READ-BENCH: BENCHMARKING HISTORICAL INSTANCE RETRIEVAL FOR TIME-SERIES DIAGNOSIS

Anonymous authors
Paper under double-blind review

ABSTRACT

Time-series diagnostic systems rarely rely on retrieving relevant historical cases, and when they do, retrieval is typically evaluated only indirectly through downstream prediction. We introduce **READ-Bench**, a benchmark for historical-case retrieval across 12 diagnostic datasets, centered on multivariate time series, that defines relevance by shared fault or event type rather than signal shape, so that visually different traces of the same fault count as relevant while similar-looking traces of different faults do not. Organizing existing datasets into queries, corpora, and explicit relevance judgments, READ-Bench is, to our knowledge, the broadest testbed to date for historical-case retrieval across diagnostic domains. Analogous to retrieval-augmented generation, we treat retrieval as a base retriever followed by a reranker, evaluating classical distances, symbolic retrievers, self-supervised and foundation-model embedders, and their fusion for search, and label-aware and language-model rerankers for reranking, under one protocol that varies supervision, normal-series pollution, and corpus scale with significance testing across datasets. Under a common channel-independent retrieval interface, pretrained representations offer no statistically detectable advantage over strong classical and symbolic baselines for search alone. The decisive factor is instead a small amount of resolved-case supervision at reranking, namely a Gaussian-process reranker that propagates a few neighbor labels in the embedding space and improves rankings far more than swapping among more sophisticated unsupervised representations or language-model reasoning, a gain that holds under corpus pollution and at full corpus scale. Guided by these findings, we introduce normal-residual scoring, which ranks each window by its departure from normal operation, and build a system that fuses a normal-residual-scored foundation-model embedder with a dynamic time warping leg by reciprocal-rank fusion, then reranks with the label-aware Gaussian-process reranker. This system improves NDCG@10 over its own search stage on all 12 datasets, by $+0.11$ from the reranking step alone and by $+0.18$ over the strongest single base retriever applied uniformly across datasets.

1 INTRODUCTION

Operational systems accumulate a rich history of prior experience, including sensor traces, process histories, supply-chain signals, and other time series, often accompanied by metadata such as annotations, maintenance records, shift notes, incident logs, and transaction histories (Liu & Hui, 2024; Varma, 1999; Zhong et al., 2018). Together, these records capture not only what happened, but also the conditions under which it happened, how it was diagnosed, and what actions followed. Effective diagnostic systems should be able to retrieve relevant historical cases from this repository and use them as evidence for interpreting new observations.

Retrieval-augmented generation (RAG) established this pattern for unstructured corpora that consists in retrieving relevant external evidence, and then reasoning over that evidence rather than relying only on the parametric knowledge of the language model (Lewis et al., 2020; Yasunaga et al., 2023). In knowledge-intensive settings, however, the quality of a RAG system depends on multiple interacting components, including similarity search, reranking, and downstream reasoning. Understanding and improving such systems therefore requires evaluating these components at the appropriate level of granularity. Aggregate end-to-end metrics can mask important differences across retrieval settings, modalities, and query types, whereas targeted evaluation can reveal where performance gains

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

and failures actually arise (Wu et al., 2024). Given the central role of the retrieval component, which commonly encompasses similarity search followed by reranking, benchmarks such as MS MARCO and BEIR (Bajaj et al., 2016; Thakur et al., 2021) helped make retrieval itself a measurable, general capability and enabled systematic comparison of retrieval methods across settings. Time-series diagnosis presents an analogous problem, but the primary evidence is often structured temporal data. A retrieved historical case can serve as the entry point to its associated metadata and unstructured records, providing richer context for diagnosis. Recent systems approach time-series diagnosis in different ways, including prompting general-purpose LLMs (Alnegheimish et al., 2024), aligning and fine-tuning time-series language models (Xie et al., 2025; Yang et al., 2026), and equipping agents with tools and memory (Tao et al., 2026). ARFBench further evaluates multimodal models on diagnostic question answering over real software incidents (Xie et al., 2026). These efforts show rapid progress in reasoning over time-series observations, but retrieving the right historical cases remains a distinct capability. Across these different system designs, retrieval provides a common mechanism for grounding diagnosis in relevant prior experience.

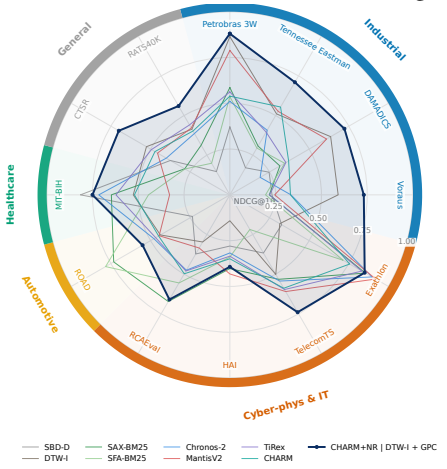


Figure 1: Per-dataset NDCG@10 over the 12 READ-Bench datasets (spokes grouped by diagnostic domain). The base retrievers, coloured by family, form a tight inner band with no clear winner, while our system (CHARM + NR | DTW-I + GPC) improves over them on most datasets.

Several lines of work address this retrieval problem. Retrieved examples support forecasting (Han et al., 2025) and anomaly detection (Liu et al., 2025). TRACE uses textual context to learn representations for retrieval (Chen et al., 2025), while comparative studies evaluate representations and unsupervised reranking (Rozin et al., 2025). TSRBench (Hu et al., 2026) provides a common evaluation framework, but its open dataset requires relevant instances to have similar shapes, its diagnostic evaluation is limited to telecom incidents, and it does not test fusion or reranking. Despite this progress, there remains limited work comparing retrieval methods across diagnostic domains, particularly when relevant instances differ in signal shape.

The role of retrieval in this setting is to surface historical cases that are diagnostically relevant to a query. This means cases corresponding to the same underlying fault, anomaly, or event. The central challenge is that diagnostic relevance is not equivalent to signal similarity. The same fault may manifest differently across operating conditions, while similar-looking signals may arise from unrelated causes. In multivariate series, the informative evidence may further be localized to only a few channels or short intervals and obscured by unrelated variation (Yeh et al., 2017). As with RAG systems, downstream prediction or end-to-end performance alone cannot isolate retrieval quality, since it depends both on which cases are retrieved and on how those cases are subsequently used. Evaluating retrieval therefore requires explicit relevance judgments and controlled analysis of the retrieval pipeline itself, including representation choice, search, reranking, corpus composition, and access to supervision.

Our contributions are:

- **READ-Bench (Retrieval Evaluation for Anomaly Diagnosis), a benchmark for diagnostic retrieval.** We introduce a 12-dataset benchmark spanning five diagnostic domains, with multivariate and complementary univariate evaluations and explicit fault- and event-level relevance judgments.
- **A systematic study of retrieval design.** We compare classical, symbolic, and pretrained retrieval methods across representations, fusion strategies, and reranking methods, and evaluate their robustness to corpus composition and scale.
- **A retrieval component for time-series diagnosis.** Guided by the findings, we combine representation-based search, complementary dynamic-time-warping search, and label-aware Gaussian-process reranking. The resulting component improves NDCG@10 over its underlying search stage on all 12 datasets and outperforms single-method retrievers across the five diagnostic domains (Figure 1).

2 Related Work

Diagnosis and retrieval augmented systems. Retrieving similar historical instances to support time-series fault diagnosis (Zhao et al., 2017) has been studied before, and recent work extends it with language and multimodal models. ChatTS (Xie et al., 2025) treats time series as a native input modality for understanding and diagnosis; Time-RA (Yang et al., 2026) frames anomaly analysis as detection, categorization, and diagnosis; ARFBench (Xie et al., 2026) evaluates question answering over real software-incident telemetry; and AnomaMind uses tool-augmented agentic reasoning for anomaly detection (Tao et al., 2026). Other systems retrieve historical examples before the downstream task, namely LLMAD for anomaly detection and explanation (Liu et al., 2025), retrieval for PHM diagnosis (Mizoguchi et al., 2023), ZARA for language-model reasoning over motion time series (Li et al., 2026), and retrieval for forecasting (Han et al., 2025; Tire et al., 2026; Ning et al., 2025). These systems are usually evaluated on the final task, making it difficult to separate retrieval quality from how retrieved instances are later used.

Similarity and representations. Time-series retrieval can use several representations and similarity measures. Euclidean distance compares aligned values, dynamic time warping (DTW) (Sakoe & Chiba, 1978) allows local temporal alignment, and shape-based distance (SBD) (Paparrizos & Gravano, 2015) compares shape using normalized cross-correlation. SAX (Lin et al., 2007) and SFA (Schäfer & Högqvist, 2012) convert signals into symbolic sequences searchable with BM25 (Robertson & Zaragoza, 2009). For multivariate series, similarity also depends on how channels are combined (Shokoohi-Yekta et al., 2017; d'Hondt et al., 2025), and irrelevant channels can hide useful patterns (Yeh et al., 2017). Self-supervised encoders (Yue et al., 2022; Franceschi et al., 2019; Tonekaboni et al., 2021) and pretrained time-series models such as CHARM, MantisV2, Chronos-2, and TiRex (Dutta et al., 2026; Feofanov et al., 2026; Ansari et al., 2025; Auer et al., 2025) provide learned representations usable for retrieval, even though search is not their primary training objective.

Retrieval and reranking. Several works study time-series retrieval. Deep r-RSJBE and DUBCNs learn supervised and unsupervised representations for multivariate retrieval (Song et al., 2018; Zhu et al., 2020); CTSR retrieves time series together with metadata (Yeh et al., 2023); and TRACE connects time series with textual context for time-series and cross-modal retrieval (Chen et al., 2025). Other works study the latter ranking stages by comparing representations and unsupervised reranking (Rozin et al., 2025), combining rank information across channels (Rozin & Pedronette, 2026), or fusing multiple rankings before DTW refinement (Barros et al., 2026).

Benchmarks. TS-Haystack evaluates retrieval and reasoning over sparse events within long time-series contexts rather than ranking historical cases from a corpus (Zumarraga et al., 2026). TSR-Bench is closer to corpus retrieval, covering similar-series and industrial incident retrieval (Hu et al., 2026), but its open similar-series setting uses univariate UCR data and ties relevance closely to signal shape, while its multivariate diagnostic setting is limited to a single telecom domain.

3 READ-Bench

3.1 Problem Formulation

Data model. The unit of retrieval is a window $x \in \mathbb{R}^{T\times C}$ of T timesteps over C channels, with row $x_t \in \mathbb{R}^C$ the reading at timestep t and column $x^{(c)} \in \mathbb{R}^T$ the c-th channel. A window is univariate when $C = 1$ and multivariate when $C > 1$. Length and channel count are per-window, so windows live in $\mathcal{X} = \bigcup_{T,C} \mathbb{R}^{T\times C}$ and a query and a corpus item need not share either.

Retrieval task. Given a corpus of windows $\mathcal{C} = \{x_1, \ldots, x_N\}$ and a query window $q \in \mathcal{X}$, a retrieval method scores each corpus item with $s(q, \cdot) : \mathcal{C} \to \mathbb{R}$ and ranks $\mathcal{C}$ in descending order of that score, returning the top-K items. A larger score means greater estimated relevance, so a distance d enters as $s = -d$. We write π_q for the induced ranking and $\mathrm{rank}(q, x)$ for the position of x in it, and measure quality by where the relevant items land in π_q. A second-stage reranker produces a new ranking of a candidate pool $\mathcal{P} \subseteq \mathcal{C}$ that a first-stage retriever returns, so it is scored on π_q by the same criterion. The method families that instantiate s are defined in Section 3.3.

Relevance. Every window carries a single label y, either one of the dataset's fault types or normal $\emptyset$, derived from the window's per-timestep annotations by the rule of Section 3.2. An item x is relevant to a query q when the two share a non-normal fault type,

$$\mathrm{rel}(q, x) \;=\; \mathbb{1}\big[\, y_x = y_q \,\wedge\, y_q \neq \emptyset \,\big] \;\in\; \{0, 1\}, \qquad (1)$$

so queries are anomalous by construction, corpus anomalies of the query's type are the retrieval targets, and normal windows are never relevant (Figure 2). Relevance is binary and multi-target, so a query typically has several relevant items in the corpus. These labels define the evaluation target only. The fault label of a query is never revealed to a retrieval method, and corpus labels are exposed solely to the methods designated as supervised or label-aware in Section 3.3.

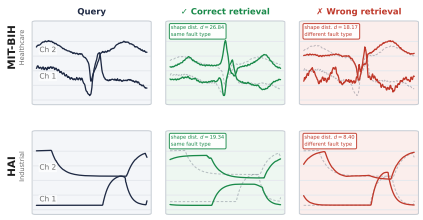


Figure 2: The correct retrieval shares the query's fault type but differs in shape, while a wrong retrieval of a different fault is closer in raw shape, so relevance is not signal similarity.

3.2 Benchmark Construction

Datasets. READ-Bench is built on 12 multivariate datasets drawn from community and industrial sources that, to our knowledge, have not previously been curated for retrieval. These are CTSR (Yeh et al., 2023; Dau et al., 2019), DAMADICS (Bartyś et al., 2006), Exathlon (Jacob et al., 2021), HAI (Shin et al., 2020), MIT-BIH (Moody & Mark, 2001), Petrobras 3W (Vargas et al., 2019), RATS40K (Yang et al., 2026), RCAEval (Pham et al., 2025), ROAD (Verma et al., 2024), TelecomTS (Feng et al., 2025), Tennessee Eastman (Downs & Vogel, 1993), and Voraus (Brockmann et al., 2023). They span condition monitoring, industrial control, microservice and telecom operations, automotive intrusion, and physiological signals, with fault signatures that differ markedly across and within domains (Figure 5 and Figure 6). Channel count ranges from a single channel to 664, window length from 16 to 1024 timesteps, and the number of fault classes per dataset from four to ninety-four, with per-class corpus sizes spanning three orders of magnitude and a max-to-min class ratio that reaches 342:1 on the most skewed dataset (Appendix A.1.2). Prior time-series retrieval evaluations are built on a single univariate archive with balanced classes and no channel axis, whereas this variation across 12 independent sources keeps a method from succeeding by overfitting one domain. The per-dataset composition is given in Table 3.

