,p) = \frac{2}{12} > \frac{2+20d^9}{22} = F1_{PAdf}(g,\hat{p})$.

Prop. 2 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000100000000 \\ \hat{p} &= 000100001000 \end{aligned}$$

Then it holds $F1_{PAdf}(g,p) = 1 = F1_{PAdf}(g,\hat{p})$.

Prop. 3 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000000000000 \\ \hat{p} &= 010000000000 \end{aligned}$$

Then it holds $F1_{PAdf}(g,p) = 0 = F1_{PAdf}(g,\hat{p})$.

Prop. 4 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 001000000000 \\ \hat{p} &= 101000000000 \end{aligned}$$

Then it holds $F1_{PAdf}(g,p) = 0 = F1_{PAdf}(g,\hat{p})$.

Prop. 5 Let $W \in A(|g-1|)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $|\hat{p}_W^{-1}(1)| = |p_W^{-1}(1)|$, and $|A(\hat{p}_W)| = |A(p_W)|$. Then every component is equal and thus $F1_{PAdf}(g,p) = F1_{PAdf}(g,\hat{p})$.

Prop. 6 **Case** $d = 1$ Then $F1_{PAdf}(g,p) = F1_{PA}(g,p)$ and this property is fulfilled following Thm. D.6.

Case $d < 1$ Consider the following example: Let $a \in \mathbb{N}$ be large enough such that $d^{a-2} < \frac{2a-2}{2a-1}$. $g = \mathbb{1}_{\{0\}} + \mathbb{1}_{[2,a]}$, $p = \mathbb{1}_{\{a\}}$, and $\hat{p} = \mathbb{1}_{\{0\}} + \mathbb{1}_{\{2\}}$. Then it holds $F1_{PAdf}(g,p) = d^{a-2} < \frac{2a-2}{2a-1} = F1_{PAdf}(g,\hat{p})$.

For example for $d \leq 0.99$:

$$\begin{aligned} g &= 001111111111 \\ p &= 000000000001 \\ \hat{p} &= 101000000000 \end{aligned}$$

Then it holds $F1_{PAdf}(g,p) = d^9 < \frac{20}{21} = F1_{PAdf}(g,\hat{p})$.

Prop. 7 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000100000000 \\ \hat{p} &= 000110000000 \end{aligned}$$

Then it holds $F1_{PAdf}(g,p) = 1 = F1_{PAdf}(g,\hat{p})$.

Prop. 8 **Case** $d = 1$ Then $F1_{PAdf}(g,p) = F1_{PA}(g,p)$ and this property is not fulfilled following Thm. D.6.

Case $d < 1$ Let $W \in A(g)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $|A(\hat{p}_W)| = |A(p_W)|$, $|\hat{p}^{-1}(1)| = |p^{-1}(1)|$, and $\min\limits_{i \in W\colon p(i)=1} i < \min\limits_{i \in W\colon \hat{p}(i)=1} i$. Then $(\min\limits_{i \in W \cap \hat{p}^{-1}(1)} i - \min\limits_{i \in W} i) < (\min\limits_{i \in W \cap p^{-1}(1)} i - \min\limits_{i \in W} i)$, while all other components remain the same. Thus $F1_{PAdf}(g,p) < F1_{PAdf}(g,\hat{p})$.

Prop. 9 **Case** $d = 1$ Then $F1_{PAdf}(g,p) = F1_{PA}(g,p)$ and this property is not fulfilled following Thm. D.6.

Case $d < 1$ Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000100001000 \\ \hat{p} &= 000110000000 \end{aligned}$$

Then it holds $F1_{PAdf}(g,p) = 1 = F1_{PAdf}(g,\hat{p})$. □

Theorem D.18. ***Reduced-Length F1-Score***

$$F1_{reduced-length}(g,p) = \frac{2 \sum\limits_{\substack{W \in A(g)\colon \\ W \cap p^{-1}(1) \neq \emptyset}} \ln |W|}{2 \sum\limits_{\substack{W \in A(g)\colon \\ W \cap p^{-1}(1) \neq \emptyset}} \ln |W| + |p^{-1}(1) \cap g^{-1}(0)| + \sum\limits_{\substack{W \in A(g)\colon \\ W \cap p^{-1}(1) = \emptyset}} \ln |W|}$$

has properties Prop. 1, Prop. 5, and Prop. 6.

Proof.

Let $: [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p}\colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 Let $W \in A(g)$ an interval, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $p_W = 0$, and $|\hat{p}_W^{-1}(1)| > 0$. Then $\{W \in A(g)\colon W \cap \hat{p}^{-1}(1) \neq \emptyset\} = \{W \in A(g)\colon W \cap p^{-1}(1) \neq \emptyset\} \cup \{W\}$ with $|W| > 0$. Thus, $F1_{reduced-length}(g,p) < F1_{reduced-length}(g,\hat{p})$.

Prop. 2 Consider the following example:

$$g = 000111111000$$
$$p = 000111000000$$
$$\hat{p} = 000111011000$$

Then it holds $F1_{reduced-length}(g,p) = 1 = F1_{reduced-length}(g,\hat{p})$.

Prop. 3 Consider the following example:

$$g = 000000000001$$
$$p = 000000010000$$
$$\hat{p} = 000010010000$$

Then it holds $F1_{reduced-length}(g,p) = 0 = F1_{reduced-length}(g,\hat{p})$.

Prop. 4 Consider the following example:

$$g = 000000111000$$
$$p = 010010010000$$
$$\hat{p} = 001100010000$$

Then it holds $F1_{reduced-length}(g,p) = \frac{\ln 3}{1+\ln 3} = F1_{reduced-length}(g,\hat{p})$.

Prop. 5 Let $W \subset W' \in A(|g-1|)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $|\hat{p}_W^{-1}(1)| = |p_W^{-1}(1)|$, and $|A(\hat{p}_W)| = |A(p_W)|$. Then $\{W \in A(g)\colon W \cap \hat{p}^{-1}(1) \neq \emptyset\} = \{W \in A(g)\colon W \cap p^{-1}(1) \neq \emptyset\}$ and $|\hat{p}^{-1}(1) \cap g^{-1}(0)| = |p^{-1}(1) \cap g^{-1}(0)|$. Thus, $F1_{reduced-length}(g,p) = F1_{reduced-length}(g,\hat{p})$.

Prop. 6 Let $W \in A(g)$, $\hat{W} \in A(|g-1|)$, $|A(\hat{p}_W)| = |A(p_W)|$, $\hat{p}_{[n]\setminus(W\cup\hat{W})} = p_{[n]\setminus(W\cup\hat{W})}$, $p_{\hat{W}} = 0$, and $\hat{p}_{\hat{W}} = \mathbb{1}_{\{i^*\}}$ for some $i^* \in \hat{W}$. Then $\{W \in A(g)\colon W \cap \hat{p}^{-1}(1) \neq \emptyset\} = \{W \in A(g)\colon W \cap p^{-1}(1) \neq \emptyset\}$ and $|\hat{p}^{-1}(1) \cap g^{-1}(0)| = |p^{-1}(1) \cap g^{-1}(0)| + 1$. Thus, $F1_{reduced-length}(g,p) > F1_{reduced-length}(g,\hat{p})$.

Prop. 7 Consider the following example:

$$g = 000111111000$$
$$p = 000010000000$$
$$\hat{p} = 000011000000$$

Then it holds $F1_{reduced-length}(g,p) = 1 = F1_{reduced-length}(g,\hat{p})$.

Prop. 8 Consider the following example:

$$g = 000111111000$$
$$p = 000110000000$$
$$\hat{p} = 000011000000$$

Then it holds $F1_{reduced-length}(g,p) = 1 = F1_{reduced-length}(g,\hat{p})$.

Prop. 9 Consider the following example:

$$g = 000111111000$$
$$p = 000010010000$$
$$\hat{p} = 000110000000$$

Then it holds $F1_{reduced-length}(g,p) = 1 = F1_{reduced-length}(g,\hat{p})$. $\square$

Theorem D.19. ***Balanced Point Adjusted F1-Score***

$$F1_{BA}(g,p) = \frac{2TP(g,p)}{2TP(g,p) + FP(g,p) + FN(g,p)}$$

where $B \in 2\mathbb{N}$ is an even integer and

$$\begin{aligned}
TP(g,p) &= \sum_{\substack{W \in A(g):\\ W \cap p^{-1}(1) \neq \emptyset}} |W| + \sum_{\substack{W \in A(g):\\ W \cap p^{-1}(1) = \emptyset}} |\{i \in W : p^{-1}(1) \cap g^{-1}(0) \cap [i - \frac{B}{2}, i + \frac{B}{2} + 1]\} | \neq \emptyset\}| \\
FP(g,p) &= |\{i \in p^{-1}(0) \cap g^{-1}(0) : p^{-1}(1) \cap g^{-1}(0) \cap [i - \frac{B}{2}, i + \frac{B}{2} + 1] \neq \emptyset\}| + |p^{-1}(1) \cap g^{-1}(0)| \\
FN(g,p) &= \sum_{\substack{W \in A(g):\\ W \cap p^{-1}(1) = \emptyset}} |\{i \in W : p^{-1}(1) \cap g^{-1}(0) \cap [i - \frac{B}{2}, i + \frac{B}{2} + 1] = \emptyset\}|
\end{aligned}$$

has no properties.

Proof.

For simplicity we slightly modify the above definition to use just an integer $b \in \mathbb{N}$ and we consider the intersection with intervals above to be entirely balanced. Namely, the intervals of the form $[i - \frac{B}{2}, i + \frac{B}{2} + 1]$ are changed to $[i - B, i + B]$. This changes the definition as follows:

$$\begin{aligned}
TP(g,p) &= \sum_{\substack{W \in A(g):\\ W \cap p^{-1}(1) \neq \emptyset}} |W| + \sum_{\substack{W \in A(g):\\ W \cap p^{-1}(1) = \emptyset}} |\{i \in W : p^{-1}(1) \cap g^{-1}(0) \cap [i - B, i + B]\} | \neq \emptyset\}| \\
FP(g,p) &= |\{i \in p^{-1}(0) \cap g^{-1}(0) : p^{-1}(1) \cap g^{-1}(0) \cap [i - B, i + B] \neq \emptyset\}| + |p^{-1}(1) \cap g^{-1}(0)| \\
FN(g,p) &= \sum_{\substack{W \in A(g):\\ W \cap p^{-1}(1) = \emptyset}} |\{i \in W : p^{-1}(1) \cap g^{-1}(0) \cap [i - B, i + B] = \emptyset\}|,
\end{aligned}$$

where $B \in \mathbb{N}$. All the following proof can be modified to prove the original definition's lack of properties by slightly modifying each of the counterexamples. We leave this slight modification to the reader.

Let $: [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p} \colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 Consider the following example: $g = \mathbb{1}_{\{a\} \cup [a+2B+1, a+4B+1]}$ with $a \geq B$, $p = \mathbb{1}_{\{a+3B+1\}}$, and $\hat{p} = \mathbb{1}_{\{a\}} + \mathbb{1}_{\{a+3B+1\}}$. Then it holds $F1_{BA}(g,p) = \frac{2(2B+1)}{2(2B+1)+1} > \frac{2(2B+2)}{2(2B+2)+2B} = F1_{BA}(g,\hat{p})$.

For $B = 1$:

$$\begin{aligned}
g &= 000001000111 \\
p &= 000000000010 \\
\hat{p} &= 000001000010
\end{aligned}$$

Then it holds $F1_{BA}(g,p) = \frac{6}{7} > \frac{4}{5} = F1_{BA}(g,\hat{p})$.

Prop. 2 Consider the following example: $g = \mathbb{1}_{[a, a+4B+1]}$ with $a > 0$, $p = \mathbb{1}_{\{a\}}$, and $\hat{p} = \mathbb{1}_{\{a\}} + \mathbb{1}_{\{a+2B+1\}}$. Then it holds $F1_{BA}(g,p) = 1 = F1_{BA}(g,\hat{p})$.

For $B \leq 1$:

$$\begin{aligned}
g &= 000111111000 \\
p &= 000010000000 \\
\hat{p} &= 000010010000
\end{aligned}$$

Then it holds $F1_{BA}(g,p) = 1 = F1_{BA}(g,\hat{p})$.

Prop. 3 Consider the following example: $g = \mathbb{1}_{\{a\}}$ with $a > B+1$, $p = \mathbb{1}_{\{a-B-1\}}$, and $\hat{p} = \mathbb{1}_{[a-B-1,a-B]}$. Then it holds $F1_{BA}(g,p) = 0 < \frac{2}{2B+3} = F1_{BA}(g,\hat{p})$.

For $B = 1$:

$$\begin{aligned} g &= 000000100000 \\ p &= 000010000000 \\ \hat{p} &= 000011000000 \end{aligned}$$

Then it holds $F1_{BA}(g,p) = 0 < \frac{2}{5} = F1_{BA}(g,\hat{p})$.