Windowing and labeling. We respect each dataset's existing conventions, keeping the windows and labels of any dataset that ships pre-windowed and pre-labeled, and otherwise segmenting each series into fixed-length windows. Datasets provide per-timestep labels $\ell_1, \ldots, \ell_T$ over a fault vocabulary in which 0 is the normal state, and where a dataset codes a fault's transient onset separately from its steady phase we merge the two so each fault carries one label. A window takes a fault label only when that fault holds a strict majority of its timesteps, with ties broken toward the lower-indexed label so the normal state is preferred; that is, $y = c$ when $|\{t : \ell_t = c\}| > T/2$ for some fault c, and $y = \emptyset$ (normal) otherwise. Because a fault shorter than the window can never hold this global majority, we relax the rule so that a fault also claims the window when it fills a majority of its own extent within the window. Writing $e_c = [\min\{t : \ell_t = c\}, \max\{t : \ell_t = c\}]$ for the span from a fault c's first to last occurrence in the window and $|e_c|$ for its length, the label is

$$y = \begin{cases} c, & \text{if } \big|\{\, t : \ell_t = c \,\}\big| > T/2 \text{ for some fault } c, \\ c, & \text{else if } \big|\{\, t \in e_c : \ell_t = c \,\}\big| > |e_c|/2 \text{ for some fault } c, \\ \emptyset, & \text{otherwise,} \end{cases} \qquad (2)$$

with $\emptyset$ denoting a normal window. The relaxation recovers short faults without changing any window a global majority already decides. A handful of datasets label at a coarser granularity that the data dictates, such as a heartbeat window taking the majority class of the beats it spans, or a robot operation whose fault flag is constant over the recording. Appendix A.1.3 documents how each dataset's native annotations yield its fault-type labels.

Split protocol. For every dataset we draw queries from the test split and build the corpus from the train split, so no query window appears in the corpus it is ranked against. We respect each dataset's

native split convention where one exists and otherwise split at a grouping coarser than the window (per fault instance, recording, entity, run, or patient), so train and test stay disjoint; per-dataset split units are listed in Table 4. When a split must be subsampled to the target size, we sample anomalies stratified by fault type so the native class imbalance is preserved rather than flattened, and draw normal windows uniformly at random. All splits use a fixed seed, and all headline results in Section 4 are macro-averaged over the 12 datasets.

Corpus composition. The corpus a query is ranked against pairs its relevant set, the same-type anomalies of Equation 1, with distractors of two kinds. Anomalies of a different fault type probe whether a method separates fault types rather than merely detecting abnormality, and normal windows are the easy negatives that dominate any real repository. We control the second kind explicitly, drawing normal windows from the training split to report three pollution levels ρ, the fraction of the corpus that is normal, at roughly ρ=0%, 10%, and 20%, so robustness to easy negatives can be read off directly. Because fault types are class-imbalanced, the number of relevant items per query varies widely, and the metrics of Section 3.4 are chosen to remain meaningful under that imbalance.

3.3 METHODS

READ-Bench organizes the retrieval pipeline into three stages, namely a base retriever that ranks the full corpus, fusion that combines complementary rankings, and a reranker that refines a short candidate pool. We also include fully supervised classifiers as a reference ceiling. The methods therefore differ along two axes, namely the representation used to compare windows and the amount of label information available at ranking time. Unless stated otherwise, methods are label-free. Exact formulations and implementation details are given in Appendix A.2.

Base retrievers. All base retrievers instantiate the scoring function of Section 3.1, but encode similarity differently. Raw-window distances compare the series directly, where Euclidean distance is lock-step, SBD uses shape-aligned cross-correlation, and DTW allows elastic alignment, either independently across channels (DTW-I) or under a shared alignment (DTW-D). Symbolic retrievers first discretize each channel and then retrieve with BM25, using either SAX value symbols or SFA frequency symbols. To compare windows of differing length or channel count, the lock-step distance and the symbolic retrievers reconcile a pair onto a common grid (zero-padding channels and interpolating length), the elastic and cross-correlation distances align unequal lengths natively, and the embedders mean-pool over channels and resample length to the encoder's admissible grid (Appendix A.2). Foundation-model (FM) embedders—CHARM, Chronos-2, MantisV2, and TiRex—map each window to a fixed-dimensional embedding and rank corpus items in that space. We use Euclidean distance for the base embedding score; cosine produces nearly identical rankings in our setting, consistent with prior time-series retrieval benchmarks. .

Absolute embedding distance can, however, be dominated by the operating regime shared by normal and faulty windows, while diagnostic relevance depends on how a window departs from normal operation. The Normal-Residual (NR) scorer changes this reference point without changing the encoder, keeping only the part of a window's embedding that its normal expectation leaves unexplained. Writing $\hat{u} = u/\|u\|$ for ℓ_2 normalization, let $\mathbb{E}[\hat{\phi}(u) \mid \text{normal}]$ be the embedding we would expect for u if it were normal; the residual is

$$r(u) = \hat{\phi}(u) - \mathbb{E}\left[\hat{\phi}(u) \mid \text{normal}\right], \qquad s_{\text{NR}}(q, x) = \hat{r}(q)^\top \hat{r}(x). \tag{3}$$

We approximate this conditional expectation by a k-nearest-neighbour estimate over a train-split pool $\mathcal{N}$ of normal windows: $\mathbb{E}[\hat{\phi}(u) \mid \text{normal}] \approx \frac{1}{k} \sum_{x \in \mathcal{N}_k(u)} \hat{\phi}(x)$, the mean of the k normal embeddings nearest $\hat{\phi}(u)$ in cosine similarity (we use k=30). Averaging a small neighbourhood rather than the single nearest normal is a variance-reduction step, since one nearest normal injects its own embedding noise into every residual, whereas the local mean cancels that noise while staying specific to the window's operating point. NR then compares residual directions by cosine, so $r(u)$ is what remains of a window after removing its expected-if-normal component, and NR measures whether the query and a candidate depart from normal in a similar direction. It requires normal windows but no fault labels and applies to any fixed embedder ϕ; the per-embedder factorial is reported in Table 10.

Supervised reference. To estimate how much fault-type structure is recoverable with full label access, we also train MR-Hydra (Dempster et al., 2023) and RDST (Guillaume et al., 2022), both strong recent time-series classifiers (Middlehurst et al., 2024), on the labeled corpus. Their class posteriors are converted to rankings as detailed in Appendix A.2. Because these models are trained as supervised classifiers rather than retrieval systems, we treat them as a reference ceiling rather than as directly comparable retrievers.

Fusion. Learned embeddings and shape or symbolic retrievers encode different notions of similarity, so we test whether their rankings contain complementary relevant items. This mirrors text retrieval, where fusing a lexical retriever such as BM25 with a dense one recovers relevant documents that neither leg ranks alone (Karpukhin et al., 2020; Formal et al., 2021); our symbolic BM25 legs and embedding retrievers play the analogous sparse and dense roles here. Each fusion pairs one embedder with one shape or symbolic retriever, and combines them with one of three standard rank- or score-level schemes, reciprocal-rank fusion (RRF) (Cormack et al., 2009), a z-scored weighted sum, and a two-stage cascade, none of which requires the two legs to share a score scale. RRF is our default, since it needs no comparable scores and performs best here (Section 4.3). The exact combination rules are standard and are given in Appendix A.2; the full embedding$\times$shape/symbolic grid is evaluated in Section 4.3.

Rerankers. Modern retrieval systems are two-stage, where a cheap first stage returns a candidate pool that a stronger reranker then re-orders (Thakur et al., 2021; Khattab & Zaharia, 2020). A reranker receives the candidate pool $\mathcal{P}$ returned by a base or fused retriever and changes only its ordering. We avoid training a global relevance model on benchmark labels and instead study three forms of additional information. Pseudo-relevance feedback (PRF) (Rocchio, 1971) is label-free and re-scores candidates relative to a pseudo-relevance centroid. Label-aware rerankers use only a small, query-local set of candidate labels. Purity rewards candidates whose local neighbourhood is label-consistent, Majority-Vote infers the query class from nearby labeled items, and query-conditioned Purity (QPurity) combines the two. This setting models a diagnostic repository in which some resolved incidents have confirmed labels without assuming labels for the full corpus (Yu et al., 2024; del Campo Barraza et al., 2021).

The Gaussian-process classifier (GPC) reranker (Rasmussen & Williams, 2006) uses the same local supervision but replaces voting with a learned local decision boundary. It fits a GPC on the labeled candidates in $\mathcal{P}$ and ranks a candidate x by the probability that it shares the query's inferred class:

$$s_{\mathrm{GPC}}(q, x) = \Pr(y_x = y_q \mid q, x, \mathcal{P}), \tag{4}$$

where the query label y_q is latent and is never provided to the reranker. GPC is deliberately a cheap, domain-agnostic reranker, since it operates on the base representation with no per-dataset tuning and adds only a fraction of a second per query, in contrast to language-model rerankers (Sun et al., 2023) that read the raw series through a hosted model at far greater cost.

Finally, we test language-model rerankers that read the retrieved series directly, namely Toto-1.0-QA (Cohen et al., 2024), ChatTS, Claude, and Claude + TSAD following LLMAD. Together, these methods let us separate gains from the base representation, complementary retrieval signals, and additional information introduced only at reranking time.

3.4 Evaluation Protocol

Metrics. We report precision at K (P@K), hit rate at K (HR@K), and normalized discounted cumulative gain at K (NDCG@K). P@K and HR@K measure the precision of the retrieved set without regard to order, namely the fraction of the top-K that is relevant and whether any relevant item appears in the top-K, while NDCG@K additionally rewards ranking relevant items above irrelevant ones. We report all three because a method can satisfy one axis while failing the other, for instance by retrieving every relevant item but ranking it last. Under the binary, multi-target relevance of Equation 1 the ideal ranking places all relevant items first, so NDCG@K is normalized against a query-specific ideal that accounts for the varying number of relevant items. Formal definitions of all three metrics are given in Appendix A.3.

Significance testing. Because averaged metric differences can arise from noise rather than a genuine gap, we compare methods with a significance test rather than raw score differences. We use

a Friedman test with a Nemenyi post-hoc test, blocking on the 12 datasets, so that a difference is reported as significant only when the methods' mean ranks differ by more than the Nemenyi critical difference. The test statistic and threshold are given in Appendix A.4.

Per-dataset breakdown. Beyond the pooled tests we report per-dataset average metrics as a descriptive robustness check across domains. We treat this as a diagnostic rather than a formal test, since 12 datasets are too few to serve as a reliable blocking unit for hypothesis testing. The full per-dataset tables appear in Appendix A.8, and the macro-averaged grids in Appendix A.7.

4 Experiments and Analysis

With the task, methods, and protocol fixed in Section 3, we ask what retrieves the right cases, whether fusion and reranking improve those rankings, and how far this reaches toward full supervision. We proceed in the order the pipeline is built. We start with the label-free base retrievers, add NR scoring and test its robustness to corpus pollution, then treat fusion and reranking as separate additions, and finally combine them. The results tell a consistent story. The base retrievers land on a tight plateau that NR scoring lifts only modestly, while a small amount of label information lifts it decisively. Fusion helps only selectively, and language-model reranking does not help at all. Headline numbers are NDCG@10 macro-averaged over the 12 datasets, reported with the significance test of Section 3.4. Embedders are scored by Euclidean distance unless stated otherwise.[1]

4.1 Base retrievers

We evaluate the label-free retrievers of Section 3, namely the raw-window distances, the symbolic BM25 retrievers, and the frozen foundation-model embedders. Table 1 reports their NDCG@10 next to the supervised reference, on each of the three pollution levels ρ of Section 3.2, the fraction of the corpus made up of normal series (0%, $\approx 10\%$, $\approx 20\%$). To keep runtime and memory comparable across datasets, each corpus is capped to a common target size (the per-dataset sizes are in Table 3); the remaining metrics (P@$\{1, 5\}$, HR@10, NDCG@$\{5, 20\}$, and macro-F1) are in Appendix A.7.1. The cap does not change the conclusion below, because for the methods that scale to the full uncapped corpus the scores barely move (Appendix A.7.2).

All of the label-free retrievers score far below the supervised reference, and they score close to one another (Figure 7b). The significance test separates only the weakest of them, SBD-D, from the rest; the top four (MantisV2, SAX-BM25, CHARM, TiRex) are statistically indistinguishable. The differences among the leaders are smaller than the variation from one dataset to the next, so the choice among a pretrained embedding, a classical distance, and a symbolic retriever does not change retrieval quality. The gap to the supervised reference, in contrast, is large and statistically clear. The rest of the section asks which additions can lift a retriever above this level, starting with NR scoring.

4.2 Normal-residual scoring and robustness to pollution

Representation choice starts to matter once the corpus is polluted with normal series (the pollution levels of Table 1). We compare every base retriever against the NR scorer of Section 3.3, which subtracts each window's expected-if-normal embedding so that similarity reflects departure from normal operation.

Pollution degrades every base retriever, but by very different amounts. The raw distances fall fastest, while NR is both the most accurate base retriever and the most robust. It lifts clean accuracy on every embedder it wraps except TiRex, which is essentially unchanged (the largest gains are on CHARM and MantisV2), and roughly halves the degradation under pollution, so the NR retrievers stay well above the raw distances, symbolic retrievers, and ED embedders as the normal fraction grows (Figure 3). MantisV2 + NR is the winner on both counts, the strongest single base retriever clean and the one that gives up the least under pollution (the per-embedder sweep is in Appendix A.7.3).

[1] We name a composed retriever by its parts, so a bare embedder name (for example CHARM) denotes that default Euclidean score, an embedder with normal-residual scoring appends the residual marker (for example CHARM + NR), a fused leg joins with a bar (CHARM + NR | DTW-I), and a reranker joins with a plus (CHARM + NR | DTW-I + GPC).

Table 1: NDCG@10 of every base retriever, our best composed system, and the supervised reference, macro-averaged over the 12 datasets at each pollution level ρ, the fraction of the corpus diluted with normal-class series ($0\%/\approx 10\%/\approx 20\%$). The supervised rows are a classification reference rather than retrieval systems. In each column the best value is bold and the second best underlined, over the base retrievers and our system (the supervised reference excluded).

Method	Family	ρ=0%	ρ=10%	ρ=20%
Random	Trivial floor	0.189	0.168	0.155
Majority		0.246	0.242	0.166
SBD-D	Raw-window distance	0.328	0.271	0.245
DTW-I		0.480	0.414	0.395
SAX-BM25	Symbolic	0.489	0.457	0.431
SFA-BM25		0.409	0.386	0.376
Chronos-2	FM embedder	0.471	0.455	0.432
MantisV2		<u>0.510</u>	<u>0.474</u>	<u>0.454</u>
TiRex		0.483	0.449	0.433
CHARM		0.486	0.461	0.437
CHARM + NR | DTW-I + GPC	Our best system	**0.687**	**0.640**	**0.619**
MR-Hydra	Supervised reference	0.776	0.672	0.658
RDST		0.711	0.634	0.590

4.3 Fusion and Reranking

Two additions can improve a base retriever. Fusion combines it with a complementary shape or symbolic leg, and reranking re-orders its candidate pool using extra information. We report each here, then combine them in Section 4.4.

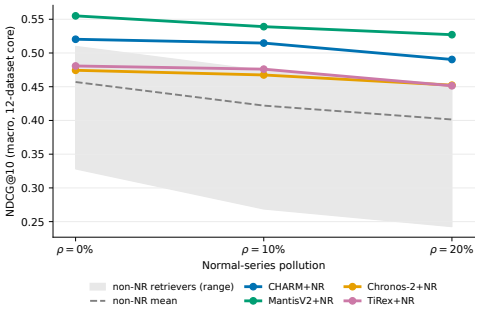


Figure 3: The NR embedder variants against the non-NR field as the corpus is diluted with normal series (NDCG@10, macro-averaged over the 12 datasets). The shaded band spans the non-NR retrievers (raw distances, symbolic BM25, and ED embedders) with their mean dashed; the four NR embedder variants (bold) sit above it and stay high at every pollution level while the field degrades. MantisV2 + NR is the strongest base retriever throughout.

Fusion. Because the base retrievers are close, we test whether pairing an embedder with a complementary leg helps, fusing by RRF, which needs no comparable scores and outperforms weighted-sum and two-stage fusion. Fusion gives a small, mostly on-plateau gain, the best leg for every embedder being DTW-I, and the gain is consistent across embedders rather than specific to one (Appendix A.7.4). Fusion alone does not leave the plateau, but a fused pool is a stronger starting point than either leg alone and reappears as a building block later.

Reranking. Reranking re-orders a fixed pool with the methods of Section 3, grouped by the label information they use (none, neighbour labels, or a language model). We apply the rerankers to a representative set of pools rather than every one, namely the strongest raw distance (DTW-I), the four foundation-model embedders, and their NR variants; the weaker distances (SBD-D, DTW-D) and symbolic retrievers add little as pools and are omitted. We use the language model two ways, reranking the retrieved pool and, as a no-retrieval baseline, classifying the anomaly type directly from the series. Appendix A.7.6 collects the statistical rerankers and Table 2 the language-model results. Three tiers emerge (visualised in Figure 4; full grids across pollution levels in Appendix A.7.7, and the language-model rerankers in Table 2). Label-aware reranking is the only addition that reliably helps, since a GPC reranker improves every pool it is applied to, with a significant gain on every pool (Appendix A.7.6), and narrows the gap to the supervised reference without closing it. Label-free reranking hurts most pools, so we drop it from later use. Language-model reranking of the fixed CHARM pool barely moves NDCG@10 across all four models (Toto-1.0-QA, ChatTS,

Table 2: Language-model methods on the CHARM pool (NDCG@10). The first block reranks the retrieved pool and the second classifies the fault type with no retrieval. Best per block bold, second best underlined.

Method	NDCG@10
CHARM (base)	0.486
Reranking the CHARM pool	
Toto-1.0-QA	<u>0.493</u>
ChatTS	0.475
Claude	**0.496**
Claude + TSAD	0.492
Classification, no retrieval	
Toto-1.0-QA	0.159
ChatTS	0.169
Claude	<u>0.180</u>
Claude + TSAD	**0.184**

Claude, Claude + TSAD), an order of magnitude less than the statistical reranker, and the same models asked to name the type without retrieval score near the random floor. We use CHARM here for budget and because its low cross-dataset variance (Appendix A.7.1) makes it a stable pool on which any real effect should be easiest to see, so the pool choice is not to blame. On this task, then, retrieval is what makes the language model useful, its reasoning over the raw series adds little, while a small amount of label information at rerank time adds the most.

4.4 Combining fusion and reranking

The preceding subsections added one component at a time, either a fused shape leg or a label-aware reranker. Because the two draw on different signals, namely a second view of the series and a few neighbour labels, we now combine them, fusing a base retriever with its best leg (DTW-I) and reranking the fused pool with GPC at each pollution level (Appendix A.7.8). We build these stacks on the NR variants and include the plain ED embedders as a non-NR reference.

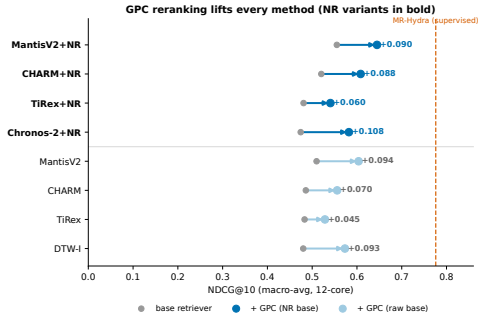


Figure 4: Effect of GPC reranking across base retrievers (NDCG@10, macro-averaged over the 12 datasets), the full grid is in Table 14. Each arrow runs from a base retriever to its GPC-reranked version, and the label-aware reranker gives a consistent gain on every base retriever. The dashed line marks the supervised reference (MR-Hydra), which no single-pool system reaches.

The two additions are close to additive, since reranking a fused pool adds about the same margin it adds to an unfused one. The best combined system, CHARM + NR | DTW-I + GPC, reaches NDCG@10 of 0.687/0.640/0.619 at ρ=0/10/20% and is the strongest at every level. Its per-dataset envelope sits above the individual retrievers on most datasets and across all five diagnostic domains (Figure 1). Normal-residual scoring is what carries it clear of the plain-embedding stacks, as CHARM | DTW-I + GPC reaches only 0.631, while fusion contributes a smaller final increment on top of the base representation and the reranker.

Our combined method runs an NR-scored foundation embedder, fuses in a DTW-I leg, and reranks with GPC, none of which needs task-specific training, and it remains inexpensive. GPC adds only a fraction of a second per query, so the combined systems run far below the supervised classifiers while closing much of the gap to them (Figure 7a).

5 Conclusion

We introduced READ-Bench, a rigorous framework for evaluating time-series anomaly diagnosis through information retrieval, addressing the gap between raw anomaly detection and actionable root-cause localization across 12 diverse industrial domains. Our experiments yield three main conclusions. First, pretrained embedding models show no advantage over strong classical and symbolic baselines, both falling well below a supervised reference. Second, leveraging additional information, specifically small amounts of label data or normal periods, serves as a decisive lever; non-parametric reranking and Gaussian-process reranking improve whichever retrieval pool is strongest, proving to be the most reliable additions even under normal-series pollution. Third, score fusion helps only in select settings, and language-model reranking of a fixed pool shows almost no detectable effect, indicating that LLMs provide almost no benefit in this setting.

Our evaluation is bounded by the original labeling fidelity and categorization granularity of the source repositories, which may occasionally obscure subtle cascading failures or multi-fault interactions, alongside choices in labeling and window strategies that naturally influence conclusions. Looking beyond static retrieval, our immediate roadmap centers on evolving READ-Bench into a dynamic, multi-turn QA environment that mirrors real-world diagnostic workflows. Ideal diagnostic usage requires interactive troubleshooting where a system must inspect telemetry windows, consult documentation, and iteratively refine its conclusions. To support this, we aim to explore coding agents, the ideation of possible reasons for failures, and robust uncertainty quantification.

Ethics Statement

READ-Bench is assembled entirely from publicly available time-series datasets, released by their originators for research use; we redistribute only curated queries, corpora, and relevance judgments derived from them, and retain each source's original license and attribution. No new data were collected from human subjects, and the physiological source (MIT-BIH) is already de-identified by its providers. The benchmark is intended to measure how well retrieval methods recover relevant historical cases for time-series diagnosis, a step that can support human operators rather than replace their judgment; as with any diagnostic tool, retrieved evidence should be reviewed by a qualified expert before it informs a decision. We are not aware of harmful dual-use risks beyond those already present in the underlying public datasets.

AI Use Statement

We used generative AI tools solely to refine and polish the writing of this paper, namely to improve readability, tighten phrasing, and suggest a title and keywords. We did not use generative AI tools to generate synthetic data, develop theoretical models or conceptual frameworks, formulate mathematical claims or proofs, propose or refine hypotheses, design experiments or methodology, implement methods, clean or reformat datasets, or interpret results; these were carried out by the authors. We have reviewed all AI-assisted text, and we take responsibility for the final content of this work, including all text, claims, and artifacts.

Reproducibility Statement

We designed READ-Bench for reproducibility. The task, relevance definition, and windowing rules are specified in Section 3, with per-dataset curation in Appendix A.1.1 and the labeling and split protocols in Appendix A.1.3. Formal definitions of every method are given in Appendix A.2, the evaluation metrics in Appendix A.3, and the significance-testing procedure in Appendix A.4; hyperparameters and the compute budget are listed in Appendix A.5. All splits and subsampling use a fixed seed (Section 3.2). Upon acceptance we will release the curated corpora, queries, and relevance judgments, together with the code for building corpora, running the retrieval methods, and reproducing every table and figure, and will additionally publish the benchmark as a versioned dataset on the Hugging Face Hub.

References

Sarah Alnegheimish, Linh Nguyen, Laure Berti-Equille, and Kalyan Veeramachaneni. Can large language models be anomaly detectors for time series? In *IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, 2024. arXiv:2405.14755.

Abdul Fatir Ansari, Oleksandr Shchur, Jaris Küken, Andreas Auer, Boran Han, Pedro Mercado, Syama Sundar Rangapuram, Huibin Shen, Lorenzo Stella, Xiyuan Zhang, Mononito Goswami, Shubham Kapoor, Danielle C Maddix, Pablo Guerron, Tony Hu, Junming Yin, Nick Erickson, Prateek Mutalik Desai, Hao Wang, Huzefa Rangwala, George Karypis, Yuyang Wang, and Michael Bohlke-Schneider. Chronos-2: From univariate to universal forecasting. *arXiv preprint arXiv:2510.15821*, 2025.

Andreas Auer, Patrick Podest, Daniel Klotz, Sebastian Böck, Günter Klambauer, and Sepp Hochreiter. Tirex: Zero-shot forecasting across long and short horizons with enhanced in-context learning. *Advances in Neural Information Processing Systems*, 38:57529–57580, 2025.

Payal Bajaj, Daniel Campos, Nick Craswell, Li Deng, Jianfeng Gao, Xiaodong Liu, Rangan Majumder, Andrew McNamara, Bhaskar Mitra, Tri Nguyen, Mir Rosenberg, Xia Song, Alina Stoica, Saurabh Tiwary, and Tong Wang. MS MARCO: A human generated machine reading comprehension dataset. *arXiv preprint arXiv:1611.09268*, 2016.

Gustavo Barros, Noel Appel, Felix Thomas, and Bernd Kuhlenkötter. Data-driven trajectory performance prediction for industrial robots via multi-modal retrieval. *The International Journal of*

Advanced Manufacturing Technology, 145:4201–4221, 2026. doi: 10.1007/s00170-026-18597-2. URL https://doi.org/10.1007/s00170-026-18597-2.

Michał Bartyś, Ron Patton, Michał Syfert, Salvador de las Heras, and Joseba Quevedo. Introduction to the DAMADICS actuator FDI benchmark study. In *Control Engineering Practice*, volume 14, pp. 577–596, 2006.

Jan Thieß Brockmann, Marco Rudolph, Bodo Rosenhahn, and Bastian Wandt. The voraus-AD dataset for anomaly detection in robot applications. *IEEE Transactions on Robotics*, 2023.

Jialin Chen, Ziyu Zhao, Gaukhar Nurbek, Aosong Feng, Ali Maatouk, Leandros Tassiulas, Yifeng Gao, and Rex Ying. TRACE: Grounding time series in context for multimodal embedding and retrieval. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025. arXiv:2506.09114.