Prop. 4 Consider the following example:

$$\begin{aligned} g &= 000000000000 \\ p &= 000010100000 \\ \hat{p} &= 000011000000 \end{aligned}$$

Then it holds $F1_{BA}(g,p) = 0 = F1_{BA}(g,\hat{p})$.

Prop. 5 Consider the following example: $g = \mathbb{1}_{\{a\}}$ with $a > B+1$, $p = \mathbb{1}_{\{a-B-1\}}$, and $\hat{p} = \mathbb{1}_{\{a-B\}}$. Then it holds $F1_{BA}(g,p) = 0 < \frac{2}{2B+2} = F1_{BA}(g,\hat{p})$.

For $B = 1$:

$$\begin{aligned} g &= 000000100000 \\ p &= 000010000000 \\ \hat{p} &= 000001000000 \end{aligned}$$

Then it holds $F1_{BA}(g,p) = 0 < \frac{1}{2} = F1_{BA}(g,\hat{p})$.

Prop. 6 Consider the following example: $g = \mathbb{1}_{[a,a+B]} + \mathbb{1}_{[a+B+2,a+2B+2]}$ with $a > B$, $p = \mathbb{1}_{\{a,a+2B+2\}}$, and $\hat{p} = \mathbb{1}_{[a+B,a+B+2]}$. Then it holds $F1_{BA}(g,p) = \frac{4}{5} < \frac{8B}{8B+1} = F1_{BA}(g,\hat{p})$.

For $B = 1$:

$$\begin{aligned} g &= 001101100000 \\ p &= 001000100000 \\ \hat{p} &= 000111000000 \end{aligned}$$

Then it holds $F1_{BA}(g,p) = \frac{4}{5} < \frac{8}{9} = F1_{BA}(g,\hat{p})$.

Prop. 7 Consider the following example: $g = \mathbb{1}_{[a,a+2B]}$ with $a > B$, $p = \mathbb{1}_{\{a+B\}}$, and $\hat{p} = \mathbb{1}_{[a+B,a+B+1]}$. Then it holds $F1_{BA}(g,p) = 1 > \frac{2(2B+1)}{2(2B+1)+1} = F1_{BA}(g,\hat{p})$.

For $B = 1$:

$$\begin{aligned} g &= 000111000000 \\ p &= 000010000000 \\ \hat{p} &= 000011000000 \end{aligned}$$

Then it holds $F1_{BA}(g,p) = 1 > \frac{6}{7} = F1_{BA}(g,\hat{p})$.

Prop. 8 Consider the following example: $g = \mathbb{1}_{[a,a+2B+1]}$ with $a > 0$, $p = \mathbb{1}_{\{a+B+1\}}$, and $\hat{p} = \mathbb{1}_{\{a+B\}}$. Then it holds $F1_{BA}(g,p) = 1 = F1_{BA}(g,\hat{p})$.

For $B = 1$:

$$g = 000111100000$$
$$p = 000001000000$$
$$\hat{p} = 000010000000$$

Then it holds $F1_{BA}(g, p) = 1 = F1_{BA}(g, \hat{p})$.

Prop. 9 Consider the following example: $g = \mathbb{1}_{[a,a+2B+2]}$ with $a > 0$, $p = \mathbb{1}_{\{a+B,a+B+2\}}$, and $\hat{p} = \mathbb{1}_{[a+B,a+B+1]}$. Then it holds $F1_{BA}(g, p) = 1 = F1_{BA}(g, \hat{p})$.

For $B = 1$:

$$g = 000111110000$$
$$p = 000010100000$$
$$\hat{p} = 000011000000$$

Then it holds $F1_{BA}(g, p) = 1 = F1_{BA}(g, \hat{p})$. □

Theorem D.20. ***Range-Based Precision*** *defined as*

$$Precision_{TS}(g, p) = \frac{1}{|A(p)|} \sum_{W \in A(p)} \gamma'(W, p) \sum_{W_g \in A(g)} \omega(W, W \cap W_g, \delta)$$

where

$$\begin{aligned}
\delta(i; W) &= |W| - \big(i - \min(W)\big) + 1 \geq 1, \\
0 \leq \omega\big(W, S, \delta\big) &= \frac{\sum_{i \in W} \delta(i; W)\, \mathbb{1}_{\{i \in S\}}}{\sum_{i \in W} \delta(i; W)} \leq 1, \\
k(W, g) &= |\{W_g \in A(g) \colon W \cap W_g \neq \emptyset\}|, \\
\gamma'(W, g) &= \frac{1}{k(W, g)} \mathbb{1}_{k(W,g)>1}(W) + \mathbb{1}_{k(W,g)\leq 1}(W).
\end{aligned}$$

Proof.

Notice that for δ usually $i \in W = [\ell, u]$ and thus $\delta(i; W) = u - (i - 1) + 1$. Let $g \colon [n] \to \{0, 1\}^n$ be the ground truth and $p, \hat{p} \colon [n] \to \{0, 1\}^n$ predictions.

Prop. 1 Consider the following example:

$$g = 000111000111$$
$$p = 000111000000$$
$$\hat{p} = 000111000001$$

Then it holds $\text{Precision}_{TS}(g, p) = 1 = \text{Precision}_{TS}(g, \hat{p})$.

Prop. 2 Consider the following example:

$$g = 000111111000$$
$$p = 000100000000$$
$$q = 000110011000$$

Then it holds $\text{Precision}_{TS}(g, p) = 1 = \text{Precision}_{TS}(g, q)$.

Prop. 3 Consider the following example:

$$\begin{aligned} g &= 000000000000, \\ p &= 000001100000, \\ q &= 000011100000. \end{aligned}$$

$\text{Precision}_{TS}(g, p) = 0 = \text{Precision}_{TS}(g, q)$.

Prop. 4 Consider the following example:

$$\begin{aligned} g &= 000000000001, \\ p &= 000100000000, \\ q &= 000100100000. \end{aligned}$$

Then it follows $\text{Precision}_{TS}(g, p) = 0 = \text{Precision}_{TS}(g, q)$.

Prop. 5 Consider the following example:

$$\begin{aligned} g &= 000011110000, \\ p &= 110011000000, \\ q &= 001111000000. \end{aligned}$$

Then it holds $\text{Precision}_{TS}(g, p) = 0.5 > 0.3 = \text{Precision}_{TS}(g, q)$

Prop. 6 Let $W \in A(g)$, $\hat{W} \in A(|g-1|)$, $|A(\hat{p}_W)| = |A(p_W)|$, $\hat{p}_{[n]\setminus(W\cup\hat{W})} = p_{[n]\setminus(W\cup\hat{W})}$, $p_{\hat{W}} = 0$, and $\hat{p}_{\hat{W}} = \mathbb{1}_{\{i^*\}}$ for some $i^* \in \hat{W}$. Consider the following example:

$$\begin{aligned} g &= 011111111110, \\ p &= 000000001011, \\ q &= 111111111011. \end{aligned}$$

Then it follows $\text{Precision}_{TS}(g, p) = 0.6667 < 0.7222 = \text{Precision}_{TS}(g, q)$.

Prop. 7 Consider the following example:

$$\begin{aligned} g &= 000111000000, \\ p &= 000110000000, \\ q &= 000100000000. \end{aligned}$$

Then it holds $\text{Precision}_{TS}(g, p) = 1 = \text{Precision}_{TS}(g, q)$.

Prop. 8 Consider the following example:

$$\begin{aligned} g &= 000111100000, \\ p &= 000110000000, \\ q &= 000011000000. \end{aligned}$$

Then it holds $\text{Precision}_{TS}(g, p) = 1.0 = \text{Precision}_{TS}(g, q)$.

Prop. 9 Consider the following example:

$$\begin{aligned} g &= 000111100000, \\ p &= 000100000000, \\ q &= 000000100000. \end{aligned}$$

Then it holds $\text{Precision}_{TS}(g, p) = 1.0 = \text{Precision}_{TS}(g, q)$. $\square$

Theorem D.21. ***Range-Based Recall*** *defined as*

$$Recall_{TS}(g,p) = \frac{1}{|A(g)|} \sum_{W \in A(g)} (\alpha \mathbb{1}_{\exists W_p \in A(p): W \cap W_p \neq \emptyset} + (1-\alpha)\gamma'(W,p) \sum_{W_p \in A(p)} \omega(W, W \cap W_p, \delta))$$

where $0 \leq \alpha \leq 1$,

$$\begin{aligned}
\delta(i;W) &= |W| - \big(i - \min(W)\big) + 1 \geq 1, \\
0 \leq \omega\big(W,S,\delta\big) &= \frac{\sum_{i \in W} \delta(i;W)\, \mathbb{1}_{\{i \in S\}}}{\sum_{i \in W} \delta(i;W)} \leq 1, \\
k(W,p) &= |\{W_p \in A(p) \colon W \cap W_p \neq \emptyset\}|, \\
\gamma'(W,p) &= \frac{1}{k(W,p)} \mathbb{1}_{k(W,p)>1}(W) + \mathbb{1}_{k(W,p)\leq 1}(W).
\end{aligned}$$

Proof.

Let $g \colon [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p} \colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 Let $W \in A(g)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $p_W = 0$, and $|\hat{p}_W^{-1}(1)| > 0$. Then there is $i_0 \in W$ with $\hat{p}(i_0) = 1$. Then

$$\begin{aligned}
0 &= \alpha \mathbb{1}_{\exists W_p \in A(p): W \cap W_p \neq \emptyset} + (1-\alpha)\gamma'(W,p) \sum_{W_p \in A(p)} \omega(W, W \cap W_p, \delta) \\
0 &< \alpha \mathbb{1}_{\exists W_{\hat{p}} \in A(\hat{p}): W \cap W_{\hat{p}} \neq \emptyset} + (1-\alpha)\gamma'(W,\hat{p}) \sum_{W_{\hat{p}} \in A(\hat{p})} \omega(W, W \cap W_{\hat{p}}, \delta)
\end{aligned}$$

While all other summands are equal and thus $\text{Recall}_{TS}(g,\hat{p}) > \text{Recall}_{TS}(g,p)$.

Prop. 2 Consider the following example:

$$\begin{aligned}
g &= 111111111111 \\
p &= 100000000000 \\
\hat{p} &= 110011111111
\end{aligned}$$

Then it holds $Recall_{TS}(g,p) = 0.1667 < 0.8846 = Recall_{TS}(g,\hat{p})$.

Prop. 3 Consider the following example:

$$\begin{aligned}
g &= 000001000000 \\
p &= 000010000000 \\
\hat{p} &= 000110000000
\end{aligned}$$

Then it holds $Recall_{TS}(g,p) = 0 = Recall_{TS}(g,\hat{p})$.

Prop. 4 Consider the following example:

$$\begin{aligned}
g &= 000010000000 \\
p &= 000001000000 \\
\hat{p} &= 000101000000
\end{aligned}$$

Then it holds $Recall_{TS}(g,p) = 0 = Recall_{TS}(g,\hat{p})$.

Prop. 5 Let $W \in A(|g-1|)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $|\hat{p}_W^{-1}(1)| = |p_W^{-1}(1)|$, and $|A(\hat{p}_W)| = |A(p_W)|$. Then, since false positives do not affect the metric at all, $Recall_{TS}(g,p) = Recall_{TS}(g,\hat{p})$.

Prop. 6 Consider the following example:

$$g = 000010000000$$
$$p = 000001000000$$
$$\hat{p} = 000101000000$$

Then it holds $Recall_{TS}(g,p) = 0 = Recall_{TS}(g,q)$.

Prop. 7 Let $W \in A(g)$, $i^* \in W$, $\hat{p}_{[n]\setminus\{i^*\}} = p_{[n]\setminus\{i^*\}}$, $p(i^*) = 0$, $\hat{p}(i^*) = 1$, and $|A(\hat{p}_W)| \leq |A(p_W)|$. Then,

$$\begin{aligned}
&(\alpha \mathbb{1}_{\exists W_p \in A(p): W \cap W_p \neq \emptyset} + (1-\alpha)\gamma'(W,p) \sum_{W_p \in A(p)} \omega(W, W \cap W_p, \delta)) \\
\leq &(\alpha \mathbb{1}_{\exists W_{\hat{p}} \in A(\hat{p}): W \cap W_{\hat{p}} \neq \emptyset} + (1-\alpha)\gamma'(W,\hat{p}) \sum_{W_p \in A(p)} \omega(W, W \cap W_p, \delta)) \\
< &(\alpha \mathbb{1}_{\exists W_{\hat{p}} \in A(\hat{p}): W \cap W_{\hat{p}} \neq \emptyset} + (1-\alpha)\gamma'(W,\hat{p}) \sum_{W_{\hat{p}} \in A(\hat{p})} \omega(W, W \cap W_{\hat{p}}, \delta))
\end{aligned}$$

while all other summands remain the same. Thus $Recall_{TS}(g,p) < Recall_{TS}(g,\hat{p})$.