Ben Cohen, Emaad Khwaja, Kan Wang, Charles Masson, Elise Ramé, Youssef Doubli, and Othmane Abou-Amal. Toto: Time series optimized transformer for observability technical report. *arXiv preprint arXiv:2407.07874*, 2024.

Gordon V. Cormack, Charles L. A. Clarke, and Stefan Buettcher. Reciprocal rank fusion outperforms Condorcet and individual rank learning methods. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pp. 758–759, 2009.

Hoang Anh Dau, Anthony Bagnall, Kaveh Kamgar, Chin-Chia Michael Yeh, Yan Zhu, Shaghayegh Gharghabi, Chotirat Ann Ratanamahatana, and Eamonn Keogh. The UCR time series archive. *IEEE/CAA Journal of Automatica Sinica*, 6(6):1293–1305, 2019.

Sergio Martin del Campo Barraza, William Lindskog, Davide Badalotti, Oskar Liew, and Arash Toyser. Active learning framework for time-series classification of vibration and industrial process data. In *Annual Conference of the PHM Society*, volume 13, 2021.

Angus Dempster, Daniel F Schmidt, and Geoffrey I Webb. Hydra: competing convolutional kernels for fast and accurate time series classification: A. dempster et al. *Data Mining and Knowledge Discovery*, 37:1779–1805, 2023.

Janez Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006.

Jens E d'Hondt, Haojun Li, Fan Yang, Odysseas Papapetrou, and John Paparrizos. A structured study of multivariate time-series distance measures. *Proceedings of the ACM on Management of Data*, 3(3):1–29, 2025.

James J. Downs and Ernest F. Vogel. A plant-wide industrial process control problem. In *Computers & Chemical Engineering*, volume 17, pp. 245–255, 1993. Tennessee Eastman process; run-to-failure simulation data from Rieth et al. 2017, Harvard Dataverse DOI 10.7910/DVN/6C3JR1.

Utsav Dutta, Gerardo Pastrana, Sina Khoshfetrat Pakazad, and Henrik Ohlsson. Giving sensors a voice: Multimodal JEPA for semantic time-series embeddings. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=lNoaqrOXti.

Austin Feng, Andreas Varvarigos, Ioannis Panitsas, Daniela Fernandez, Jinbiao Wei, Yuwei Guo, Jialin Chen, Ali Maatouk, Leandros Tassiulas, and Rex Ying. TelecomTS: A multi-modal observability dataset for time series and language analysis, 2025.

Vasilii Feofanov, Songkang Wen, Jianfeng Zhang, Lujia Pan, and Ievgen Redko. Mantisv2: Closing the zero-shot gap in time series classification with synthetic data and test-time strategies. *arXiv preprint arXiv:2602.17868*, 2026.

Thibault Formal, Benjamin Piwowarski, and Stéphane Clinchant. SPLADE: Sparse lexical and expansion model for first stage ranking. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, pp. 2288–2292, 2021.

Jean-Yves Franceschi, Aymeric Dieuleveut, and Martin Jaggi. Unsupervised scalable representation learning for multivariate time series. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2019. Curran Associates Inc.

Antoine Guillaume, Christel Vrain, and Wael Elloumi. Random dilated shapelet transform: A new approach for time series shapelets. In *International Conference on Pattern Recognition and Artificial Intelligence (ICPRAI)*, pp. 653–664. Springer, 2022.

Sungwon Han, Seungeon Lee, Meeyoung Cha, Sercan Ö. Arik, and Jinsung Yoon. Retrieval augmented time series forecasting (RAFT). In *International Conference on Machine Learning (ICML)*, 2025. arXiv:2505.04163.

Cenjie Hu, Hang Cui, Zexin Wang, Juncheng Bao, Jingwen Yang, Jingjing Li, Changhua Pei, Dan Pei, and Gaogang Xie. TSRBench: Benchmarking time-series retrieval. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2026. DOI 10.1145/3770855.3817598.

Vincent Jacob, Fei Song, Arnaud Stiegler, Bijan Rad, Yanlei Diao, and Nesime Tatbul. Exathlon: a benchmark for explainable anomaly detection over time series. *Proc. VLDB Endow.*, 14(11): 2613–2626, July 2021. ISSN 2150-8097. doi: 10.14778/3476249.3476307. URL https://doi.org/10.14778/3476249.3476307.

Kalervo Järvelin and Jaana Kekäläinen. Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems (TOIS)*, 20(4):422–446, 2002.

Vladimir Karpukhin, Barlas Oğuz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. Dense passage retrieval for open-domain question answering. In *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*, pp. 6769–6781, 2020.

Omar Khattab and Matei Zaharia. ColBERT: Efficient and effective passage search via contextualized late interaction over BERT. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, pp. 39–48, 2020.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Haojun Li and John Paparrizos. Mufasa: Fast and accurate multivariate time-series clustering. *Proceedings of the ACM on Management of Data*, 4:1–29, 2026.

Zechen Li, Baiyu Chen, Hao Xue, and Flora D. Salim. ZARA: Training-free motion time-series reasoning via evidence-grounded LLM agents. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 14986–15008, San Diego, California, United States, 2026. Association for Computational Linguistics. doi: 10.18653/v1/2026.acl-long.684. URL https://aclanthology.org/2026.acl-long.684/.

Jessica Lin, Eamonn Keogh, Li Wei, and Stefano Lonardi. Experiencing SAX: a novel symbolic representation of time series. *Data Mining and Knowledge Discovery*, 15:107–144, 2007.

Jun Liu, Chaoyun Zhang, Jiaxu Qian, Minghua Ma, Si Qin, Chetan Bansal, Qingwei Lin, Saravan Rajmohan, and Dongmei Zhang. Large language models can deliver accurate and interpretable time series anomaly detection. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2025. DOI 10.1145/3711896.3737239; arXiv:2405.15370 (2024).

Zengkun Liu and Justine Hui. Advancing predictive maintenance: a deep learning approach to sensor and event-log data fusion. *Sensor Review*, 44(5):563–574, 2024.

Matthew Middlehurst, Patrick Schäfer, and Anthony Bagnall. Bake off redux: a review and experimental evaluation of recent time series classification algorithms. *Data Mining and Knowledge Discovery*, 38:1958–2031, 2024. arXiv:2304.13029.

Takehiko Mizoguchi, Yuji Kobayashi, and Yasuhiro Ajiro. Unsupervised retrieval based multivariate time series anomaly detection and diagnosis with deep binary coding models. In *PHM Society Asia-Pacific Conference*, volume 4, 2023.

George B. Moody and Roger G. Mark. The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, 20:45–50, 2001.

Kanghui Ning, Zijie Pan, Yu Liu, Yushan Jiang, James Zhang, Kashif Rasul, Anderson Schneider, Lintao Ma, Yuriy Nevmyvaka, and Dongjin Song. Ts-rag: Retrieval-augmented generation based time series foundation models are stronger zero-shot forecaster. *Advances in Neural Information Processing Systems*, 38:163170–163199, 2025.

John Paparrizos and Luis Gravano. k-shape: Efficient and accurate clustering of time series. In *Proceedings of the 2015 ACM SIGMOD international conference on management of data*, pp. 1855–1870, 2015.

John Paparrizos, Haojun Li, Fan Yang, Kaize Wu, Jens E d'Hondt, and Odysseas Papapetrou. A survey on time-series distance measures. *arXiv preprint arXiv:2412.20574*, 2024.

Luan Pham, Hongyu Zhang, Huong Ha, Flora Salim, and Xiuzhen Zhang. Rcaeval: A benchmark for root cause analysis of microservice systems with telemetry data. In *Companion Proceedings of the ACM on Web Conference 2025*, pp. 777–780, 2025.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

Stephen Robertson and Hugo Zaragoza. The probabilistic relevance framework: BM25 and beyond. *Foundations and Trends in Information Retrieval*, 4(1-2):1–174, 2009.

J. J. Rocchio. Relevance feedback in information retrieval. In *The SMART Retrieval System: Experiments in Automatic Document Processing*, pp. 313–323. Prentice-Hall, 1971.

Bionda Rozin and Daniel Carlos Guimarães Pedronette. A ranked-based framework based on manifold learning for multivariate time series retrieval and classification. *Pattern Recognition Letters*, 202:36–43, 2026. doi: 10.1016/j.patrec.2026.01.026.

Bionda Rozin, Daniel Carlos Guimarães Pedronette, and Ricardo da Silva Torres. Re-ranking and representations for time series retrieval: A comparative study. *IEEE Access*, 13:186103–186120, 2025. doi: 10.1109/ACCESS.2025.3623930.

Hiroaki Sakoe and Seibi Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, 26(1):43–49, 1978.

Patrick Schäfer and Mikael Högqvist. SFA: a symbolic fourier approximation and index for similarity search in high dimensional datasets. In *International Conference on Extending Database Technology (EDBT)*, pp. 516–527, 2012.

Hyeok-Ki Shin, Woomyo Lee, Jeong-Han Yun, and HyoungChun Kim. {HAI} 1.0:{HIL-based} augmented {ICS} security dataset. In *13Th USENIX workshop on cyber security experimentation and test (CSET 20)*, 2020.

Mohammad Shokoohi-Yekta, Bing Hu, Hongxia Jin, Jun Wang, and Eamonn Keogh. Generalizing dtw to the multi-dimensional case requires an adaptive approach. *Data mining and knowledge discovery*, 31(1):1–31, 2017.

Dongjin Song, Ning Xia, Wei Cheng, Haifeng Chen, and Dacheng Tao. Deep r-th root of rank supervised joint binary embedding for multivariate time series retrieval. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2018.

Weiwei Sun, Lingyong Yan, Xinyu Ma, Shuaiqiang Wang, Pengjie Ren, Zhumin Chen, Dawei Yin, and Zhaochun Ren. Is ChatGPT good at search? investigating large language models as re-ranking agents. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 14918–14937, 2023. doi: 10.18653/v1/2023.emnlp-main.923.

Xiaoyu Tao, Yuchong Wu, Mingyue Cheng, Ze Guo, and Tian Gao. AnomaMind: Agentic time series anomaly detection with tool-augmented reasoning. *arXiv preprint arXiv:2602.13807*, 2026.

Nandan Thakur, Nils Reimers, Andreas Rücklé, Abhishek Srivastava, and Iryna Gurevych. BEIR: A heterogeneous benchmark for zero-shot evaluation of information retrieval models. In *Advances in Neural Information Processing Systems (NeurIPS) Datasets and Benchmarks Track*, 2021.

Kutay Tire, Ege Onur Taga, Muhammed Emrullah Ildiz, and Samet Oymak. Retrieval augmented time series forecasting (RAF). In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2026. arXiv:2411.08249; OpenReview:UD76JhLswg.

Sana Tonekaboni, Danny Eytan, and Anna Goldenberg. Unsupervised representation learning for time series with temporal neighborhood coding. In *International Conference on Learning Representations (ICLR)*, 2021.

Ricardo Emanuel Vaz Vargas, Celso José Munaro, Patrick Marques Ciarelli, et al. A realistic and public dataset with rare undesirable real events in oil wells. *Journal of Petroleum Science and Engineering*, 181:106223, 2019. Petrobras 3W dataset, v2.0.0.

Anil Varma. Icarus: design and deployment of a case-based reasoning system for locomotive diagnostics. In *International Conference on Case-Based Reasoning*, pp. 581–596. Springer, 1999.

Miki E. Verma, Robert A. Bridges, Michael D. Iannacone, Samuel C. Hollifield, Pablo Moriano, Steven C. Hespeler, Bill Kay, and Frank L. Combs. A comprehensive guide to CAN IDS data and introduction of the ROAD dataset. *PLOS ONE*, 19:e0296879, 2024.

Ian Wu, Sravan Jayanthi, Vijay Viswanathan, Simon Rosenberg, Sina Pakazad, Tongshuang Wu, and Graham Neubig. Synthetic multimodal question generation. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, 2024.

Stephan Xie, Ben Cohen, Mononito Goswami, Junhong Shen, Emaad Khwaja, Chenghao Liu, David Asker, Othmane Abou-Amal, and Ameet Talwalkar. ARFBench: Benchmarking time series question answering ability for software incident response, 2026.

Zhe Xie, Zeyan Li, Xiao He, Longlong Xu, Xidao Wen, Tieying Zhang, Jianjun Chen, Rui Shi, and Dan Pei. ChatTS: Aligning time series with LLMs via synthetic data for enhanced understanding and reasoning. In *Proceedings of the VLDB Endowment (VLDB)*, 2025. arXiv:2412.03104.

Yiyuan Yang, Zichuan Liu, Lei Song, Kai Ying, Stephen Wang, Joshua Thomas Bamford, Svitlana Vyetrenko, Jiang Bian, and Qingsong Wen. Time-RA: Towards time series reasoning for anomaly diagnosis with LLM feedback. In Maria Liakata, Viviane P. Moreira, Jiajun Zhang, and David Jurgens (eds.), *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 11591–11616, San Diego, California, United States, July 2026. Association for Computational Linguistics. ISBN 979-8-89176-395-1. doi: 10.18653/v1/2026.findings-acl.562. URL https://aclanthology.org/2026.findings-acl.562/.

Michihiro Yasunaga, Armen Aghajanyan, Weijia Shi, Rich James, Jure Leskovec, Percy Liang, Mike Lewis, Luke Zettlemoyer, and Wen-tau Yih. Retrieval-augmented multimodal language modeling. In *International Conference on Machine Learning (ICML)*, 2023.

Chin-Chia Michael Yeh, Nickolas Kavantzas, and Eamonn Keogh. Matrix profile VI: Meaningful multidimensional motif discovery. In *IEEE International Conference on Data Mining (ICDM)*, pp. 565–574, 2017.

Chin-Chia Michael Yeh, Huiyuan Chen, Xin Dai, Yan Zheng, Junpeng Wang, Vivian Lai, Yujie Fan, Audrey Der, Zhongfang Zhuang, Liang Wang, Wei Zhang, and Jeff M. Phillips. An efficient content-based time series retrieval system. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, CIKM '23, pp. 4909–4915, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400701245. doi: 10.1145/3583780.3614655. URL https://doi.org/10.1145/3583780.3614655.

Rongwei Yu, Yong Wang, and Wang Wang. Amad: Active learning-based multivariate time series anomaly detection for large-scale it systems. *Computers & Security*, 137:103603, 2024.

Zhihan Yue, Yujing Wang, Juanyong Duan, Tianmeng Yang, Congrui Huang, Yunhai Tong, and Bixiong Xu. TS2Vec: Towards universal representation of time series. In *AAAI Conference on Artificial Intelligence*, volume 36, pp. 8980–8987, 2022.

Hui Zhao, Jingwei Liu, Wei Dong, Xinya Sun, and Yindong Ji. An improved case-based reasoning method and its application on fault diagnosis of Tennessee Eastman process. *Neurocomputing*, 249:266–276, 2017. doi: 10.1016/j.neucom.2017.04.022. URL https://doi.org/10.1016/j.neucom.2017.04.022.

Zhiwang Zhong, Tianhua Xu, Feng Wang, and Tao Tang. Text case-based reasoning framework for fault diagnosis and predication by cloud computing. *Mathematical Problems in Engineering*, 2018(1):9464971, 2018.

Dixian Zhu, Dongjin Song, Yuncong Chen, Cristian Lumezanu, Wei Cheng, Bo Zong, Jingchao Ni, Takehiko Mizoguchi, Tianbao Yang, and Haifeng Chen. Deep unsupervised binary coding networks for multivariate time series retrieval. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 1403–1411, 2020.

Nicolas Zumarraga, Thomas Kaar, Ning Wang, William Tennien, Alpay Hasanli, Max Rosenblattl, Fan Wu, Kevin Riehl, Maxwell A. Xu, Markus Kreft, Kevin O'Sullivan, Elgar Fleisch, Paul Schmiedmayer, Robert Jakob, and Patrick Langer. TS-Haystack: A multi-scale retrieval benchmark for time series language models. arXiv preprint arXiv:2602.14200, 2026.

A Appendix

Contents

A.1 Dataset Details . . . 16
A.1.1 Curation protocol . . . 16
A.1.2 Query and corpus construction . . . 17
A.1.3 Label construction . . . 17
A.2 Method Suite Formal Definitions . . . 18
A.2.1 Trivial baselines . . . 20
A.2.2 Raw-window distances . . . 20
A.2.3 Embedding-space scorers . . . 20
A.2.4 Symbolic bag-of-words retrievers . . . 21
A.2.5 Foundation-model embedders . . . 21
A.2.6 Supervised classifiers . . . 22
A.2.7 Score fusion . . . 22
A.2.8 Rerankers . . . 22
A.2.9 Language-model rerankers . . . 23
A.3 Evaluation Metric Definitions . . . 23
A.4 Statistical Testing Definitions . . . 24
A.5 Hyperparameters and Compute Budget . . . 24
A.6 Language-Model Reranker Prompts . . . 26
A.7 Additional Results . . . 28
A.7.1 Full base-retriever metric grid . . . 29
A.7.2 Scaling to the full corpus . . . 29
A.7.3 Normal-residual scoring by embedder . . . 29
A.7.4 Choosing the fusion leg and mode . . . 30
A.7.5 Complete fusion grid . . . 30
A.7.6 Reranking a fixed pool . . . 31
A.7.7 Complete reranker grid . . . 31
A.7.8 Composing fusion and reranking . . . 31
A.8 Complete Per-Dataset Results . . . 32

A.1 Dataset Details

This appendix details the twelve datasets of READ-Bench (Section 3.2), covering how they were curated, how queries and corpora are constructed from them, and how each dataset's native annotations are turned into fault-type labels.

A.1.1 Curation protocol

Each dataset enters the benchmark through the same pipeline. We start from a public source, retain the channels and label vocabulary of the originators, and keep the windows of any dataset that ships

pre-windowed, otherwise segmenting each series into fixed-length windows sized to its sampling rate. Each window takes one fault-type label by Equation 2. Missing values are cleaned within each series before windowing, never across a fault boundary, and by one of two source conventions. Most datasets forward-fill a gap with the last valid reading, back-fill a leading gap, and set a wholly-absent channel to zero, since a missing sample means no fresh reading rather than a true zero. Petrobras 3W instead follows its upstream drop convention, coercing out-of-range sensor sentinels to missing and dropping the affected rows without interpolating them.

CTSR is the one univariate dataset we retain, derived from the UCR archive (Dau et al., 2019). Its classes are not anomalies, but diagnostic relevance in the multivariate datasets often concentrates in a subset of channels, so the single-channel case is a close analog and provides a well-established, shape-driven reference point alongside them. Because CTSR has far more relevance classes than the query budget, its corpus is drawn to keep at least ten same-class entries per query, so scores stay stable rather than resting on a single relevant item.

A.1.2 Query and corpus construction

For every dataset we draw queries from the test split and build the retrieval corpus from the train split. Queries are anomalous windows; the corpus is either anomalies alone (the clean setting) or anomalies mixed with a controlled fraction of normal windows drawn from the training split. We report three pollution levels, roughly 0%, 10%, and 20% normal windows, so that robustness to easy negatives can be read off directly. All splits use a fixed seed for reproducibility, and the split unit of Table 4 is always coarser than a window, so no query window, or a near-duplicate of one, enters the corpus it is ranked against. Table 3 lists the resulting corpus and query sizes.

Table 3: Per-dataset composition of READ-Bench. C is the channel count, T the window length, and "classes" the number of non-normal fault types. $|\mathcal{C}|$ is the corpus size at the three pollution levels ($0/10/20\%$ normal windows), $|\mathcal{Q}|$ the query count, and "rel/q" the mean number of same-class corpus entries per query at ρ=0%. "Imb." is the class-imbalance ratio (most frequent over least frequent fault class, at ρ=0%).

Dataset	C	T	Classes	$\lvert\mathcal{C}\rvert$ $(0/10/20\%)$	$\lvert\mathcal{Q}\rvert$	rel/q	Imb.
CTSR (Yeh et al., 2023)	1	512	94	1000/1111/1250	100	10.6	1.2
DAMADICS (Bartyś et al., 2006)	32	256	4	1000/1111/1250	18	308.8	83.9
Exathlon (Jacob et al., 2021)	51	256	6	1000/1111/1250	100	215.9	13.2
HAI (Shin et al., 2020)	79	256	6	305/339/381	28	60.8	10.0
MIT-BIH (Moody & Mark, 2001)	2	256	4	1000/1111/1250	100	336.9	10.7
Petrobras 3W (Vargas et al., 2019)	2–7	256	9	1000/1111/1250	100	146.8	23.7
RATS40K (Yang et al., 2026)	1–9	16–128	20	1000/1111/1250	100	151.2	342.0
RCAEval (Pham et al., 2025)	429	256	5	700/778/875	50	140.0	1.0
ROAD (Verma et al., 2024)	664	256	6	71/79/89	25	22.4	37.0
TelecomTS (Feng et al., 2025)	18	128	11	977/1086/1221	100	117.8	5.3
Tennessee Eastman (Downs & Vogel, 1993)	52	256	20	1000/1111/1250	100	50.0	1.0
Voraus (Brockmann et al., 2023)	130	1024	12	528/587/660	100	68.9	15.6

The datasets span three axes of difficulty that Section 3.1 identifies, and their fault signatures vary widely across domains (Figure 5). Channel count ranges from univariate ($C = 1$ for CTSR) to several hundred channels (ROAD at 664, RCAEval at 429), window length from 16 to 1024 timesteps, and class balance from perfectly uniform (RCAEval and Tennessee Eastman, both balanced by construction) to severely skewed (RATS40K and DAMADICS, whose rarest fault type has one to a few corpus windows against hundreds for the most common). Two datasets are themselves channel-heterogeneous. RATS40K mixes a univariate variant with a multivariate one, so within it C ranges over 1, 3, and 9 channels and T over 16, 32, 64, and 128 timesteps, and Petrobras 3W drops any of its 7 nominal sensors that a well never recorded, so its windows carry 2 to 7 channels. The rarest classes in RATS40K and ROAD contain a single corpus window, and the split protocol keeps such a class in the corpus rather than the query set when it cannot supply both, so every retained query still has a same-class neighbour to retrieve. The fault classes themselves are also visually diverse, both across datasets and among the classes within one dataset (Figure 6), which is what makes fault-type retrieval harder than detecting mere abnormality.

A.1.3 Label construction

The datasets differ in how their native annotations define a fault type, and therefore in how the fault-type label y of Equation 2 is derived. Table 4 records, for each dataset, the source of the fault-type label, the rule that resolves a window's label, and the split unit that keeps queries disjoint from their

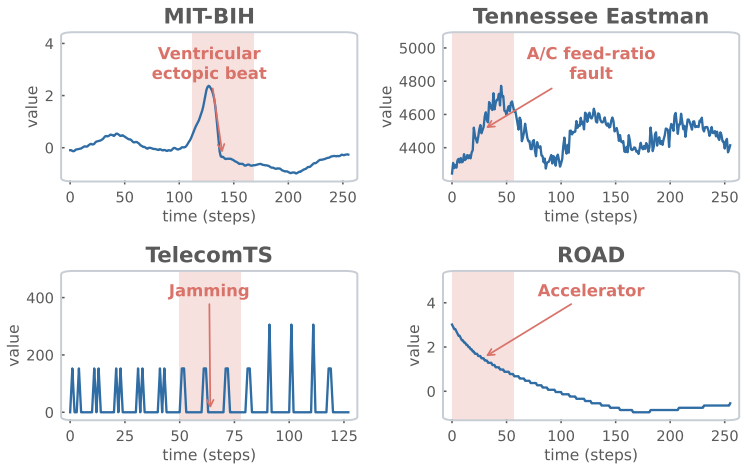


Figure 5: Example fault signatures across four READ-Bench domains, one per panel, with the anomalous span shaded and the most informative channel shown.

corpus. The label rule takes one of three forms. Most datasets take the strict majority of the per-timestep labels within a window ("window majority"), and MIT-BIH is the same rule over a coarser unit, cut around an annotated heartbeat and labeled by the majority beat class among the beats it contains. Datasets whose faults are often shorter than a window, namely the intrusion, control, and disturbance datasets whose anomalies are brief injected events, additionally apply the short-fault relaxation of Section 3.2, labeling a window by a fault that fills a majority of its own extent even without a window majority ("majority or coverage"). The remaining datasets carry a "native label" that their annotations dictate, where Voraus, RATS40K, and TelecomTS inherit a single as-shipped label per recording or window.

A few datasets annotate a fault's brief transient onset with a code separate from the one for the same fault in its steady phase, as Tennessee Eastman, for instance, marks the initial ramp of a disturbance apart from its post-onset regime. Because both codes denote the same underlying fault, we merge each onset code into its steady fault before applying Equation 2, so a window is labeled by the fault itself regardless of whether it captures the onset or the settled phase.

Table 4: How each dataset's fault-type label is derived. "Label source" is the native annotation the label is read from; "label rule" is how a window's fault type is resolved from it, one of "window majority" (a fault holds over half the window), "majority or coverage" (window majority or the short-fault relaxation of Section 3.2), or "native label" (the window inherits its as-shipped annotation); "split unit" is the granularity at which train and test are kept disjoint so that queries are held out from their corpus.

Dataset	Label source	Label rule	Split unit
CTSR	UCR dataset and class label	native label	random hold-out
DAMADICS	actuator-fault schedule	majority or coverage	fault instance
Exathlon	disturbance interval	majority or coverage	trace
HAI	attack schedule	majority or coverage	attack instance
MIT-BIH	per-beat class	window majority	patient recording
Petrobras 3W	per-timestep event class	window majority	instance
RATS40K	per-window anomaly type	native label	native split
RCAEval	per-timestep fault type	window majority	service × fault
ROAD	attack family and variant	majority or coverage	capture
TelecomTS	per-window anomaly type	native label	shuffled split
Tennessee Eastman	per-timestep fault mode	window majority	simulation run
Voraus	per-recording anomaly category	native label	recording

A.2 Method Suite Formal Definitions

We give the exact formulation of every method family of Section 3.3. Let x, y denote two windows, where $x_t \in \mathbb{R}^C$ is the value at timestep t of a C-channel window of length T, and $x^{(c)}$ is channel c. A query is q and a corpus item is $x \in \mathcal{C}$, and every method induces a score $s(q, x)$ that is ranked in descending order to return the top-K.



Figure 6: Representative anomaly classes across the READ-Bench datasets. Each row is one dataset and each cell shows a representative window of one fault type (red) against a normal window (grey).

A.2.1 TRIVIAL BASELINES

Two label-blind floors bound the achievable range. Random assigns each corpus item an independent uniform score $s(q,x) \sim \mathrm{Unif}[0,1]$, giving a chance-level ranking. Majority scores every query by corpus-class prevalence, $s(q,x) = \mathbb{1}[\,y_x = y^\star\,]$ with $y^\star$ the most frequent corpus label, so the ranking places the majority class first and is identical across queries.