Prop. 8 Consider the following example:

$$g = 111111111111$$
$$p = 100000000001$$
$$\hat{p} = 010100000000$$

Then it holds $Recall_{TS}(g,p) = 0.1923 < 0.2821 = Recall_{TS}(g,\hat{p})$.

Prop. 9 Let $W \in A(g)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, Let $p, \hat{p}$ be two predictions with $|A(\hat{p}_W)| \leq |A(p_W)|$, let $i^* \in W$ with $p(i^*) = 1$, let $i^{**} < i^* \in W$ with $p(i^{**}) = 0$, and $p'_W(i) = \mathbb{1}_{\{i^{**}\}}(i) + p(i)\mathbb{1}_{W\setminus\{i^*,i^{**}\}}(i)$. Then,

$$\begin{aligned}
&\alpha \mathbb{1}_{\exists W_p \in A(p): W \cap W_p \neq \emptyset} + (1-\alpha)\gamma'(W,p) \sum_{W_p \in A(p)} \omega(W, W \cap W_p, \delta) \\
\leq &\alpha \mathbb{1}_{\exists W_{\hat{p}} \in A(\hat{p}): W \cap W_{\hat{p}} \neq \emptyset} + (1-\alpha)\gamma'(W,\hat{p}) \sum_{W_p \in A(p)} \omega(W, W \cap W_p, \delta)
\end{aligned}$$

Now $\sum_{W_p \in A(p)} \omega(W, W \cap W_p, \delta) = \frac{\sum_{i \in I} \delta(i;W)}{d}$ For some set I and $d = \sum_{i \in W} \delta(i;W)$. However, $\sum_{W_{\hat{p}} \in A(\hat{p})} \omega(W, W \cap W_{\hat{p}}, \delta) = \frac{\sum_{i \in I \setminus i^* \cup i^{**}} \delta(i;W)}{d}$ with $i^{**} < i^*$. Thus,

$$\begin{aligned}
&\alpha \mathbb{1}_{\exists W_{\hat{p}} \in A(\hat{p}): W \cap W_{\hat{p}} \neq \emptyset} + (1-\alpha)\gamma'(W,\hat{p}) \sum_{W_p \in A(p)} \omega(W, W \cap W_p, \delta) \\
< &\alpha \mathbb{1}_{\exists W_{\hat{p}} \in A(\hat{p}): W \cap W_{\hat{p}} \neq \emptyset} + (1-\alpha)\gamma'(W,\hat{p}) \sum_{W_{\hat{p}} \in A(\hat{p})} \omega(W, W \cap W_{\hat{p}}, \delta)
\end{aligned}$$

Notice that all other summands remain the same. Thus, finally $Recall_{TS}(g,p) < Recall_{TS}(g,\hat{p})$.

□

Theorem D.22. ***Range-Based F1-Score*** *defined as*

$$F1_{TS}(g,p) = 2\frac{Precision_{TS}(g,p)Recall_{TS}(g,p)}{Precision_{TS}(g,p) + Recall_{TS}(g,p)}$$

where

$$Precision_{TS}(g,p) = \frac{1}{|A(p)|} \sum_{W \in A(p)} \gamma'_p(W,g) \sum_{W_g \in A(g)} \omega_p(W, W \cap W_g, \delta_p)$$

where $\delta_p \geq 1$, $0 \leq \omega_p \leq 1$, *and*

$$\gamma'_p(W,p) = \gamma_p(W,g)\mathbb{1}_{|\{W_p \in A(p)\colon W \cap W_p \neq \emptyset\}| \leq 1}(W) + \mathbb{1}_{|\{W_p \in A(p)\colon W \cap W_p \neq \emptyset\}| > 1}(W)$$

where $0 \leq \gamma_p \leq 1$ *and*

$$Recall_{TS}(g,p) = \frac{1}{|A(g)|} \sum_{W \in A(g)} (\alpha \mathbb{1}(\sum_{W_p \in A(p)} |W \cap W_p| > 0) + (1-\alpha)\gamma'_r(W,p) \sum_{W_p \in A(p)} \omega_r(W, W \cap W_p, \delta_r)$$

where $0 \leq \alpha \leq 1$, $\delta_r \geq 1$, $0 \leq \omega_r \leq 1$, *and*

$$\gamma'_r(W,p) = \gamma_r(W,p)\mathbb{1}_{|\{W_p \in A(p)\colon W \cap W_p \neq \emptyset\}| \leq 1}(W) + \mathbb{1}_{|\{W_p \in A(p)\colon W \cap W_p \neq \emptyset\}| > 1}(W)$$

where $0 \leq \gamma_r \leq 1$

Proof.

Let $g\colon [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p}\colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 Let $W \in A(g)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $p_W = 0$, and $|\hat{p}_W^{-1}(1)| > 0$.

$$\begin{aligned}
\text{Precision}_{TS}(g,\hat{p}) &\geq \text{Precision}_{TS}(g,p), \\
\text{Recall}_{TS}(g,\hat{p}) &> \text{Recall}_{TS}(g,p), \\
F1_{TS}(g,p) &= 2\,\frac{\text{Precision}_{TS}(g,p) \times \text{Recall}_{TS}(g,p)}{\text{Precision}_{TS}(g,p) + \text{Recall}_{TS}(g,p)},
\end{aligned}$$

from the above it follows $F1_{TS}(g,\hat{p}) > F1_{TS}(g,p)$.

Prop. 2 Consider the following example:

$$\begin{aligned}
g &= 111111111111 \\
p &= 100000000000 \\
\hat{p} &= 110011111111
\end{aligned}$$

Then it holds $F1_{TS}(g,p) = 0.2667 < 0.5488 = F1_{TS}(g,\hat{p})$.

Prop. 3 Consider the following example:

$$\begin{aligned}
g &= 000001000000 \\
p &= 000010000000 \\
\hat{p} &= 000110000000
\end{aligned}$$

Then it holds $F1_{TS}(g,p) = 0.0000 = F1_{TS}(g,\hat{p})$.

Prop. 4 Consider the following example:

$$\begin{aligned}
g &= 000010000000 \\
p &= 000001000000 \\
\hat{p} &= 000101000000
\end{aligned}$$

Then it holds $F1_{TS}(g,p) = 0.0000 = F1_{TS}(g,\hat{p})$.

Prop. 5 Consider the following example:

$$\begin{aligned} g &= 000001000000 \\ p &= 000101000000 \\ \hat{p} &= 000011000000 \end{aligned}$$

Then it holds $F1_{TS}(g,p) = 0.6667 > 0.5000 = F1_{TS}(g,\hat{p})$.

Prop. 6 Consider the following example:

$$\begin{aligned} g &= 000010000000 \\ p &= 000001000000 \\ \hat{p} &= 000101000000 \end{aligned}$$

Then it holds $F1_{TS}(g,p) = 0.000 = 0.0000 = F1_{TS}(g,\hat{p})$.

Prop. 7 Let $W \in A(g)$, $\hat{p}_{[n]\setminus\{i^*\}} = p_{[n]\setminus\{i^*\}}$ for some $i^* \in W$, $\hat{p}(i^*) \neq p(i^*) = 0$, and $|A(\hat{p}_W)| \leq |A(p_W)|$.

$$\begin{aligned} \text{Precision}_{TS}\big(g,\hat{p}\big) &= \text{Precision}_{TS}\big(g,p\big), \\ \text{Recall}_{TS}\big(g,\hat{p}\big) &> \text{Recall}_{TS}\big(g,p\big), \\ F1_{TS}\big(g,p\big) &= 2\,\frac{\text{Precision}_{TS}\big(g,p\big) \times \text{Recall}_{TS}\big(g,p\big)}{\text{Precision}_{TS}\big(g,p\big) + \text{Recall}_{TS}\big(g,p\big)}, \end{aligned}$$

from the above it follows $F1_{TS}\big(g,\hat{p}\big) > F1_{TS}\big(g,p\big)$.

Prop. 8 Consider the following example:

$$\begin{aligned} g &= 111111111111 \\ p &= 100000000001 \\ \hat{p} &= 010100000000 \end{aligned}$$

Then it holds $F1_{TS}(g,p) = 0.1538 < 0.2273 = F1_{TS}(g,\hat{p})$.

Prop. 9 Let $W \in A(g)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, Consider the following example:

$$\begin{aligned} g &= 000000000011 \\ p &= 111111111110 \\ \hat{p} &= 111111111101 \end{aligned}$$

Then it holds $F1_{TS}(g,p) = 0.0296 < 0.4000 = F1_{TS}(g,\hat{p})$.

□

Theorem D.23. ***Time-Series Precision*** *defined as*

$$TPrec(g,p) = \frac{1}{\sum\limits_{W_p \in A(p)} |W_p|} \sum_{W \in A(p)} |W| \gamma(|\{W_g \in A(g) \colon W \cap W_g \neq \emptyset\}|) \sum_{W_g \in A(g)} \frac{\sum\limits_{i \in W_g \cap W} \delta(i - \min W, |W|)}{\sum\limits_{i \in W} \delta(i - \min W, |W|)}$$

where $\delta \geq 1$, *and* γ *monotonically decreasing with* $\gamma(1) = 1$ *and*

$$\gamma(1,W) = 1, \qquad \gamma(n,W) = \max_{0<m<n} \frac{\sum\limits_{i \in W} \delta(i - \min W, |W|) - n + m}{\sum\limits_{i \in W} \delta(i - \min W, |W|)} \gamma(m,W)$$

$$\Longrightarrow \delta(pos, |W|) \;=\; |W| \;-\; pos \;+\; 1 \quad (\textit{front bias for each } i \in W)$$

Proposition D.24 (Front-Bias $\implies$ Strictly Positive Coverage)**.**

$$i \in W = [m, m + |W| - 1] \implies 0 \;\leq\; i - m \;\leq\; |W| - 1 \implies \delta(i; W) \;=\; |W| - \big(i - m\big) + 1 \;>\; 0.$$

$$(W \cap W_p) \;\neq\; \varnothing \implies \sum_{i \in (W \cap W_p)} \delta(i; W) \;>\; 0, \quad \sum_{j \in W} \delta(j; W) \;>\; 0 \implies \frac{\sum_{i \in (W \cap W_p)} \delta(i; W)}{\sum_{j \in W} \delta(j; W)} \;>\; 0. \qquad \square$$

Proof.

Let $g\colon [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p}\colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 Consider the following example:

$$\begin{aligned} g &= 000111000111 \\ p &= 000111000000 \\ \hat{p} &= 000111000001 \end{aligned}$$

Then it holds $TPrec(g, p) = 1 = TPrec(g, \hat{p})$.

Prop. 2 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000100000000 \\ q &= 000110011000 \end{aligned}$$

Then it holds $TPrec(g, p) = 1 = TPrec(g, q)$.

Prop. 3 Consider the following example:

$$\begin{aligned} g &= 000000000000, \\ p &= 000001100000, \\ q &= 000011100000. \end{aligned}$$

$TPrec(g, p) = 0 = TPrec(g, q)$.

Prop. 4 Consider the following example:

$$\begin{aligned} g &= 000000000001, \\ p &= 000100000000, \\ q &= 000100100000. \end{aligned}$$

Then it follows $TPrec(g, p) = 0 = TPrec(g, q)$.

Prop. 5 Consider the following example:

$$\begin{aligned} g &= 000011110000, \\ p &= 110011000000, \\ q &= 001111000000. \end{aligned}$$

Then it holds $TPrec(g, p) = 0.5 > 0.3 = TPrec(g, q)$

Prop. 6 Consider the following example:

$$\begin{aligned} g &= 011111111110, \\ p &= 000000001011, \\ q &= 111111111011. \end{aligned}$$

Then it follows $TPrec(g,p) = 0.6667 < 0.7222 = TPrec(g,q)$.

Prop. 7 Consider the following example:

$$\begin{aligned} g &= 000111000000, \\ p &= 000110000000, \\ q &= 000100000000. \end{aligned}$$

Then it holds $TPrec(g,p) = 1.0 = TPrec(g,q)$.

Prop. 8 Consider the following example:

$$\begin{aligned} g &= 000111100000, \\ p &= 000110000000, \\ q &= 000011000000. \end{aligned}$$

Then it holds $TPrec(g,p) = 1.0 = TPrec(g,q)$.