A.2.2 RAW-WINDOW DISTANCES

Distances d are turned into scores by $s(q,x) = -d(q,x)$. The lock-step (ED) and symbolic (SAX/SFA-BM25) measures require the query and corpus window to share a length and channel count; when they differ we right-pad the narrower window's channels with zeros up to the shared maximum C and linearly interpolate both windows' lengths onto a shared grid (the maximum observed length, so no measured point is discarded) before scoring. The elastic (DTW) and cross-correlation (SBD) measures align series of unequal length natively and need only the zero-channel-padding step.

Euclidean (ED). A lock-step measure whose multivariate extension is inherently channel-independent (Paparrizos et al., 2024), $d_{\mathrm{ED}}(x,y) = \big(\sum_{c=1}^{C}\sum_{t=1}^{T}(x_t^{(c)} - y_t^{(c)})^2\big)^{1/2}$.

Shape-Based Distance (SBD) (Paparrizos & Gravano, 2015). A sliding measure that finds the best alignment between two time series via normalized cross-correlation (NCC). Following the same channel-dependency taxonomy, the channel-independent variant (SBD-I) applies the univariate NCC to each channel separately; for channel c, $\mathrm{NCC}(x^{(c)}, y^{(c)})[s^{(c)}] = \dfrac{(x^{(c)} \star y^{(c)})[s^{(c)}]}{\|x^{(c)}\|\,\|y^{(c)}\|}$, accelerated in the frequency domain through the Fast Fourier Transform (FFT) and its inverse (IFFT), and SBD-I sums the resulting channel scores, $d_{\text{SBD-I}}(x,y) = \sum_{c=1}^{C}\big(1 - \max_{s^{(c)}} \mathrm{NCC}(x^{(c)}, y^{(c)})[s^{(c)}]\big)$. The channel-dependent variant (SBD-D) instead correlates x, y jointly, as a single $T \times C$ matrix, under one shared lag s; the multivariate cross-correlation is $\mathrm{MCC}(x,y) = \mathrm{IFFT2}\big(\mathrm{FFT2}(x) \odot \overline{\mathrm{FFT2}(y)}\big)$, where $\odot$ denotes the elementwise product and the overline the complex conjugate. Normalized by the norms of the two windows, $\|x\|_F = \big(\sum_{c=1}^{C}\sum_{t=1}^{T}(x_t^{(c)})^2\big)^{1/2}$, this gives $\mathrm{MNCC}(x,y)[s] = \dfrac{\mathrm{MCC}(x,y)[s]}{\|x\|_F\,\|y\|_F}$ and $d_{\text{SBD-D}}(x,y) = 1 - \max_s \mathrm{MNCC}(x,y)[s]$. We use SBD-D throughout due to its better performance compared to SBD-I when evaluated under both classification and clustering settings (Li & Paparrizos, 2026).

Dynamic Time Warping (DTW) (Sakoe & Chiba, 1978). An elastic measure that permits one-to-many point matching, rather than the one-to-one matching of lock-step measures, to achieve local alignment. DTW finds the optimal alignment path and the minimum distance between two series via the accumulated cost $\gamma(i,j) = \delta(i,j) + \min\{\gamma(i-1,j), \gamma(i,j-1), \gamma(i-1,j-1)\}$ and total distance $\gamma(T,T)$, optionally under a Sakoe–Chiba band of radius r. Following the same channel-dependency taxonomy, the independent variant (DTW-I) sets $\delta^{(c)}(i,j) = (x_i^{(c)} - y_j^{(c)})^2$ separately for each channel c, giving each channel its own optimal alignment and per-channel distance $d_{\mathrm{DTW}}(x^{(c)}, y^{(c)}) = \gamma^{(c)}(T,T)$, and sums the channel distances, $d_{\text{DTW-I}}(x,y) = \sum_{c=1}^{C} d_{\mathrm{DTW}}(x^{(c)}, y^{(c)})$ (Shokoohi-Yekta et al., 2017). The dependent variant (DTW-D) instead finds a single alignment shared by every channel; treating x, y as one $T \times C$ matrix, the joint cost $\delta(i,j) = \sum_{c=1}^{C}(x_i^{(c)} - y_j^{(c)})^2$ feeds the same recursion, giving $d_{\text{DTW-D}}(x,y) = \gamma(T,T)$.

A.2.3 EMBEDDING-SPACE SCORERS

Let $\phi(x) \in \mathbb{R}^D$ be a fixed (precomputed) embedding.

Cosine. $s_{\cos}(q,x) = \dfrac{\phi(q)^\top \phi(x)}{\|\phi(q)\|\,\|\phi(x)\|}$.

Normal-Residual (NR). NR treats a fault as a deviation from the embedding expected under normal operation rather than as an absolute location in embedding space. It first ℓ_2-normalizes every embedding, $\hat{\phi}(x) = \phi(x)/\|\phi(x)\|$, and subtracts from each window the embedding it would be expected to have if it were normal, $\mathbb{E}[\hat{\phi}(u) \mid \text{normal}]$. We approximate this conditional expectation by a k-nearest-neighbour estimate over a train-split pool $\mathcal{N}$ of normal windows, letting $\mathcal{N}_k(u) \subseteq \mathcal{N}$ be the k normal windows whose normalized embeddings have the highest cosine similarity to $\hat{\phi}(u)$, and take their mean $\mu_k(u) = \frac{1}{k}\sum_{x\in\mathcal{N}_k(u)} \hat{\phi}(x)$ (we use k=30). Averaging the k nearest normals rather than the single closest is a variance-reduction step, since one nearest normal injects its own embedding noise into every residual, whereas the local mean cancels that noise while remaining specific to the window's operating point. A k-sweep across the 12 datasets finds k=1 consistently weakest and a mid-range k (≈ 30) best. Subtracting this normal reference gives the residual $r(u) = \hat{\phi}(u) - \mu_k(u)$ (the part of a window its expected-if-normal embedding does not explain), which is renormalized and scored by cosine,

$$\hat{r}(u) = \frac{r(u)}{\|r(u)\|}, \qquad s_{\mathrm{NR}}(q, x) = \hat{r}(q)^\top \hat{r}(x).$$

Removing the expected-if-normal component leaves only what departs from normal, so similarity is measured relative to the normal operating envelope rather than absolute position. NR is defined for any fixed embedder ϕ; the per-embedder factorial is given in Table 10.

A.2.4 Symbolic bag-of-words retrievers

Each channel is discretized into a word sequence and scored with Okapi BM25 (Robertson & Zaragoza, 2009). For a query word set with term frequencies and a corpus "document" x of length $|x|$ (number of words) with average length $\overline{|x|}$,

$$\mathrm{BM25}(q, x) = \sum_{w\in q} \mathrm{IDF}(w)\, \frac{f(w, x)\,(k_1 + 1)}{f(w, x) + k_1\big(1 - b + b\,|x|/\overline{|x|}\big)},$$

with $\mathrm{IDF}(w) = \log \frac{N - n(w) + 0.5}{n(w) + 0.5}$, $n(w)$ the number of corpus items containing w, and defaults $k_1 = 1.5$, $b = 0.75$. Scores are averaged over channels.

SAX-BM25. Words are SAX symbols (Lin et al., 2007), a piecewise-aggregate approximation of the z-normalized window followed by Gaussian breakpoints.

SFA-BM25. Words are SFA symbols (Schäfer & Högqvist, 2012), low-frequency DFT coefficients discretized by multiple-coefficient binning (MCB).

A.2.5 Foundation-model embedders

The frozen encoders CHARM, Chronos-2, MantisV2, and TiRex each map a window to $\phi(x)$ and are retrieved by Euclidean distance on the embeddings, $s(q, x) = -d_{\mathrm{ED}}(\phi(q), \phi(x))$, with $s_{\cos}$ giving near-identical rankings (Hu et al., 2026). Because windows vary in both length and channel count across (and within) datasets while each encoder expects a fixed interface, we apply a uniform adaptation along both axes. **Channels.** Each channel $x^{(c)}$ is embedded independently to $\phi(x^{(c)}) \in \mathbb{R}^D$, and the per-channel embeddings are averaged over the channel axis, $\phi(x) = \frac{1}{C}\sum_{c=1}^{C} \phi(x^{(c)})$, so the output dimension is D regardless of C (mean-pooling, rather than concatenation, is what makes the representation channel-count invariant). **Length.** When an encoder constrains the admissible input length, the window is resampled to the nearest admissible length by linearly interpolating each channel onto a shared time grid before encoding, adding interpolated points for shorter windows without discarding any measured point, and never zero-padding the time axis (MantisV2 requires the length to be a multiple of its patch grid; TiRex and Chronos-2 additionally mean-pool over their internal patch axis). This keeps $\phi(x) \in \mathbb{R}^D$ fixed-dimensional regardless of the window's native T and C, so a query and a corpus item are always comparable even when their shapes differ.

MantisV2 (Feofanov et al., 2026). A Transformer encoder pretrained on synthetic data. Each channel is tokenized into a fixed grid of 32 patches, with each patch token formed by concatenating

three descriptors, a convolutional feature of the raw signal, the same feature computed on its first-order difference, and patch-level mean and standard-deviation statistics, linearly projected and layer-normalized. These tokens, with a prepended class token, are processed by 6 Transformer layers, and $\phi(x^{(c)})$ is read from the class token's final hidden state.

Chronos-2 (Ansari et al., 2025). An encoder-only Transformer pretrained for zero-shot forecasting. The series is split into fixed-length patches and embedded by a residual network, then processed by alternating time attention (self-attention along the temporal axis) and group attention layers, the latter aggregating information across related series or channels, $h_i = \sum_{j:\, g_j = g_i} \alpha_{ij}\, v_j$, for in-context learning; $\phi(x)$ is read from the resulting patch embeddings.

TiRex (Auer et al., 2025). An xLSTM-based recurrent sequence model pretrained for zero-shot forecasting. Patches are embedded by a residual block and processed sequentially by stacked sLSTM blocks in place of self-attention, $h_p = \mathrm{xLSTM}(h_{p-1}, u_p)$, with $\phi(x)$ read from the resulting hidden states.

CHARM (Dutta et al., 2026). Trained with a Joint-Embedding Predictive Architecture (JEPA). A contextual temporal convolutional network first extracts per-timestep, per-channel features $\mathbf{X} = \mathrm{TCN}_\theta(x)$ conditioned on textual channel descriptions, and a stack of contextual attention layers fuses the channel and temporal dimensions, $\mathbf{Y}[i,j,:] = \sum_{j'} \alpha_{j,j'}(E_d)\, \mathbf{X}[i,j',:]$, to produce $\phi(x)$.

The per-encoder embedding dimension and context handling are reported in Appendix A.5.

A.2.6 Supervised classifiers

MR-Hydra (Dempster et al., 2023) and RDST (Guillaume et al., 2022) are trained on the labeled corpus to produce a class posterior $g_\theta(\cdot)$. Used as a retrieval scorer, an item x receives the query's posterior mass on its class, $s(q,x) = [g_\theta(q)]_{y_x}$. Because they ultimately emit a label $\hat{y} = \arg\max_y [g_\theta(q)]_y$ rather than a ranking, we also report macro-F1 $= \frac{1}{|\mathcal{Y}|} \sum_{y \in \mathcal{Y}} \mathrm{F1}_y$ as an upper-bound reference. The kernel and shapelet counts and the training regime are given in Appendix A.5.

A.2.7 Score fusion

Given two base scorers A, B over $\mathcal{C}$, we form a fused score in three ways.

Reciprocal-rank fusion (RRF) (Cormack et al., 2009). With $\mathrm{rank}_A(q,x)$ the rank of x under A, $s_{\mathrm{RRF}}(q,x) = \dfrac{1}{k_0 + \mathrm{rank}_A(q,x)} + \dfrac{1}{k_0 + \mathrm{rank}_B(q,x)}$, default $k_0 = 60$.

Weighted sum (WS). $s_{\mathrm{WS}}(q,x) = \alpha\, z(s_A) + (1-\alpha)\, z(s_B)$, where $z(\cdot)$ is per-query z-scoring and $\alpha \in [0,1]$.

Two-step cascade. Retrieve the top-K' by A, then re-rank that shortlist by B.

The fusion grid pairs each of the four FM embedders with each of five shape/symbolic legs under these three modes.

A.2.8 Rerankers

A reranker adjusts the base score on the retrieved pool $\mathcal{P}$ (the top-k neighbours of q). Let y_x be the label of x.

PRF (pseudo-relevance feedback) (Rocchio, 1971). Form the centroid $\bar{c} = \frac{1}{k} \sum_{x \in \mathcal{P}} \phi(x)$ of the top-k and re-score $s'(q,x) = \beta\, s_{\mathrm{base}}(q,x) - (1-\beta)\, \|\phi(x) - \bar{c}\|$.

Purity. Let $\pi(x)$ be the fraction of x's k nearest neighbours sharing its label; $s'(q,x) = \alpha\, z(s_{\mathrm{base}}) + (1-\alpha)\, z(\pi)$.

QPurity (query-conditioned purity). Predict the query label $\hat{y}_q$ by a vote over q's top-k, then credit purity only on candidates with $y_x = \hat{y}_q$.

Majority-Vote. Predict $\hat{y}_q$ from the labels in $\mathcal{P}$ and promote candidates with $y_x = \hat{y}_q$ above the rest.

GPC (Gaussian-process classifier) (Rasmussen & Williams, 2006). Fit a GP classifier on the labeled neighbours in $\mathcal{P}$ and set $s'(q, x) = \Pr(y_x = y_q \mid q, x, \mathcal{P})$, its predicted probability that x shares the query's class. The kernel and PCA settings are given in Appendix A.5.