Prop. 9 Consider the following example:

$$\begin{aligned} g &= 000111100000, \\ p &= 000100000000, \\ q &= 000000100000. \end{aligned}$$

Then it holds $TPrec(g,p) = 1.0 = TPrec(g,q)$. □

Theorem D.25. ***Time-Series Recall*** *defined as*

$$TRec(g,p) = \frac{1}{|A(g)|} \sum_{W \in A(g)} \gamma(|\{W_p \in A(p) \colon W \cap W_p \neq \emptyset\}|) \sum_{W_p \in A(p)} \frac{\sum_{i \in W_p \cap W} \delta(i - \min W, |W|)}{\sum_{i \in W} \delta(i - \min W, |W|)}$$

where $\delta \geq 1$*, and* γ *monotonically non-increasing with* $\gamma(k) = 1/k$ *and*

$$\gamma(1, W) = 1, \qquad \gamma(n, W) = \max_{0<m<n} \frac{\sum_{i \in W} \delta(i - \min W, |W|) - n + m}{\sum_{i \in W} \delta(i - \min W, |W|)} \gamma(m, W)$$

Proposition D.26 (Front-Bias $\implies$ Strictly Positive Coverage)**.**

$$i \in W = [m, m + |W| - 1] \implies 0 \leq i - m \leq |W| - 1 \implies \delta(i; W) = |W| - \big(i - m\big) + 1 > 0.$$

$$(W \cap W_p) \neq \varnothing \implies \sum_{i \in (W \cap W_p)} \delta(i; W) > 0, \quad \sum_{j \in W} \delta(j; W) > 0 \implies \frac{\sum_{i \in (W \cap W_p)} \delta(i; W)}{\sum_{j \in W} \delta(j; W)} > 0.$$ □

Proof.

Let $g \colon [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p} \colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 Let $W \in A(g)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $p_W = 0$, and $|\hat{p}_W^{-1}(1)| > 0$. Then $TRec(g,p)$ and $TRec(g,\hat{p})$ differ in one summand

$$0 = \gamma(|\{W_p \in A(p) \colon W \cap W_p \neq \emptyset\}|) \sum_{W_p \in A(p)} \frac{\sum\limits_{i \in W_p \cap W} \delta(i - \min W, |W|)}{\sum\limits_{i \in W} \delta(i - \min W, |W|)}$$

$$0 < \gamma(|\{W_{\hat{p}} \in A(\hat{p}) \colon W \cap W_{\hat{p}} \neq \emptyset\}|) \sum_{W_{\hat{p}} \in A(\hat{p})} \frac{\sum\limits_{i \in W_{\hat{p}} \cap W} \delta(i - \min W, |W|)}{\sum\limits_{i \in W} \delta(i - \min W, |W|)}$$

Thus $TRec(g,p) < TRec(g,\hat{p})$.

Prop. 2 Consider the following example:

$$g = 111111111111$$
$$p = 100000000000$$
$$\hat{p} = 110011111111$$

Then it holds $TRec_(g,p) = 0.1667 < 0.8846 = TRec(g,\hat{p})$.

Prop. 3 Consider the following example:

$$g = 000001000000$$
$$p = 000010000000$$
$$\hat{p} = 000110000000$$

Then it holds $TRec(g,p) = 0 = TRec(g,\hat{p})$.

Prop. 4 Consider the following example:

$$g = 000010000000$$
$$p = 000001000000$$
$$\hat{p} = 000101000000$$

Then it holds $TRec(g,p) = 0 = TRec(g,\hat{p})$.

Prop. 5 Let $W \in A(|g-1|)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $|\hat{p}_W^{-1}(1)| = |p_W^{-1}(1)|$, and $|A(\hat{p}_W)| = |A(p_W)|$. Let $W \in A(|g-1|)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $|\hat{p}_W^{-1}(1)| = |p_W^{-1}(1)|$, and $|A(\hat{p}_W)| = |A(p_W)|$. Then, since false positives do not affect the metric at all, $TRec(g,p) = TRec(g,\hat{p})$.

Prop. 6 Consider the following example:

$$g = 000010000000$$
$$p = 000001000000$$
$$\hat{p} = 000101000000$$

Then it holds $TRec(g,p) = 0 = TRec(g,q)$.

Prop. 7 Let $W \in A(g)$, $\hat{p}_{[n]\setminus\{i^*\}} = p_{[n]\setminus\{i^*\}}$ for some $i^* \in W$, $\hat{p}(i^*) \neq p(i^*) = 0$, and $|A(\hat{p}_W)| \leq |A(p_W)|$. Then

$$\begin{aligned}
&\gamma(|\{W_p \in A(p) \colon W \cap W_p \neq \emptyset\}|) \sum_{W_p \in A(p)} \frac{\sum\limits_{i \in W_p \cap W} \delta(i - \min W, |W|)}{\sum\limits_{i \in W} \delta(i - \min W, |W|)} \\
\leq &\gamma(|\{W_{\hat{p}} \in A(\hat{p}) \colon W \cap W_{\hat{p}} \neq \emptyset\}|) \sum_{W_p \in A(p)} \frac{\sum\limits_{i \in W_p \cap W} \delta(i - \min W, |W|)}{\sum\limits_{i \in W} \delta(i - \min W, |W|)} \\
< &\gamma(|\{W_{\hat{p}} \in A(\hat{p}) \colon W \cap W_{\hat{p}} \neq \emptyset\}|) \sum_{W_{\hat{p}} \in A(\hat{p})} \frac{\sum\limits_{i \in W_{\hat{p}} \cap W} \delta(i - \min W, |W|)}{\sum\limits_{i \in W} \delta(i - \min W, |W|)}
\end{aligned}$$

Thus $TRec(g, p) < TRec(g, \hat{p})$.

Prop. 8 Consider the following example:

$$\begin{aligned}
g &= 111111111111 \\
p &= 100000000001 \\
\hat{p} &= 010100000000
\end{aligned}$$

Then it holds $TRec(g, p) = 0.1923 < 0.2821 = TRec(g, \hat{p})$.

Prop. 9 Let $W \in A(g)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, Let $p, \hat{p}$ be two predictions with $|A(\hat{p}_W)| \leq |A(p_W)|$, let $i^* \in W$ with $p(i^*) = 1$, let $i^{**} < i^* \in W$ with $p(i^{**}) = 0$, and $p'_W(i) = \mathbb{1}_{\{i^{**}\}}(i) + p(i)\mathbb{1}_{W\setminus\{i^*, i^{**}\}}(i)$. Then

$$\begin{aligned}
&\gamma(|\{W_p \in A(p) \colon W \cap W_p \neq \emptyset\}|) \sum_{W_p \in A(p)} \frac{\sum\limits_{i \in W_p \cap W} \delta(i - \min W, |W|)}{\sum\limits_{i \in W} \delta(i - \min W, |W|)} \\
\leq &\gamma(|\{W_{\hat{p}} \in A(\hat{p}) \colon W \cap W_{\hat{p}} \neq \emptyset\}|) \sum_{W_p \in A(p)} \frac{\sum\limits_{i \in W_p \cap W} \delta(i - \min W, |W|)}{\sum\limits_{i \in W} \delta(i - \min W, |W|)}
\end{aligned}$$

Now

$$\begin{aligned}
&\sum_{W_p \in A(p)} \frac{\sum\limits_{i \in W_p \cap W} \delta(i - \min W, |W|)}{\sum\limits_{i \in W} \delta(i - \min W, |W|)} = \frac{\sum_{i \in I} \delta(i - minW, |W|)}{d} \\
< &\frac{\sum_{i \in I \setminus i^* \cup i^{**}} \delta(i - minW, |W|)}{d} = \sum_{W_{\hat{p}} \in A(\hat{p})} \frac{\sum\limits_{i \in W_{\hat{p}} \cap W} \delta(i - \min W, |W|)}{\sum\limits_{i \in W} \delta(i - \min W, |W|)}
\end{aligned}$$

for $d = \sum\limits_{i \in W} \delta(i - \min W, |W|)$ and some I. Thus

$$\begin{aligned}
&\gamma(|\{W_{\hat{p}} \in A(\hat{p}) \colon W \cap W_{\hat{p}} \neq \emptyset\}|) \sum_{W_p \in A(p)} \frac{\sum\limits_{i \in W_p \cap W} \delta(i - \min W, |W|)}{\sum\limits_{i \in W} \delta(i - \min W, |W|)} \\
< &\gamma(|\{W_{\hat{p}} \in A(\hat{p}) \colon W \cap W_{\hat{p}} \neq \emptyset\}|) \sum_{W_{\hat{p}} \in A(\hat{p})} \frac{\sum\limits_{i \in W_{\hat{p}} \cap W} \delta(i - \min W, |W|)}{\sum\limits_{i \in W} \delta(i - \min W, |W|)}
\end{aligned}$$

Thus finally $TRec(g, p) < TRec(g, \hat{p})$. $\square$

Theorem D.27. ***Time-Series F1-Score*** *defined as*

$$TF1(g,p) = 2\frac{TPrec(g,p)TRec(g,p)}{TPrec(g,p)+TRec(g,p)}$$

where

Proof.

Let $g\colon [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p}\colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 Let $W \in A(g)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $p_W = 0$, and $|\hat{p}_W^{-1}(1)| > 0$.

$$\begin{aligned} TPrec\big(g,\hat{p}\big) &\leq TPrec\big(g,p\big), \\ TRec\big(g,\hat{p}\big) &> TRec\big(g,p\big), \\ TF1\big(g,p\big) &= 2\,\frac{TPrec\big(g,p\big)\times TRec\big(g,p\big)}{TPrec\big(g,p\big)+TRec\big(g,p\big)}, \end{aligned}$$

from the above it follows $TF1\big(g,\hat{p}\big) > TF1\big(g,p\big)$.

Prop. 2 Consider the following example:

$$\begin{aligned} g &= 111111111111 \\ p &= 100000000000 \\ \hat{p} &= 110011111111 \end{aligned}$$

Then it holds $TF1(g,p) = 0.2667 < 0.5488 = TF1(g,\hat{p})$.

Prop. 3 Consider the following example:

$$\begin{aligned} g &= 000001000000 \\ p &= 000010000000 \\ \hat{p} &= 000110000000 \end{aligned}$$

Then it holds $TF1(g,p) = 0.0000 = TF1(g,\hat{p})$.

Prop. 4 Consider the following example:

$$\begin{aligned} g &= 000010000000 \\ p &= 000001000000 \\ \hat{p} &= 000101000000 \end{aligned}$$

Then it holds $TF1(g,p) = 0 = TF1(g,\hat{p})$.

Prop. 5 Consider the following example:

$$\begin{aligned} g &= 000001000000 \\ p &= 000101000000 \\ \hat{p} &= 000011000000 \end{aligned}$$

Then it holds $TF1(g,p) = 0.6667 > 0.5000 = TF1(g,\hat{p})$.

Prop. 6 Consider the following example:

$$\begin{aligned} g &= 000010000000 \\ p &= 000001000000 \\ \hat{p} &= 000101000000 \end{aligned}$$

Then it holds $TF1(g,p) = 0 = TF1(g,\hat{p})$.

Prop. 7 Let $W \in A(g)$, $\hat{p}_{[n]\setminus\{i^*\}} = p_{[n]\setminus\{i^*\}}$ for some $i^* \in W$, $\hat{p}(i^*) \neq p(i^*) = 0$, and $|A(\hat{p}_W)| \leq |A(p_W)|$.

$$\begin{aligned} TPrec\big(g,\hat{p}\big) &= TPrec\big(g,p\big), \\ TRec\big(g,\hat{p}\big) &> TRec\big(g,p\big), \\ TF1\big(g,p\big) &= 2\,\frac{TPrec\big(g,p\big) \times TRec\big(g,p\big)}{TPrec\big(g,p\big) + TRec\big(g,p\big)}, \end{aligned}$$

from the above it follows $TF1\big(g,\hat{p}\big) > TF1\big(g,p\big)$.

Prop. 8 Consider the following example:

$$\begin{aligned} g &= 111111111111 \\ p &= 100000000001 \\ \hat{p} &= 010100000000 \end{aligned}$$

Then it holds $TF1(g,p) = 0.1538 < 0.2273 = TF1(g,\hat{p})$.

Prop. 9

$$\begin{aligned} g &= 000000000011 \\ p &= 111111111110 \\ \hat{p} &= 111111111101 \end{aligned}$$

Then it holds $TF1(g,p) = 0.0296 < 0.4000 = TF1(g,\hat{p})$. □

Theorem D.28. ***NAB-Score*** *defined as*

$$NAB(g,p) = 100\frac{s_{NAB}(g,p) - s_{NAB}(g,0)}{s_{NAB}(g,g) - s_{NAB}(g,0)}$$

$$\begin{aligned} s_{NAB}(g,p) =& A_{FN}|\{W \in A(g)\colon W \cap p^{-1}(1) = \emptyset\}| + \sum_{\substack{W \in A(g)\colon \\ W \cap p^{-1}(1) \neq \emptyset}} A_{TP}\left(\frac{2}{1+e^{5(\min W \cap p^{-1}(1) - \max W)}} - 1\right) \\ &+ \sum_{\substack{i \in p^{-1}(1) \cap g^{-1}(0)\colon \\ i > \min g^{-1}(1)}} A_{FP}\left(\frac{2}{1+e^{5(\min\limits_{\substack{W \in A(g)\colon \\ \max W < i}} (i - \max W))}} - 1\right) - A_{FP}|\{i \in p^{-1}(1) \cap g^{-1}(0)\colon i < \min g^{-1}(1)\}| \end{aligned}$$

where $A_{TP} \geq 0$, $A_{FP} \geq 0$, $A_{FN} \leq 0$ *satisfies*

Proof.

For all properties it suffices to show the relation of s_{NAB}, as the normalization does not change the order. Let $g\colon [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p}\colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 Let $W \in A(g)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $p_W = 0$, and $|\hat{p}_W^{-1}(1)| > 0$. Then $s_{NAB}(g,p)$ and $s_{NAB}(g,\hat{p})$ only differ in the terms

$$\begin{aligned} \sum_{\substack{W_1 \in A(g)\colon \\ W_1 \cap \hat{p}^{-1}(1) \neq \emptyset}} A_{TP}\left(\frac{2}{1+e^{5(\min W_1 \cap \hat{p}^{-1}(1) - \max W_1)}} - 1\right) =& \sum_{\substack{W_1 \in A(g)\colon \\ W_1 \cap p^{-1}(1) \neq \emptyset}} A_{TP}\left(\frac{2}{1+e^{5(\min W_1 \cap p^{-1}(1) - \max W_1)}} - 1\right) \\ &+ \underbrace{A_{TP}\left(\frac{2}{1+e^{5(\min W \cap \hat{p}^{-1}(1) - \max W)}} - 1\right)}_{>0}. \end{aligned}$$

and

$$A_{FN}|\{W \in A(g)\colon W \cap \hat{p}^{-1}(1) = \emptyset\}| = A_{FN}|\{W \in A(g)\colon W \cap p^{-1}(1) = \emptyset\}| - A_{FN}$$

Thus, it follows

$$s_{NAB}(g,\hat{p}) - s_{NAB}(g,p) = A_{TP}\left(\frac{2}{1+e^{5(\min W \cap \hat{p}^{-1}(1) - \max W)}} - 1\right) - A_{FN} \geq 0$$

Where equality is achieved only, if $\hat{p}^{-1}(1) = \{\max W\}$ or $A_{TP} = 0$, and $A_{FN} = 0$. Thus, if $A_{TP} > 0$ and $A_{FN} < 0$, it holds $s_{NAB}(g,p) < s_{NAB}(g,\hat{p})$.

Prop. 2 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000111000000 \\ \hat{p} &= 000111011000 \end{aligned}$$

Then it holds $s_{NAB}(g,p) = A_{TP}\left(\frac{2}{1+e^{-30}} - 1\right) = s_{NAB}(g,\hat{p})$.

Prop. 3 Let $W \in A(|g-1|)$, $\hat{p}_{[n]\setminus\{i^*\}} = p_{[n]\setminus\{i^*\}}$ for some $i^* \in W$, $\hat{p}(i^*) \neq p(i^*) = 0$, and $|A(\hat{p}_W)| = |A(p_W)|$.
Case 1: $i^* < \min g^{-1}(1)$ Then $s_{NAB}(g,p)$ and $s_{NAB}(g,\hat{p})$ only differ in the term

$$|\{i \in \hat{p}^{-1}(1) \cap g^{-1}(0)\colon i < \min g^{-1}(1)\}| = |\{i \in p^{-1}(1) \cap g^{-1}(0)\colon i < \min g^{-1}(1)\}| + 1$$

Thus, it follows $s_{NAB}(g,p) > s_{NAB}(g,\hat{p})$, if $A_{FP} > 0$.
Case 2: $i^* > \min g^{-1}(1)$ Then $s_{NAB}(g,p)$ and $s_{NAB}(g,\hat{p})$ only differ in the third term by

$$\underbrace{\left(\frac{2}{1+e^{5(\min\limits_{\substack{W \in A(g):\\ \max W < i^*}} (i^* - \max W))}} - 1\right)}_{<0}.$$

Thus, it follows $s_{NAB}(g,p) > s_{NAB}(g,\hat{p})$, if $A_{FP} > 0$.

Prop. 4 Consider the following example:

$$\begin{aligned} g &= 000110000000 \\ p &= 000000000011 \\ \hat{p} &= 000001000010 \end{aligned}$$

Then it holds $s_{NAB}(g,p) = A_{FP}\left(\frac{2}{1+e^{30}} - 1 + \frac{2}{1+e^{35}} - 1\right) \leq A_{FP}\left(\frac{2}{1+e^{30}} - 1 + \frac{2}{1+e^{5}} - 1\right) = s_{NAB}(g,\hat{p})$.

Prop. 5 Consider the following example:

$$\begin{aligned} g &= 000110000000 \\ p &= 000000000001 \\ \hat{p} &= 000001000000 \end{aligned}$$

Then it holds $s_{NAB}(g,p) = A_{FP}\left(\frac{2}{1+e^{35}} - 1\right) < A_{FP}\left(\frac{2}{1+e^{5}} - 1\right) = s_{NAB}(g,\hat{p})$, if $A_{FP} > 0$.

Prop. 6 Let $W \in A(g)$, $\hat{W} \in A(|g-1|)$, $|A(\hat{p}_W)| = |A(p_W)|$, $\hat{p}_{[n]\setminus(W\cup\hat{W})} = p_{[n]\setminus(W\cup\hat{W})}$, $p_{\hat{W}} = 0$, and $\hat{p}_{\hat{W}} = \mathbb{1}_{\{i^*\}}$ for some $i^* \in \hat{W}$. Then $s_{NAB}(g,p)$ and $s_{NAB}(g,\hat{p})$ only differ in the terms

$$0 \leq \underbrace{\left(\frac{2}{1+e^{5(\min W \cap p^{-1}(1) - \max W)}} - 1\right)}_{x_p} \leq 1$$

and

$$-1 \leq \underbrace{\left(\frac{2}{1+e^{5(\min\limits_{\substack{W_1 \in A(g):\\ \max W_1 < i^*}} (i^* - \max W_1))}} - 1 \right) - \mathbb{1}(i^* < \min g^{-1}(1))}_{y_p} \leq \left(\frac{2}{1+e^5} - 1 \right) < 0$$

For the property to be satisfied, we need

$$\begin{aligned} A_{TP} x_{\hat{p}} + A_{FP} y_{\hat{p}} &< A_{TP} x_p \\ A_{TP}(x_{\hat{p}} - x_p) &< A_{FP} y_{\hat{p}} \end{aligned}$$

Since, $A_{TP} x_p \geq 0$ and $A_{FP} y_{\hat{p}} \leq 0$, this does not holds if $x_{\hat{p}} - x_p > 0$, i.e. $\min \hat{p}_W^{-1}(1) < \min p_W^{-1}(1)$.

Prop. 7 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000100000000 \\ \hat{p} &= 000111111000 \end{aligned}$$

Then it holds $s_{NAB}(g,p) = A_{TP}\left(\frac{2}{1+e^{-30}} - 1\right) = s_{NAB}(g,\hat{p})$.

Prop. 8 Let $W \in A(g)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $|A(\hat{p}_W)| = |A(p_W)|$, $|\hat{p}^{-1}(1)| = |p^{-1}(1)|$, and $\min\limits_{i \in W:\, p(i)=1} i < \min\limits_{i \in W:\, \hat{p}(i)=1} i$. Then it holds

$$\begin{aligned} s_{NAB}(g,p) - s_{NAB}(g,\hat{p}) &= A_{TP}\left(\frac{2}{1+e^{5(\min W \cap p^{-1}(1) - \max W)}} - 1\right) - A_{TP}\left(\frac{2}{1+e^{5(\min W \cap \hat{p}^{-1}(1) - \max W)}} - 1\right) \\ &= A_{TP}\left(\frac{2}{1+e^{5(\min W \cap p^{-1}(1) - \max W)}} - \frac{2}{1+e^{5(\min W \cap \hat{p}^{-1}(1) - \max W)}}\right) > 0. \end{aligned}$$

Thus, it holds $s_{NAB}(g,p) > s_{NAB}(g,\hat{p})$.

Prop. 9 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000100001000 \\ \hat{p} &= 000110000000 \end{aligned}$$

Then it holds $s_{NAB}(g,p) = A_{TP}\left(\frac{2}{1+e^{-30}} - 1\right) = s_{NAB}(g,\hat{p})$.

□

Theorem D.29. ***Time-Series Aware Precision*** *defined as*

$$TaP(g,p) = \alpha \frac{|\{W \in A(p) \colon \frac{1}{|W|} \sum\limits_{W_g \in A(g)} O(W_g, W) \geq \theta\}|}{|A(p)|} + (1-\alpha) \frac{1}{|A(p)|} \sum_{W \in A(p)} \frac{\sum\limits_{W_g \in A(g)} O(W_g, W)}{|W|}$$

where $\alpha \in [0,1]$, $\delta \geq 0$, $\theta > 0$, *and*

$$O(W, W_p) = |W \cap W_p| + \sum_{j \in [\max W + 1, \max W + \delta] \cap W_p} \frac{1}{1 + e^{\frac{12(j - 2 - \max W)}{\delta - 1} - 6}}$$

satisfies no properties.

Proof.

Let $g \colon [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p} \colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 Consider the example $g = \mathbb{1}_{[b,b+1]}$, $p = \mathbb{1}_{[a,b-1]} + \mathbb{1}_{\{b+2\}}$, and $\hat{p} = \mathbb{1}_{[a,b+2]}$ with $a < b$. Then, the second term for p is

$$(1-\alpha)\frac{1}{|A(p)|}\sum_{W\in A(p)}\frac{\sum_{W_g\in A(g)} O(W_g, W)}{|W|} = (1-\alpha)\frac{1}{2}\frac{1}{1+e^{\frac{-12}{\delta-1}-6}}$$

and for $\hat{p}$

$$(1-\alpha)\frac{1}{|A(\hat{p})|}\sum_{W\in A(\hat{p})}\frac{\sum_{W_g\in A(g)} O(W_g, W)}{|W|} = (1-\alpha)\frac{1}{|b+2-a|}\left(1+\frac{1}{1+e^{\frac{-12}{\delta-1}-6}}\right)$$

For all δ, we can select a small enough and b larger enough, such that

$$\frac{1}{2}\frac{1}{1+e^{\frac{-12}{\delta-1}-6}} > \frac{1}{|b+2-a|}\left(1+\frac{1}{1+e^{\frac{-12}{\delta-1}-6}}\right)$$

which implies the second term of TaP is smaller for $\hat{p}$ than for p. Additionally,

$$\frac{1}{1+e^{\frac{-12}{\delta-1}-6}} > \frac{1}{|b+2-a|}\left(1+\frac{1}{1+e^{\frac{-12}{\delta-1}-6}}\right)$$

implies that the first term of TaP for $\hat{p}$ can never be larger than for p. Thus, it holds $TaP(g,p) > TaP(g,\hat{p})$.

Prop. 2 Consider the following example:

$$g = 000111111000$$
$$p = 000110000000$$
$$\hat{p} = 000110011000$$

The first term is equal to α for predictions, $1 \geq \delta$, end 0 otherwise. The second term is equal to $(1-\alpha)$ for both predictions. Thus, it holds $TaP(g,p) = TaP(g,\hat{p})$.

Prop. 3 Consider the following example:

$$g = 000000111000$$
$$p = 000100000000$$
$$\hat{p} = 000111000000$$

Then it holds $TaP(g,p) = 0 = TaP(g,\hat{p})$.

Prop. 4 Consider the following example:

$$g = 000000111000$$
$$p = 000100000000$$
$$\hat{p} = 000101000000$$

Then it holds $TaP(g,p) = 0 = TaP(g,\hat{p})$.

Prop. 5 Consider the example $g = \mathbb{1}_{\{a\}}$, $p = \mathbb{1}_{\{a+1\}}$, and $\hat{p} = \mathbb{1}_{\{a+\delta+1\}}$. Then it holds $TaP(g,p) > \alpha\mathbb{1}(\theta = 0) = TaP(g,\hat{p})$.

Prop. 6 Consider the following example:

$$g = 000000111000$$
$$p = 000000000100$$
$$\hat{p} = 000100000100$$

Then it holds $TaP(g,p) = 0 = TaP(g,\hat{p})$.

Prop. 7 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000110000000 \\ \hat{p} &= 000111111000 \end{aligned}$$

Then it holds $TaP(g,p) = \alpha \mathbb{1}(1 \geq \theta) + (1-\alpha) = TaP(g,\hat{p})$.

Prop. 8 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000111000000 \\ \hat{p} &= 000000111000 \end{aligned}$$

Then it holds $TaP(g,p) = \alpha \mathbb{1}(1 \geq \theta) + (1-\alpha) = TaP(g,\hat{p})$.