A.2.9 LANGUAGE-MODEL RERANKERS

Toto-1.0-QA (Cohen et al., 2024), ChatTS (Xie et al., 2025), Claude, and Claude + TSAD (after LLMAD (Liu et al., 2025)) reorder the candidate pool $\mathcal{P}$ by reading the retrieved series directly rather than by evaluating a closed-form score. For all four methods $\mathcal{P}$ is the top-20 pool returned by the CHARM base retriever, so $|\mathcal{P}|$=20, and each candidate is presented with its class label, so these rerankers are label-aware in the sense of Section 3.3 (the query's own label is never shown). Each window is first reduced to its k_{ch}=20 highest-variance channels, and this reduction is applied identically to every representation the model receives (numeric tensor, placeholder token, or rendered image) so that a reduced window is never shown as mismatched modalities of the same series. The methods differ in how a ranking is elicited from $\mathcal{P}$.

The pointwise rerankers, Toto-1.0-QA and ChatTS, issue one request per pair (q, x) with $x \in \mathcal{P}$. Each request returns a scalar relevance judgment, a binary match with an associated confidence for Toto-1.0-QA and a single similarity score on the 0 to 100 scale for ChatTS, and the resulting $|\mathcal{P}|$ scores are sorted in descending order to induce the reranking. A pointwise reranker therefore issues $|\mathcal{P}|$ model calls per query.

The listwise rerankers, Claude and Claude + TSAD, issue a single request per query that presents q together with all of $\mathcal{P}$ and returns the ranked candidate identifiers directly, so one query induces one model call. Claude + TSAD differs from Claude only in its input representation, comparing the per-channel deseasonalized residual of each series rather than the raw series, so that ranking is driven by anomalous deviation rather than shared periodic shape.

Each model's output is turned into the same (q, x) score matrix that every other method produces, so the LLM rerankers are scored by the identical metrics of Appendix A.3. A reranker promotes the candidates the model returned, in the model's order, above the base retriever's remaining ordering of the pool, leaving corpus items outside $\mathcal{P}$ at their base scores. The no-retrieval classification baseline instead reads a per-class confidence and assigns it to every corpus item of that class, so an unrecognized class contributes zero, matching how the supervised classifiers are scored.

The serving, decoding, and series-rendering settings for all four methods are given in Appendix A.5, and the exact prompts in Appendix A.6.

A.3 EVALUATION METRIC DEFINITIONS

Let q be a query, $\text{rel}(q, x) \in \{0, 1\}$ indicate whether corpus item x is relevant to q (Section 3.1), and let $x_1, \ldots, x_K$ denote the top-K items returned for q, ranked by descending score.

Precision at K (P@K).

$$\text{P@}K(q) = \frac{1}{K} \sum_{i=1}^{K} \text{rel}(q, x_i).$$

Hit Rate at K (HR@K).

$$\text{HR@}K(q) = \mathbb{1}\left[\sum_{i=1}^{K} \text{rel}(q, x_i) > 0\right],$$

i.e. whether at least one relevant item appears in the top K.

Normalized Discounted Cumulative Gain (NDCG@K) (Järvelin & Kekäläinen, 2002).

$$\mathrm{DCG@}K(q) = \sum_{i=1}^{K} \frac{\mathrm{rel}(q, x_i)}{\log_2(i+1)}, \qquad \mathrm{NDCG@}K(q) = \frac{\mathrm{DCG@}K(q)}{\mathrm{IDCG@}K(q)},$$

where $\mathrm{IDCG@}K(q)$ is $\mathrm{DCG@}K$ under the ideal ranking (all relevant items first). Unlike P@K/HR@K, NDCG@K is sensitive to the order of retrieved items, not just set membership.

All three metrics are averaged over queries (and, where noted, over datasets) to produce the per-method scores compared in Section 4.

A.4 Statistical Testing Definitions

To compare methods beyond raw score differences we use a Friedman test with a Nemenyi post-hoc test, blocking on the 12 datasets (Demšar, 2006). Let m be the number of methods compared and $n = 12$ the number of datasets.

Friedman test. For each dataset j, rank the m methods by their metric value, giving rank r_i^j to method i. The Friedman statistic is

$$\chi_F^2 = \frac{12n}{m(m+1)} \left[\sum_{i=1}^{m} \bar{r}_i^2 - \frac{m(m+1)^2}{4} \right], \qquad \bar{r}_i = \frac{1}{n} \sum_{j=1}^{n} r_i^j,$$

tested against a χ^2 distribution with $m-1$ degrees of freedom; rejection indicates that at least one method differs significantly from the others.

Nemenyi post-hoc test. Following a significant Friedman result, two methods i, i' are declared significantly different when their mean ranks differ by more than the critical difference,

$$|\bar{r}_i - \bar{r}_{i'}| > q_\alpha \sqrt{\frac{m(m+1)}{6n}},$$

where q_α is the critical value of the studentized range distribution at family-wise significance level $\alpha = 0.05$.

A.5 Hyperparameters and Compute Budget

This appendix records the exact search space or fixed setting and the final chosen value for every method family of Section 3.3, together with the compute budget. Values fixed by a pretrained checkpoint or inherited from a library default are marked as such in the "source" column, so that our own choices are distinguishable from upstream defaults. All retrieval metrics are computed at cutoffs $K \in \{1, 3, 5, 10, 20\}$.

Distance and symbolic baselines. Euclidean distance and SBD carry no tunable settings. SBD is run in its dependent (SBD-D) form, and DTW is banded (Sakoe–Chiba) and defaults to its independent (DTW-I) channel mode. The symbolic retrievers discretize each channel with a sliding window and score the resulting words with Okapi BM25, SAX with piecewise-aggregate approximation (PAA) over Gaussian breakpoints and SFA with discrete-Fourier-transform (DFT) coefficients under equi-depth multiple-coefficient binning (MCB). Table 5 lists the tunable settings.

Table 5: Distance and symbolic baseline settings. T is the window length, so the DTW band scales with the window. The symbolic vocabulary size is alphabet$^{\text{word length}}$.

Method	Hyperparameter	Value
DTW	Sakoe–Chiba band radius	$\max(1, \mathrm{round}(0.1\,T))$
NR	nearest-normal count k	30
SAX-BM25	window / stride	96/8
	PAA word length	6
	alphabet size	4
SFA-BM25	window / stride	128/8
	word length (DFT coefficients)	8
	alphabet size	3
BM25	k_1 / b	1.5/0.75

Foundation-model embedders. Each embedder is a frozen pretrained encoder, so its embedding dimension and context handling are fixed by the released checkpoint rather than tuned here. We embed each channel independently and mean-pool over channels, so the representation is channel-count invariant, and we set only the inference batch size at the call site. Table 6 records the per-encoder settings.

Table 6: Foundation-model embedder settings. The embedding dimension and any patch-grid constraint are fixed by the released checkpoint; pooling and inference batch size are set by us.

Encoder	Hyperparameter	Value
CHARM	embedding dimension	384
	pooling / batch	mean / 32
Chronos-2	embedding dimension	768
	pooling / batch	mean / 32
MantisV2	per-channel width	256
	patch grid	32 patches (T a multiple of 32, $T \geq 64$)
	pooling / batch	mean / 32
TiRex	per-channel width	12×512
	pooling / batch	mean / 512

Supervised classifiers. MR-Hydra and RDST are the aeon implementations, run at their library defaults for the kernel, shapelet, and ensemble counts. MR-Hydra averages the class posteriors of its MultiRocket and Hydra members in equal weight.

Fusion and reranking. Fusion pairs one embedder with one shape or symbolic leg, and each reranker adjusts a base score on a retrieved pool. Table 7 records the constants.

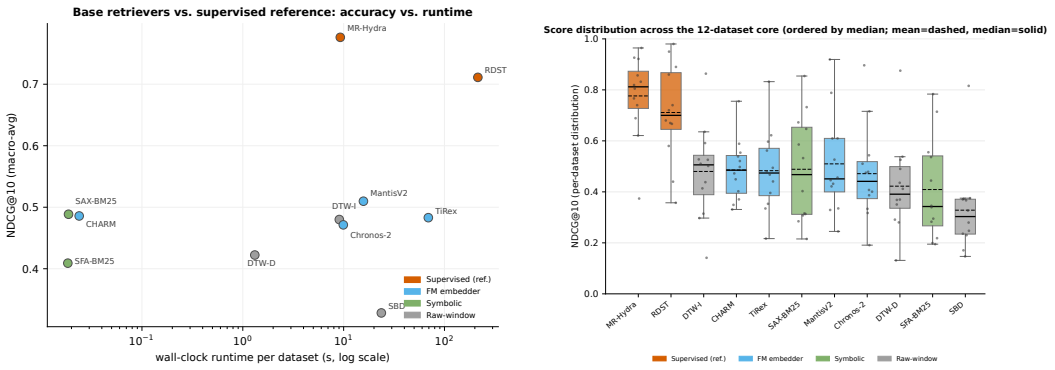
Table 7: Fusion and reranking settings. The weighted sum and the purity-based rerankers combine per-query z-scored terms with weight α.

Method	Hyperparameter	Value
RRF	fusion constant k_0	60
Weighted sum	weight α	0.5
Two-step cascade	shortlist size K'	100
PRF	neighbourhood k / weight β	10/0.5
Purity	neighbourhood k / weight α	10/0.5
Majority-Vote	neighbourhood k	10
QPurity	neighbourhood k / vote k / weight α	10/10/0.5
GPC	fit-window K	20
	PCA variance retained	0.95
	kernel	L1-norm, length scale 1.0 (fixed)

Language-model rerankers. The method mechanics of the four language-model rerankers, namely the shared top-20 CHARM pool, the pointwise versus listwise scoring, and the highest-variance channel cap, are given in Appendix A.2, and their exact prompts, including the no-retrieval classification baseline of Table 2, in Appendix A.6. Toto-1.0-QA[2] and ChatTS[3] run locally from their released checkpoints and decode greedily (no sampling), up to 2000 and 512 new tokens respectively. Claude and Claude + TSAD are Claude Sonnet 5 (1M-context) called through the AWS Bedrock Messages API at the default temperature with extended thinking disabled and up to 2048 output tokens, retried up to four times under exponential backoff on a throttling or empty response. Each retrieved window is rendered for the model as one plot per channel and a numeric series, and Claude + TSAD substitutes the per-channel deseasonalized residual (following LLM-TSAD)[4] for the raw series.

Compute and hardware. All experiments ran on a single node with two NVIDIA A100-SXM4-80GB GPUs, an Intel Xeon Platinum 8275CL (96 cores) and 1.1 TiB of RAM. The foundation-model embedders, supervised classifiers, and the two locally-hosted language-model rerankers (Toto-1.0-QA, ChatTS) use the GPU, while the distance, symbolic, and fusion or reranking stages

[2] https://huggingface.co/Datadog/Toto-1.0-QA-Experimental
[3] https://huggingface.co/bytedance-research/ChatTS-14B
[4] https://github.com/junwoopark92/LLM-TSAD



(a) **(b)**

Figure 7: Base retrievers and the supervised reference over the 12 datasets, both at ρ=0% and family-coloured. (a) Retrieval accuracy against wall-clock scoring time (log scale, median per-dataset runtime). The base retrievers span two orders of magnitude in scoring time yet cluster in a narrow accuracy band well below the supervised reference, so among label-free methods the choice of a classical distance, a symbolic retriever, or a pretrained embedder trades more compute than retrieval quality. (b) The matching per-dataset NDCG@10 distribution, one box per method ordered by median (dashed line is the mean, solid the median, dots the 12 per-dataset scores), where the label-free retrievers overlap heavily and remain below the supervised reference.

are CPU-only; Claude and Claude + TSAD instead call the hosted AWS Bedrock API. The scoring cost of the retrieval stage itself is modest and varies widely across method families, spanning about two orders of magnitude, as Figure 7a shows for the base retrievers.

A.6 Language-Model Reranker Prompts

This appendix gives the exact system and question prompts sent to each language-model method, verbatim from the run scripts, for both the reranking task and the no-retrieval classification baseline, whose results are reported in Table 2. Every prompt below is instantiated per query; {...} placeholders are filled in at request time (a topk count, the option list, etc.). Every reranking prompt appends one candidate label line per candidate, shown inline in each prompt below, disclaiming that class membership implies similarity so the model treats the label as neutral metadata; the query's own label is never shown.

Toto-1.0-QA reranking. One request per (query, candidate) pair; the model is shown two entities (the QUERY, then this one CANDIDATE) as two separate images plus a packed numeric tensor, and returns a match/confidence judgment that is converted to a score (confidence if the match is "yes", else 100 − confidence) and sorted across the pool's 20 candidates to obtain the ranking.

Toto-1.0-QA reranking prompt

```
SYSTEM:
You are an expert observability anomaly analyst comparing time-series
patterns for retrieval reranking.

Input:
- Images: one plot per time series shown, in order -- the QUERY first,
  then each CANDIDATE.
- Time Series: the same series, embedded per channel.
- Question: the ranking instruction below.

QUESTION (per candidate):
In the following time-series, does the anomaly in the QUERY series
correlate with the anomaly in the CANDIDATE series, if anomalies
exist? Respond with ONLY a JSON object of the form
{"match": "yes"|"no", "confidence": <integer 0-100>} -- confidence
is how sure you are of your match/no-match judgment. No explanation.
```

```
(Appended to the CANDIDATE:)
For reference, this CANDIDATE is drawn from the class: {label}. This
class membership does NOT indicate similarity to the QUERY -- base
your similarity judgment ONLY on the actual time-series pattern.
```

When a dataset provides one, a short prose description of its sensors and labels is appended to the system prompt under a "Dataset context" heading, and a per-request time-series metadata string (channel names and timestamps) is appended to the question when non-empty.