Prop. 9 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000000111000 \\ \hat{p} &= 000001110000 \end{aligned}$$

Then it holds $TaP(g,p) = \alpha \mathbb{1}(1 \geq \theta) + (1-\alpha) = TaP(g,\hat{p})$. □

Theorem D.30. ***Time-Series Aware Recall*** *defined as*

$$TaR(g,p) = \alpha \frac{|\{W \in A(g) \colon \frac{1}{|W|} \sum_{W_p \in A(p)} O(W, W_p) \geq \theta\}|}{|A(g)|} + (1-\alpha) \frac{1}{|A(g)|} \sum_{W \in A(g)} \min\left\{1, \frac{\sum_{W_p \in A(p)} O(W, W_p)}{|W|}\right\}$$

where $\alpha \in [0,1]$, $\delta \geq 0$, $\theta > 0$, *and*

$$O(W, W_p) = |W \cap W_p| + \sum_{j \in [\max W + 1, \max W + \delta] \cap W_p} \frac{1}{1 + e^{-6 \frac{12(j - 2 - \max W)}{\delta - 1}}}$$

satisfies Prop. 1 and Prop. 7 for many but not all parameter combinations.

Proof.

Let $\mathrm{g} \colon [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p} \colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 Let $W \in A(g)$, $\hat{p}_{[n] \setminus W} = p_{[n] \setminus W}$, $p_W = 0$, and $|\hat{p}_W^{-1}(1)| > 0$.

In general, it holds

$$|\{W \in A(g) \colon \frac{1}{|W|} \sum_{W_p \in A(\hat{p})} O(W, W_p) \geq \theta\}| \geq |\{W \in A(g) \colon \frac{1}{|W|} \sum_{W_p \in A(p)} O(W, W_p) \geq \theta\}|$$

and

$$\sum_{W \in A(g)} \min\left\{1, \frac{\sum_{W_p \in A(\hat{p})} O(W, W_p)}{|W|}\right\} \geq \sum_{W \in A(g)} \min\left\{1, \frac{\sum_{W_p \in A(p)} O(W, W_p)}{|W|}\right\},$$

If the first term depending on θ does not change, we can achieve equality in the second term for some parameterizations, e.g. $\delta = 1$.

Prop. 2 Let $W \in A(g)$, and let $p_W(i) = 0$ for all $i \in I \subset W$, $|A(\hat{p}_W)| = |A(p_W)| + 1$, $\min_{i \in I} i > \max_{j \in w\colon p(j)=1} j$, and $\hat{p}(i) = \mathbb{1}_I(i) + p(i)\mathbb{1}_{[n]\setminus I}(i)$.

Prop. 3 Consider the following example:

$$\begin{aligned} g &= 000000111000 \\ p &= 000100000000 \\ \hat{p} &= 000111000000 \end{aligned}$$

Then it holds $TaR(g, p) = 0 = TaR(g, \hat{p})$.

Prop. 4 Consider the following example:

$$\begin{aligned} g &= 000000111000 \\ p &= 000100000000 \\ \hat{p} &= 000101000000 \end{aligned}$$

Then it holds $TaR(g, p) = 0 = TaR(g, \hat{p})$.

Prop. 5 Consider the example $g = \mathbb{1}_{\{a\}}$, $p = \mathbb{1}_{\{a+1\}}$, and $\hat{p} = \mathbb{1}_{\{a+\delta+1\}}$. Then it holds $TaR(g, p) > \alpha\mathbb{1}(\theta = 0) = TaR(g, \hat{p})$.

Prop. 6 Consider the following example:

$$\begin{aligned} g &= 000000111000 \\ p &= 000000000000 \\ \hat{p} &= 000100000000 \end{aligned}$$

Then it holds $TaR(g, p) = 0 = TaR(g, \hat{p})$.

Prop. 7 Let $W \in A(g)$, $\hat{p}_{[n]\setminus\{i^*\}} = p_{[n]\setminus\{i^*\}}$ for some $i^* \in W$, $\hat{p}(i^*) \neq p(i^*) = 0$, and $|A(\hat{p}_W)| \leq |A(p_W)|$. In general, it holds

$$|\{W \in A(g)\colon \frac{1}{|W|} \sum_{W_p \in A(\hat{p})} O(W, W_p) \geq \theta\}| \geq |\{W \in A(g)\colon \frac{1}{|W|} \sum_{W_p \in A(p)} O(W, W_p) \geq \theta\}|$$

and

$$\sum_{W \in A(g)} \min\left\{1, \frac{\sum_{W_p \in A(\hat{p})} O(W, W_p)}{|W|}\right\} \geq \sum_{W \in A(g)} \min\left\{1, \frac{\sum_{W_p \in A(p)} O(W, W_p)}{|W|}\right\},$$

where equality holds only, if

$$\sum_{j \in [\max W+1, \max W+\delta] \cap p^{-1}(1)} \frac{1}{1 + e^{-6\frac{12(j-2-\max W)}{\delta-1}}} \geq |(g \cdot p)_W^{-1}(0)|$$

There are parameterizations for which this is true, e.g. $\delta = 1$, but also parameterizations for which it is not true, e.g. $\delta = 0$.

Prop. 8 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000011000000 \\ \hat{p} &= 000110000000 \end{aligned}$$

Then it holds $TaR(g, p) = TaR(g, \hat{p})$.

Prop. 9 Consider the following example:

$$g = 000111111000$$
$$p = 000010100000$$
$$\hat{p} = 000110000000$$

Then it holds $TaR(g, p) = TaR(g, \hat{p})$.

□

Theorem D.31. ***Time-Tolerant Precision***

$$P_\delta(g, p) = \frac{|\{i \in p^{-1}(1) \colon [i - \delta, i + \delta] \cap g^{-1}(1) \neq \emptyset\}|}{|p^{-1}(1)|}$$

where $\delta \in \mathbb{Z}_{\geq 0}$ satisfies property Prop. 5 for $\delta = 0$.

Proof.

Let $\mathrm{g} \colon [n] \to \{0, 1\}^n$ be the ground truth and $p, \hat{p} \colon [n] \to \{0, 1\}^n$ predictions.

Prop. 1 Consider the following example:

$$g = 000110011000$$
$$p = 000110000000$$
$$\hat{p} = 000110001000$$

Then $\forall \delta : P_\delta(g, p) = 1 = P_\delta(g, \hat{p})$.

Prop. 2 Consider the following example:

$$g = 000111111000$$
$$p = 000111000000$$
$$\hat{p} = 000111011000$$

Then $\forall \delta : P_\delta(g, p) = 1 = P_\delta(g, \hat{p})$.

Prop. 3 Consider the following example:

$$g = 000000000000$$
$$p = 100000000000$$
$$\hat{p} = 110000000000$$

Then it holds $P_\delta(g, p) = 0 = P_\delta(g, \hat{p})$.

Prop. 4 Consider the following example:

$$g = 000000000000$$
$$p = 000110000000$$
$$\hat{p} = 000101000000$$

Then it holds $P_\delta(g, p) = 0 = P_\delta(g, \hat{p})$.

Prop. 5 **Case $\delta = 0$.** Then $P_\delta(g,p) = Precision(g,p)$ and this property is fulfilled following Thm. D.1.

Case $\delta > 0$. Consider the following example: $g = \mathbb{1}_{[a,a+1]}$ with $\delta + 1 < a < n$, $p = \mathbb{1}_{\{a-\delta,a+1\}}$, and $\hat{p} = \mathbb{1}_{\{a-\delta-1,a+1\}}$. Then it holds $P_\delta(g,p) = 1 > \frac{1}{2} = P_\delta(g,\hat{p})$.

For example, let $\delta = 1$:

$$\begin{aligned} g &= 000000110000 \\ p &= 000001010000 \\ \hat{p} &= 000010010000 \end{aligned}$$

Then it holds $P_\delta(g,p) = 1 > \frac{1}{2} = P_\delta(g,\hat{p})$.

Prop. 6 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000010000100 \\ \hat{p} &= 001011110100 \end{aligned}$$

- For $\delta = 0$: $P_\delta(g,p) = \frac{1}{2} < \frac{2}{3} = P_\delta(g,\hat{p})$
- For $\delta > 0$: $P_\delta(g,p) = 1 = P_\delta(g,\hat{p})$

Regardless of the choice of δ this property does not hold.

Prop. 7 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000010000000 \\ \hat{p} &= 000111111000 \end{aligned}$$

Then $\forall\delta : P_\delta(g,p) = 1 = P_\delta(g,\hat{p})$.

Prop. 8 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000010000000 \\ \hat{p} &= 000000010000 \end{aligned}$$

Then $\forall\delta : P_\delta(g,p) = 1 = P_\delta(g,\hat{p})$.

Prop. 9 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000010010000 \\ \hat{p} &= 000110000000 \end{aligned}$$

Then $\forall\delta : P_\delta(g,p) = 1 = P_\delta(g,\hat{p})$. □

Theorem D.32. ***Time-Tolerant Recall***

$$R_\delta(g,p) = \frac{|\{i \in g^{-1}(1) \colon [i-\delta, i+\delta] \cap p^{-1}(1) \neq \emptyset\}|}{|g^{-1}(1)|}$$

where $\delta \in \mathbb{Z}_{\geq 0}$ satisfies properties Prop. 1, Prop. 5, and Prop. 7 only for $\delta = 0$.

Proof.

Let $\mathrm{g} \colon [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p} \colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 **Case** $\delta = 0$**.** Then $R_\delta(g,p) = \mathit{Recall}(g,p)$ and this property is fulfilled following Thm. D.2.

Case $\delta > 0$**.** Consider the following example:

$$\begin{aligned} g &= 000000100000 \\ p &= 000001100000 \\ \hat{p} &= 000001000000 \end{aligned}$$

Then it holds $R_\delta(g,p) = 1 = R_\delta(g,\hat{p})$.

Prop. 2 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000110000000 \\ \hat{p} &= 000110011000 \end{aligned}$$

Then $\forall\delta : R_\delta(g,p) = \frac{1}{3} < \frac{2}{3} = R_\delta(g,\hat{p})$.

Prop. 3 Consider the following example:

$$\begin{aligned} g &= 000000111000 \\ p &= 000010000000 \\ \hat{p} &= 000011000000 \end{aligned}$$

Then it holds

- For $\delta = 0$: $R_\delta(g,p) = 0 = R_\delta(g,\hat{p})$
- For $\delta = 1$: $R_\delta(g,p) = 0 < \frac{1}{3} = R_\delta(g,\hat{p})$
- For $\delta > 1$: $R_\delta(g,p) = \frac{1}{3} < \frac{2}{3} = R_\delta(g,\hat{p})$

Regardless of the choice of δ this property does not hold.

Prop. 4 Consider the following example:

$$\begin{aligned} g &= 000000111000 \\ p &= 000110010000 \\ \hat{p} &= 000101010000 \end{aligned}$$

Then it holds

- For $\delta = 0$: $R_\delta(g,p) = \frac{1}{3} = R_\delta(g,\hat{p})$
- For $\delta > 0$: $R_\delta(g,p) = 1 = R_\delta(g,\hat{p})$

Regardless of the choice of δ this property does not hold.

Prop. 5 **Case** $\delta = 0$**.** Then $R_\delta(g,p) = \mathit{Recall}(g,p)$ and this property is fulfilled following Thm. D.2.

Case $\delta > 0$**.** Consider the following example: $g = \mathbb{1}_{[a,a+\delta+1]}$ with $\delta + 1 < a < a + \delta < n$, $p = \mathbb{1}_{\{a-\delta,a+\delta+1\}}$, and $\hat{p} = \mathbb{1}_{\{a-\delta-1,a+\delta+1\}}$. Then it holds $R_\delta(g,p) = 1 > \frac{\delta+1}{\delta+2} = R_\delta(g,\hat{p})$.

For example, let $\delta = 1$:

$$\begin{aligned} g &= 000000111000 \\ p &= 000001001000 \\ \hat{p} &= 000010001000 \end{aligned}$$

Then it holds $R_\delta(g,p) = 1 > \frac{\delta+1}{\delta+2} = \frac{2}{3} = R_\delta(g,\hat{p})$.

Prop. 6 Consider the following example:

$$g = 000000111000$$
$$p = 000000010000$$
$$\hat{p} = 000100110000$$

Then it holds

- For $\delta = 0$: $R_\delta(g, p) = \frac{1}{3} < \frac{2}{3} = R_\delta(g, \hat{p})$
- For $\delta > 0$: $R_\delta(g, p) = 1 = R_\delta(g, \hat{p})$

Regardless of the choice of δ this property does not hold.

Prop. 7 **Case $\delta = 0$.** Then $R_\delta(g, p) = Recall(g, p)$ and this property is fulfilled following Thm. D.2.

Case $\delta > 0$. Consider the following example: $g = \mathbb{1}_{[0,\delta+1]}$, $p = \mathbb{1}_{\{0\}} + \mathbb{1}_{\{\delta+2\}}$, and $p = \mathbb{1}_{[0,1]} + \mathbb{1}_{\{\delta+2\}}$. Then it holds $R_\delta(g, p) = 1 = R_\delta(g, \hat{p})$.

For example for $\delta = 1$:

$$g = 111000000000$$
$$p = 100100000000$$
$$\hat{p} = 110100000000$$

Then it follows $R_\delta(g, p) = 1 = R_\delta(g, \hat{p})$

Prop. 8 Consider the following example:

$$g = 000001100000$$
$$p = 000001000000$$
$$\hat{p} = 000000100000$$

Then it holds

- For $\delta = 0$: $R_\delta(g, p) = \frac{1}{2} = R_\delta(g, \hat{p})$
- For $\delta > 0$: $R_\delta(g, p) = 1 = R_\delta(g, \hat{p})$

Regardless of the choice of δ this property does not hold.

Prop. 9 Consider the following example:

$$g = 000011110000$$
$$p = 000011000000$$
$$\hat{p} = 000001010000$$

Then it holds

- For $\delta = 0$: $R_\delta(g, p) = \frac{1}{2} = R_\delta(g, \hat{p})$
- For $\delta = 1$: $R_\delta(g, p) = \frac{3}{4} < 1 = R_\delta(g, \hat{p})$
- For $\delta > 1$: $R_\delta(g, p) = 1 = R_\delta(g, \hat{p})$

Regardless of the choice of δ this property does not hold. □

Theorem D.33. ***Average Precision Probability*** *defined as*

$$P_{Precision}(g,p) = \frac{\sum_{\substack{W\in A(g):\\ I_W\cap p^{-1}(1)\neq\emptyset}} \frac{1}{|I_W\cap p^{-1}(1)|} \sum_{i\in I_W\cap p^{-1}(1)} \left(1 - \frac{|W|+\min\{\min_{j\in W}|i-j|, \min\{\min W-\min I_W, \max I_W-\max W\}\}+\min_{j\in W}|i-j|}{|I_W|}\right)}{|\{W\in A(g)\colon I_W\cap p^{-1}(1)\neq\emptyset\}|}$$

where $I_W = \{i\in[n]\colon \min_{j\in W}|i-\epsilon-j| < \min_{\hat{W}\in A(g)\setminus\{W\}}\min_{j\in\hat{W}}|i-\epsilon-j|\}$ *satisfies none of the properties.*

Note: We introduce $\epsilon\in(-\frac{1}{2},\frac{1}{2})$ *here to potentially break ties. The original paper (Huet et al., 2022) considers the case* $\epsilon=0$*, but does not specify how ties are broken. Furthermore, the paper does not specify the case* $|\{W\in A(g)\colon I_W\cap p^{-1}(1)\neq\emptyset\}|=0$*. This can happen, if the ground truth or the prediction does not contain a single anomaly window. In that case, the metric as stated is undefined. Thus, we assume* $|g^{-1}(1)|>0$ *and* $|p^{-1}(1)|>0$*.*

Proof.

Let g$\colon[n]\to\{0,1\}^n$ be the ground truth and $p,\hat{p}\colon[n]\to\{0,1\}^n$ predictions.

Prop. 1 Consider the following example:

$$g = 010000001000$$
$$p = 000000001000$$
$$\hat{p} = 010000001000$$

Then it holds, $P_{Precision}(g,p)=\frac{30}{35}>\frac{29}{35}=P_{Precision}(g,\hat{p})$.

Prop. 2 Consider the following example:

$$g = 000111111000$$
$$p = 000111000000$$
$$\hat{p} = 000111011000$$

Then it holds, $P_{Precision}(g,p)=\frac{1}{2}=P_{Precision}(g,\hat{p})$.

Prop. 3 Consider the following example:

$$g = 010000100000$$
$$p = 001100000000$$
$$\hat{p} = 001110000000$$

Then it holds, $P_{Precision}(g,p)=\frac{1}{8}<\frac{1}{4}=P_{Precision}(g,\hat{p})$.

Prop. 4 Consider the following example:

$$g = 010000001000$$
$$p = 000011101000$$
$$\hat{p} = 000010101000$$

Then it holds, $P_{Precision}(g,p)=\frac{3}{35}<\frac{3}{28}=P_{Precision}(g,\hat{p})$.

Prop. 5 Consider the following example:

$$g = 000000111000$$
$$p = 010000100000$$
$$\hat{p} = 000010100000$$

Then it holds, $P_{Precision}(g,p)=\frac{5}{12}<\frac{7}{12}=P_{Precision}(g,\hat{p})$.

Prop. 6 Consider the following example:

$$g = 000111111000$$
$$p = 100000001000$$
$$\hat{p} = 100111100010$$

Then it holds, $P_{Precision}(g, p) = \frac{1}{4} < \frac{13}{36} = P_{Precision}(g, \hat{p})$.

Prop. 7 Consider the following example:

$$g = 000111111000$$
$$p = 000011111000$$
$$\hat{p} = 000111111000$$

Then it holds, $P_{Precision}(g, p) = \frac{1}{2} = P_{Precision}(g, \hat{p})$.

Prop. 8 Consider the following example:

$$g = 000111111000$$
$$p = 000010000000$$
$$\hat{p} = 000000010000$$

Then it holds, $P_{Precision}(g, p) = \frac{1}{2} = P_{Precision}(g, \hat{p})$.

Prop. 9 Consider the following example:

$$g = 000111111000$$
$$p = 000010010000$$
$$\hat{p} = 000110000000$$

Then it holds, $P_{Precision}(g, p) = \frac{1}{2} = P_{Precision}(g, \hat{p})$. □

Theorem D.34. ***Average Recall Probability*** *defined as*

$$P_{Recall}(g, p) = \frac{1}{|A(g)|} \sum_{W \in A(g)} \frac{1}{|W|} \sum_{i \in W} \left(1 - \frac{\min\{d(i, I_W \cap p^{-1}(1)), \min\{i - \min I_W, \max I_W - i\}\} + d(i, I_W \cap p^{-1}(1))}{|I_W|}\right)$$

where $I_W = \{i \in [n] \colon \min_{j \in W} |i - \epsilon - j| < \min_{\hat{W} \in A(g) \setminus \{W\}} \min_{j \in \hat{W}} |i - \epsilon - j|\}$ *and*

$$d(i, I \cap p^{-1}(1)) = \begin{cases} \min_{j \in I \cap p^{-1}(1)} |i - j| & , I \cap p^{-1}(1) \neq \emptyset \\ \infty & , I \cap p^{-1}(1) = \emptyset \end{cases}$$

satisfies none of the properties.
Note: We introduce $\epsilon \in (-\frac{1}{2}, \frac{1}{2})$ *here to potentially break ties. The original paper (Huet et al., 2022) considers the case* $\epsilon = 0$*, but does not specify how ties are broken.*

Proof.

Let $g \colon [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p} \colon [n] \to \{0,1\}^n$ predictions.

Consider the following example:

$$g = 000110011000$$
$$p = xxxxxx000000$$

The term in the inner sum of P_{Recall} for the second window is equal to $-\infty$ since $I_W \cap p^{-1}(1) = \emptyset$ for any predictions that differ only in the x's. Thus, none of the properties except for Prop. 5 can hold for any two predictions that do not predict one of the ground truth alarms.

Prop. 5 Consider the following example:

$$g = 000000010000$$
$$p = 000100000000$$
$$\hat{p} = 000010000000$$

Then it holds $P_{Recall}(g,p) = \frac{1}{3} \neq \frac{1}{2} = P_{Recall}(g,\hat{p})$. □

Theorem D.35. ***Affiliation F1-Score*** *defined as*

$$F1_{affiliation}(g,p) = 2\frac{P_{Precision} \cdot P_{Recall}}{P_{Precision} + P_{Recall}}$$

satisfies none of the properties.
Note: $P_{Precision}$ is not defined, when either the ground truth or the prediction are zero. Additionally, P_{Recall} equals $-\infty$ if at least one ground truth alarm does not have any affiliated predictions. Thus, we assume $|g^{-1}(1)| > 0$, $|p^{-1}(1)| > 0$, and $I_W \cap p^{-1}(1) \neq \emptyset$ for all $W \in A(g)$.

Proof.

Let g: $[n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p}$: $[n] \to \{0,1\}^n$ predictions. Consider Thm. D.34. Similarly here, we can construct counterexamples for all properties since recall is $-\infty$. Affiliation F1-Score is undefined until there is at least one prediction in each window. Thus, none of the properties except for Prop. 5 can hold for any two predictions that do not predict one of the ground truth alarms.

Prop. 5 Consider the following example:

$$g = 000000111000$$
$$p = 010000100000$$
$$\hat{p} = 000010100000$$

$$P_{\text{Precision}}(g,p) = \frac{5}{12} \quad P_{\text{Precision}}(g,\hat{p}) = \frac{7}{12}$$
$$P_{\text{Recall}}(g,p) = \frac{1}{6} \quad P_{\text{Recall}}(g,\hat{p}) = \frac{19}{36}$$

Then it holds, $F1_{\text{affiliation}}(g,p) = \frac{5}{21} < \frac{133}{240} = F1_{\text{affiliation}}(g,\hat{p})$. □

Theorem D.36. ***Enhanced Time-Series Aware Precision*** *defined as*

$$eTaP(g,p) = \frac{1}{2}\sum_{W \in P}\left(1 + \sum_{W_g \in A}\frac{|W \cap W_g|}{|W|}\right)\frac{\sqrt{|W|}}{\sum_{W_p \in A(p)}\sqrt{|W_p|}}$$

where $\theta_r > 0$, $\theta_p > 0$, and

$$A = \{W \in A(g) \colon \sum_{W_p \in P}\frac{|W \cap W_p|}{|W|} \geq \theta_r\}$$
$$P = \{W \in A(p) \colon \sum_{W_g \in A}\frac{|W \cap W_g|}{|W|} \geq \theta_p\}$$

satisfies Prop. 5.

Proof.

Let g: $[n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p}$: $[n] \to \{0,1\}^n$ predictions.

Prop. 1 Consider the following example: Let $\mathbb{N} \ni a > 1/\theta_r$, $g = \mathbb{1}_{[1,a]}$, $p = 0$, and $\hat{p} = \mathbb{1}_{\{1\}}$. Then it holds $eTaP(g,p) = 0 = eTaP(g,\hat{p})$.

Prop. 2 Consider the following example: Let $\mathbb{N} \ni a > 2/\theta_r$, $g = \mathbb{1}_{[1,a]}$, $p = \mathbb{1}_{\{1\}}$, and $\hat{p} = \mathbb{1}_{\{1,a\}}$. Then it holds $eTaP(g,p) = 0 = eTaP(g,\hat{p})$.

Prop. 3 Consider the following example:

$$\begin{aligned} g &= 000000000000 \\ p &= 010000000000 \\ \hat{p} &= 011000000000 \end{aligned}$$

Then it holds $eTaP(g,p) = 0 = eTaP(g,\hat{p})$.

Prop. 4 Consider the following example:

$$\begin{aligned} g &= 000000000000 \\ p &= 010000000000 \\ \hat{p} &= 010100000000 \end{aligned}$$

Then it holds $eTaP(g,p) = 0 = eTaP(g,\hat{p})$.

Prop. 5 Let $W \in A(|g-1|)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $|\hat{p}_W^{-1}(1)| = |p_W^{-1}(1)|$, and $|A(\hat{p}_W)| = |A(p_W)|$. Since $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, then $A(\hat{p}) = A(p)$, and hence $eTaP(g,\hat{p}) = eTaP(g,p)$.

Prop. 6 Consider the following example:

$$\begin{aligned} g &= 010000000000 \\ p &= 100000000000 \\ \hat{p} &= 101000000000 \end{aligned}$$

Then it holds $eTaP(g,p) = 0 = eTaP(g,\hat{p})$.

Prop. 7 Consider the following example: Let $\mathbb{N} \ni a > 2/\theta_r$, $g = \mathbb{1}_{[1,a]}$, $p = \mathbb{1}_{\{1\}}$, and $\hat{p} = \mathbb{1}_{\{1,2\}}$. Then it holds $eTaP(g,p) = 0 = eTaP(g,\hat{p})$.

Prop. 8 Consider the following example: Let $\mathbb{N} \ni a > 2/\theta_r$, $g = \mathbb{1}_{[1,a]}$, $p = \mathbb{1}_{\{1\}}$, and $\hat{p} = \mathbb{1}_{\{2\}}$. Then it holds $eTaP(g,p) = 0 = eTaP(g,\hat{p})$.

Prop. 9 Consider the following example: Let $\mathbb{N} \ni a > 2/\theta_r$, $g = \mathbb{1}_{[1,a]}$, $p = \mathbb{1}_{\{1\}}$, and $\hat{p} = \mathbb{1}_{\{2\}}$. Then it holds $eTaP(g,p) = 0 = eTaP(g,\hat{p})$. □

Theorem D.37. ***Enhanced Time-Series Aware Recall*** *defined as*

$$eTaR(g,p) = \frac{1}{2|A(g)|} \sum_{W \in A} \left(1 + \sum_{W_p \in P} \frac{|W \cap W_p|}{|W|} \right)$$

where $\theta_r > 0$, $\theta_p > 0$, *and*

$$\begin{aligned} A &= \{W \in A(g) \colon \sum_{W_p \in P} \frac{|W \cap W_p|}{|W|} \geq \theta_r\} \\ P &= \{W \in A(p) \colon \sum_{W_g \in A} \frac{|W \cap W_g|}{|W|} \geq \theta_p\} \end{aligned}$$

satisfies Prop. 5.

Proof.

Let $g \colon [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p} \colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 Consider the following example: Let $\mathbb{N} \ni a > 1/\theta_r$, $g = \mathbb{1}_{[1,a]}$, $p = 0$, and $\hat{p} = \mathbb{1}_{\{1\}}$. Then it holds $eTaR(g,p) = 0 = eTaR(g,\hat{p})$.

Prop. 2 Consider the following example: Let $\mathbb{N} \ni a > 2/\theta_r$, $g = \mathbb{1}_{[1,a]}$, $p = \mathbb{1}_{\{1\}}$, and $\hat{p} = \mathbb{1}_{\{1,a\}}$. Then it holds $eTaR(g,p) = 0 = eTaR(g,\hat{p})$.

Prop. 3 Consider the following example:

$$\begin{aligned} g &= 100000000000 \\ p &= 010000000000 \\ \hat{p} &= 011000000000 \end{aligned}$$

Then it holds $eTaR(g,p) = 0 = eTaR(g,\hat{p})$.

Prop. 4 Consider the following example:

$$\begin{aligned} g &= 100000000000 \\ p &= 010000000000 \\ \hat{p} &= 010100000000 \end{aligned}$$

Then it holds $eTaR(g,p) = 0 = eTaR(g,\hat{p})$.

Prop. 5 Let $W \in A(|g-1|)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $|\hat{p}_W^{-1}(1)| = |p_W^{-1}(1)|$, and $|A(\hat{p}_W)| = |A(p_W)|$. Since $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, then $A(\hat{p}) = A(p)$, and hence $eTaR(g,\hat{p}) = eTaR(g,p)$.

Prop. 6 Consider the following example:

$$\begin{aligned} g &= 100000000000 \\ p &= 100000000000 \\ \hat{p} &= 101000000000 \end{aligned}$$

Then it holds $eTaR(g,p) = 1 = eTaR(g,\hat{p})$.

Prop. 7 Consider the following example: Let $\mathbb{N} \ni a > 2/\theta_r$, $g = \mathbb{1}_{[1,a]}$, $p = \mathbb{1}_{\{1\}}$, and $\hat{p} = \mathbb{1}_{\{1,2\}}$. Then it holds $eTaR(g,p) = 0 = eTaR(g,\hat{p})$.

Prop. 8 Consider the following example: Let $\mathbb{N} \ni a > 2/\theta_r$, $g = \mathbb{1}_{[1,a]}$, $p = \mathbb{1}_{\{1\}}$, and $\hat{p} = \mathbb{1}_{\{2\}}$. Then it holds $eTaR(g,p) = 0 = eTaR(g,\hat{p})$.

Prop. 9 Consider the following example: Let $\mathbb{N} \ni a > 2/\theta_r$, $g = \mathbb{1}_{[1,a]}$, $p = \mathbb{1}_{\{1\}}$, and $\hat{p} = \mathbb{1}_{\{2\}}$. Then it holds $eTaR(g,p) = 0 = eTaR(g,\hat{p})$. □

Theorem D.38. ***Temporal Distance*** *is defined as*

$$TD(g,p) = f_{closest}(g,p) + f_{closest}(p,g)$$

$$f_{closest}(a,b) = \sum_{i \in a^{-1}(1)} \min_{j \in b^{-1}(1)} |i-j|.$$

A Temporal Distance-based metric $m_{TD}(g,p) = f(TD(g,p))$, where f is a strictly decreasing function, satisfies properties Prop. 1, Prop. 7.

Proof.

Let $\mathrm{g}\colon [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p}\colon [n] \to \{0,1\}^n$ predictions.

Lemma D.39. *Let $W \in A(g)$, $\hat{p}_{[n]\setminus\{i^*\}} = p_{[n]\setminus\{i^*\}}$ for some $i^* \in W$, and $\hat{p}(i^*) \neq p(i^*) = 0$. Then $m_{TD}(g,\hat{p}) > m_{TD}(g,p)$.*

Proof. Consider the two $f_{closest}$ terms of TD separately:

1. Since $i^* \in \hat{p}^{-1}(1)$, we have $\min_{j \in \hat{p}^{-1}(1)} |i^* - j| = 0$. For any $i' \in g^{-1}(1) \setminus \{i^*\}$ if $|i' - i^*| < \min_{j \in p^{-1}(1)} |i' - j|$, then
$$\min_{j \in \hat{p}^{-1}(1)} |i' - j| = |i' - i^*| < \min_{j \in p^{-1}(1)} |i' - j|.$$
Thus:
$$f_{\text{closest}}(g, \hat{p}) < f_{\text{closest}}(g, p)$$

2. Note, that for $I \subset [n]$, the following holds:
$$f_{\text{closest}}(a, b) = f_{\text{closest}}(a_I, b) + f_{\text{closest}}(a_{[n]\setminus I}, b).$$
Since $i^* \in g^{-1}(1)$, we have $\min_{j \in g^{-1}(1)} |i^* - j| = 0$, thus:
$$f_{\text{closest}}(\hat{p}_W, g) = f_{\text{closest}}(p_W, g) = 0 \implies f_{\text{closest}}(p, g) = f_{\text{closest}}(\hat{p}, g)$$

Then:
$$TD(g, \hat{p}) < TD(g, p) \implies m_{TD}(g, \hat{p}) > m_{TD}(g, p).$$

□

Prop. 1 Let $W \in A(g)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $p_W = 0$, and $|\hat{p}_W^{-1}(1)| > 0$. By Thm. D.39 , it directly follows that $m_{TD}(g, \hat{p}) > m_{TD}(g, p)$.

Prop. 2 Consider the following example:

$$g = 000111111000$$
$$p = 000111000000$$
$$\hat{p} = 000111011000$$

Then it holds $TD(g, p) = 6 > 1 = TD(g, \hat{p}) \implies m_{TD}(g, p) < m_{TD}(g, \hat{p})$.

Prop. 3 Consider the following example:

$$g = 000000011000$$
$$p = 000010000000$$
$$\hat{p} = 000011000000$$

Then it holds $TD(g, p) = 10 = TD(g, \hat{p}) \implies m_{TD}(g, p) = m_{TD}(g, \hat{p})$.

Prop. 4 Consider the following example:

$$g = 000000111000$$
$$p = 001110110000$$
$$\hat{p} = 001010110000$$

Then it holds $TD(g, p) = 10 > 7 = TD(g, \hat{p}) \implies m_{TD}(g, p) < m_{TD}(g, \hat{p})$.

Prop. 5 Consider the following example:

$$g = 000000011110$$
$$p = 011001011000$$
$$\hat{p} = 001011011000$$

Then it holds $TD(g, p) = 16 > 13 = TD(g, \hat{p}) \implies m_{TD}(g, p) < m_{TD}(g, \hat{p})$.

Prop. 6

$$g = 000011110000$$
$$p = 000010000000$$
$$\hat{p} = 001011110000$$

Then it holds $TD(g,p) = 6 > 2 = TD(g,\hat{p}) \implies m_{TD}(g,p) < m_{TD}(g,\hat{p})$.

Prop. 7 Let $W \in A(g)$, $\hat{p}_{[n]\setminus\{i^*\}} = p_{[n]\setminus\{i^*\}}$ for some $i^* \in W$, $\hat{p}(i^*) \neq p(i^*) = 0$, and $|A(\hat{p}_W)| \leq |A(p_W)|$. By Thm. D.39 , it directly follows that $m_{TD}(g,\hat{p}) > m_{TD}(g,p)$.

Prop. 8

$$g = 000011111000$$
$$p = 000011000000$$
$$\hat{p} = 000000011000$$

Then it holds $TD(g,p) = TD(g,\hat{p}) = 6 \implies m_{TD}(g,p) = m_{TD}(g,\hat{p})$.

Prop. 9 Consider the following example:

$$g = 000111111000$$
$$p = 000010010000$$
$$\hat{p} = 000011000000$$

Then it holds $TD(g,p) = 4 < 7 = TD(g,\hat{p}) \implies m_{TD}(g,p) > m_{TD}(g,\hat{p})$. □

Theorem D.40. ***Average Alert Delay*** *is defined as*

$$AAD(g,p) = \frac{1}{|\{W \in A(g) \colon W \cap p^{-1}(1) \neq \emptyset\}|} \sum_{\substack{W \in A(g) \colon \\ W \cap p^{-1}(1) \neq \emptyset}} \min_{i \in W \colon p(i)=1} i - \min_{i \in W} i.$$

An Average Alert Delay-based metric $m_{AAD}(g,p) = f(AAD(g,p))$, where f is a strictly decreasing function, satisfies Prop. 5 and Prop. 8.

Proof.

Let $: [n] \to \{0,1\}^n$ be the ground truth and $p, \hat{p} \colon [n] \to \{0,1\}^n$ predictions.

Prop. 1 Consider the following example:

$$g = 000110011000$$
$$p = 000110010000$$
$$\hat{p} = 000110000000$$

Then it holds $AAD(g,p) = 0 = AAD(g,\hat{p}) \implies m_{AAD}(g,p) = m_{AAD}(g,\hat{p})$.

Prop. 2 Consider the following example:

$$g = 000111111000$$
$$p = 000010000000$$
$$\hat{p} = 000010010000$$

Then it holds $AAD(g,p) = 1 = AAD(g,\hat{p}) \implies m_{AAD}(g,p) = m_{AAD}(g,\hat{p})$.

Prop. 3 Consider the following example:

$$\begin{aligned} g &= 000000111000 \\ p &= 000010010000 \\ \hat{p} &= 000110010000 \end{aligned}$$

Then it holds $AAD(g,p) = 1 = AAD(g,\hat{p}) \implies m_{AAD}(g,p) = m_{AAD}(g,\hat{p})$.

Prop. 4 Consider the following example:

$$\begin{aligned} g &= 000000111000 \\ p &= 010010010000 \\ \hat{p} &= 000010010000 \end{aligned}$$

Then it holds $AAD(g,p) = 1 = AAD(g,\hat{p}) \implies m_{AAD}(g,p) = m_{AAD}(g,\hat{p})$.

Prop. 5 Let $W \subset W' \in A(|g-1|)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $|\hat{p}_W^{-1}(1)| = |p_W^{-1}(1)|$, and $|A(\hat{p}_W)| = |A(p_W)|$. Then $\{W \in A(g)\colon W \cap \hat{p}^{-1}(1) \neq \emptyset\} = \{W \in A(g)\colon W \cap p^{-1}(1) \neq \emptyset\}$. Thus, $AAD(g,p) = AAD(g,\hat{p}) \implies m_{AAD}(g,p) = m_{AAD}(g,\hat{p})$.

Prop. 6 Consider the following example:

$$\begin{aligned} g &= 000000111000 \\ p &= 000000010000 \\ \hat{p} &= 000010010000 \end{aligned}$$

Then it holds $AAD(g,p) = 1 = AAD(g,\hat{p}) \implies m_{AAD}(g,p) = m_{AAD}(g,\hat{p})$.

Prop. 7 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000010000000 \\ \hat{p} &= 000011000000 \end{aligned}$$

Then it holds $AAD(g,p) = 1 = AAD(g,\hat{p}) \implies m_{AAD}(g,p) = m_{AAD}(g,\hat{p})$.

Prop. 8 Let $W \in A(g)$, $\hat{p}_{[n]\setminus W} = p_{[n]\setminus W}$, $|A(\hat{p}_W)| = |A(p_W)|$, $|\hat{p}^{-1}(1)| = |p^{-1}(1)|$, and $\min\limits_{i \in W\colon p(i)=1} i < \min\limits_{i \in W\colon \hat{p}(i)=1} i$. Then $\min\limits_{i \in W\colon p(i)=1} i - \min\limits_{i \in W} i < \min\limits_{i \in W\colon \hat{p}(i)=1} i - \min\limits_{i \in W} i$, and hence $AAD(g,p) < AAD(g,\hat{p}) \implies m_{AAD}(g,p) > m_{AAD}(g,\hat{p})$.

Prop. 9 Consider the following example:

$$\begin{aligned} g &= 000111111000 \\ p &= 000010010000 \\ \hat{p} &= 000011000000 \end{aligned}$$

Then it holds $AAD(g,p) = 1 = AAD(g,\hat{p}) \implies m_{AAD}(g,p) = m_{AAD}(g,\hat{p})$. $\square$