ChatTS reranking. Same shape as Toto-1.0-QA (one request per (query, candidate) pair, scores sorted over the pool), but ChatTS has no vision input; each channel of each entity becomes one `<ts></ts>` placeholder token in the text, paired with its raw numeric series fed to the model's own time-series encoder. It rates an anchored 0–100 similarity score.

ChatTS reranking prompt

```
SYSTEM:
You are an expert observability anomaly analyst comparing time-series
patterns for retrieval reranking.

QUESTION:
Rate how closely the CANDIDATE time series resembles the QUERY time
series, considering shape, timing, magnitude, and duration of any
anomalous pattern. Use the FULL 0-100 scale -- do not default to a
high score just because the candidate was retrieved as a plausible
match; most candidates in a retrieval pool are only partial matches
and should score accordingly. Scale: 0-20 = no resemblance,
21-40 = weak/superficial resemblance only, 41-60 = partial
resemblance (some but not all features match), 61-80 = strong
resemblance (most features match), 81-100 = near-identical pattern.
Respond with ONLY a JSON object of the form {"similarity": <integer
0-100>}. No explanation.

(Appended to the CANDIDATE:)
For reference, this CANDIDATE is drawn from the class: {label}. This
class membership does NOT indicate similarity to the QUERY -- base
your similarity judgment ONLY on the actual time-series pattern.
```

Claude reranking. One request per query, with all 20 candidates shown together (each as its own image plus a serialized numeric text block) and ranked directly.

Claude reranking prompt

```
SYSTEM:
You are an expert observability anomaly analyst comparing time-series
patterns for retrieval reranking. You will be shown one plotted image
and one text block of numeric values per series -- the QUERY first,
then each CANDIDATE, in order.

QUESTION:
The first image and text block is the QUERY time series. Every
other image/text block is a CANDIDATE time series, labeled with its
candidate id (e.g. "sample_00007"). Rank the up to {topk} candidates
whose pattern most closely resembles/correlates with the query, best
match first. Respond with ONLY a JSON object of the form
{"ranked_ids": ["<candidate_id>", ...]} using the exact candidate id
strings given, most relevant first, no explanation.

(Appended to each CANDIDATE:)
For reference, this CANDIDATE is drawn from the class: {label}. This
class membership does NOT indicate similarity to the QUERY -- base
your similarity judgment ONLY on the actual time-series pattern.
```

Claude + TSAD reranking. Identical request shape to Claude above, except every series (query and every candidate) is first deseasonalized per channel (autocorrelation peak-period detection, clamped to a minimum period of 8, then an additive seasonal decomposition; channels where this is degenerate fall back to the raw values) – both the image and the numeric text show this residual, not the raw series.

Claude + TSAD reranking prompt

```
SYSTEM:
You are an expert observability anomaly analyst comparing
deseasonalized time-series residual patterns for retrieval
reranking. You will be shown one plotted image and one text block
of numeric values per series -- the QUERY first, then each
CANDIDATE, in order.

QUESTION:
The first image and text block is the QUERY time series. Every
other image/text block is a CANDIDATE time series, labeled with its
candidate id (e.g. "sample_00007"). Each series has had its
seasonal component removed where detectable (deseasonalized
residual, noted per-channel) so you are comparing the anomalous
deviation pattern, not the raw periodic shape. Rank the up to
{topk} candidates whose pattern most closely resembles/correlates
with the query, best match first. Respond with ONLY a JSON object
of the form {"ranked_ids": ["<candidate_id>", ...]} using the exact
candidate id strings given, most relevant first, no explanation.

(Appended to each CANDIDATE:)
For reference, this CANDIDATE is drawn from the class: {label}. This
class membership does NOT indicate similarity to the QUERY -- base
your similarity judgment ONLY on the actual time-series pattern.
```

No-retrieval classification. Rather than reranking a retrieved pool, the query is compared directly against one representative labeled EXAMPLE per class in the dataset's fault taxonomy (no retrieval step), and the model rates its confidence that the query matches each. The wording is shared near-verbatim across all four models, differing only in whether it refers to "image/series" or "entry" for the respective input modality.

No-retrieval classification prompt

```
SYSTEM:
You are an expert observability anomaly analyst rating how well a
query time series matches each of several labeled example patterns.
You will be shown one plotted image and one text block of numeric
values per series -- the QUERY first, then one labeled EXAMPLE per
category.

QUESTION:
Rate your confidence (0-100) that the QUERY belongs to the SAME
category as each numbered EXAMPLE below, based on how closely its
pattern matches the QUERY's:
{option_lines}

Use the full 0-100 scale -- most examples will only partially match,
so most scores should NOT be high. Respond with ONLY a JSON object
of the form {"scores": {"1": <0-100>, "2": <0-100>, ...}} covering
every number above. No explanation.
```

A.7 Additional Results

This appendix collects the complete result tables behind Section 4. They cover the full set of metrics for the base retrievers, how each method behaves as the corpus is polluted, normal-residual scoring on every embedder, the complete fusion and reranking results, their combination, and a check

that the findings hold on the uncapped corpus. All numbers are NDCG@10 unless stated, macro-averaged over the 12 datasets, at pollution levels ρ=0%/ $\approx$ 10%/ $\approx$ 20% where shown. In every table the best value in each column is bold and the second best is underlined.

A.7.1 Full base-retriever metric grid

Table 8 extends the headline NDCG@10 of Table 1 to the full metric set at ρ=0%, adding P@1, P@5, HR@10, NDCG@5, NDCG@20, and macro-F1, as well as DTW-D. Across every metric the label-free retrievers stay close and below the supervised reference, so the choice among them does not change retrieval quality on any single metric. σ, the standard deviation of NDCG@10 across the 12 datasets, is smallest for CHARM, so it is also the most consistent across domains.

Table 8: Full base-retriever metric grid at ρ=0%, macro-averaged over the 12 datasets. σ is the standard deviation of NDCG@10 across datasets. The supervised rows are a classification reference rather than retrieval systems.

Method	Family	P@1	P@5	HR@10	NDCG@5	NDCG@10	NDCG@20	macro-F1	σ
Random	Trivial floor	0.172	0.183	0.699	0.182	0.189	0.192	0.125	0.10
Majority		0.246	0.246	0.246	0.246	0.246	0.247	0.060	0.13
SBD-D	Raw-window distance	0.353	0.340	0.739	0.345	0.328	0.317	0.330	0.17
DTW-I		0.568	0.492	0.857	0.510	0.480	0.470	0.464	0.18
DTW-D		0.507	0.434	0.787	0.451	0.422	0.412	0.435	0.18
SAX-BM25	Symbolic	0.553	0.498	0.897	0.512	0.489	0.466	0.503	0.20
SFA-BM25		0.490	0.422	0.862	0.440	0.409	0.390	0.428	0.19
Chronos-2	FM embedder	0.530	0.486	0.885	0.497	0.471	0.454	0.445	0.18
MantisV2		**0.583**	**0.525**	0.878	**0.538**	**0.510**	**0.489**	**0.518**	0.19
TiRex		0.579	0.491	**0.905**	0.510	0.483	0.466	0.473	0.15
CHARM		0.550	0.505	0.897	0.515	0.486	0.471	0.481	**0.11**
MR-Hydra	Supervised reference	0.759	0.771	0.850	0.772	0.776	0.785	0.666	0.15
RDST		0.711	0.711	0.711	0.711	0.711	0.712	0.617	0.18

A.7.2 Scaling to the full corpus

The main-text results cap each corpus to a common size (Table 3) for uniformity and compute. To check that the closeness of the label-free retrievers (Section 4.1) is not an artifact of that cap, we rerun them on the uncapped corpus of each dataset, up to over a million windows. Five datasets are already at or below the cap, so for them the uncapped corpus is the one already reported; the other seven scale up substantially. The elastic distances (DTW-I, DTW-D), the shape-based distance (SBD-D), the symbolic BM25 retrievers, and the supervised classifiers do not scale to the largest corpora and are omitted; the table compares the foundation-model embedders (which do scale) against their capped scores across all 12 datasets. Absolute scores drop as the corpus grows, as expected with far more candidates to rank against, and the structure holds, with the label-free embedders still close and no single retriever uniformly best.

Table 9: Capped versus full-corpus NDCG@10, macro-averaged over the 12 datasets (five are already at or below the cap, so their full corpus is the one already reported). The elastic distances, the symbolic BM25 retrievers, and the supervised reference do not scale to the largest corpora and are omitted. In each column the best value is bold and the second best underlined.

Method	Capped	Full	Δ
CHARM	0.486	0.437	-0.049
CHARM + NR	**0.523**	0.442	-0.081
Chronos-2	0.471	0.437	-0.034
MantisV2	0.510	**0.447**	-0.063
TiRex	0.483	0.434	-0.049

A.7.3 Normal-residual scoring by embedder

Table 10 applies NR scoring to each embedder and compares it against ED at every pollution level. NR improves pollution robustness on every embedder it wraps, roughly halving the degradation in each case, and improves clean accuracy on all of them except TiRex, whose clean score is essentially unchanged ($0.483 \rightarrow 0.481$); MantisV2 + NR is the strongest single base retriever, clean and under pollution.

Table 10: NR scoring versus ED on each embedder, NDCG@10 at each pollution level ρ, where Δ is the drop from ρ=0% to ρ=20% and a smaller-magnitude Δ means more robust.

Embedder	Scoring	ρ=0%	ρ=10%	ρ=20%	Δ
CHARM	ED	0.486	0.461	0.437	−0.049
	NR	0.523	0.511	0.502	−0.021
MantisV2	ED	0.510	0.474	0.454	−0.056
	NR	**0.555**	0.539	0.527	−0.028
Chronos-2	ED	0.471	0.455	0.432	−0.039
	NR	0.474	0.467	0.452	−0.022
TiRex	ED	0.483	0.449	0.433	−0.050
	NR	0.481	0.476	0.451	−0.029

A.7.4 Choosing the fusion leg and mode

Fusion involves two choices, namely which complementary leg to pair with the embedder and which rule combines the two rankings. Table 11 sweeps both at once, crossing every leg with every mode for each embedder. Across all seven embedder blocks the best cell is the DTW-I leg fused by RRF, so the joint optimum is the same regardless of embedder. RRF is also the best mode on every leg except SBD-D, which is strongest under weighted sum in each block but never competes for the top cell, and DTW-I is the best leg under RRF throughout. RRF over a DTW-I leg is therefore the fusion recipe used in the rest of the paper. The same grid across pollution levels is in Appendix A.7.5.

Table 11: Choosing the fusion leg and mode. NDCG@10 at ρ=0%, macro-averaged over the 12 datasets, for every complementary leg (rows) crossed with every fusion mode (columns), per embedder. Modes are RRF, WS (α=0.5), and 2-step cascade (shortlist 100). Marks are per embedder block, and the DTW-I leg under RRF is the joint optimum in every block.

Embedder	Leg	RRF	WS	2-step
CHARM	DTW-I	**0.536**	0.521	0.498
	DTW-D	0.492	0.477	0.444
	SBD-D	0.445	0.454	0.397
	SAX-BM25	0.519	0.500	0.497
	SFA-BM25	0.485	0.448	0.456
CHARM + NR	DTW-I	**0.575**	0.543	0.538
	DTW-D	0.543	0.528	0.500
	SBD-D	0.476	0.520	0.438
	SAX-BM25	0.551	0.543	0.537
	SFA-BM25	0.510	0.493	0.491
MantisV2	DTW-I	**0.532**	0.509	0.509
	DTW-D	0.507	0.488	0.471
	SBD-D	0.462	0.490	0.410
	SAX-BM25	0.531	0.495	0.502
	SFA-BM25	0.496	0.448	0.465
MantisV2 + NR	DTW-I	**0.590**	0.572	0.561
	DTW-D	0.562	0.556	0.519
	SBD-D	0.493	0.541	0.440
	SAX-BM25	0.582	0.541	0.540
	SFA-BM25	0.543	0.493	0.498
Chronos-2 + NR	DTW-I	**0.551**	0.514	0.531
	DTW-D	0.528	0.500	0.497
	SBD-D	0.442	0.486	0.413
	SAX-BM25	0.538	0.533	0.529
	SFA-BM25	0.499	0.485	0.482
TiRex	DTW-I	**0.520**	0.501	0.484
	DTW-D	0.486	0.470	0.444
	SBD-D	0.445	0.463	0.399
	SAX-BM25	0.511	0.507	0.506
	SFA-BM25	0.479	0.463	0.471
TiRex + NR	DTW-I	**0.554**	0.520	0.540
	DTW-D	0.521	0.504	0.497
	SBD-D	0.442	0.492	0.413
	SAX-BM25	0.530	0.528	0.520
	SFA-BM25	0.491	0.480	0.472

A.7.5 Complete fusion grid

Table 12 gives the full fusion grid, every embedder (ED and NR) crossed with every shape or symbolic leg, at all three pollution levels. The best leg is DTW-I, MantisV2 + NR is the strongest fused pool, and the gains are consistent across embedders rather than specific to one.

Table 12: Complete fusion grid. RRF of each embedder (rows) with each shape or symbolic leg (column groups), NDCG@10 at pollution levels ρ=0/≈10/≈20%, macro-averaged over the 12 datasets. Both the ED and NR variant of each embedder are shown, so the fusion gains are not specific to a single embedder. Marks are on the clean (ρ=0%) sub-column of each leg; DTW-I is the best complementary leg and MantisV2 + NR the strongest fused pool.

	| DTW-I			| DTW-D			| SBD-D			| SAX-BM25			| SFA-BM25		
Embedder	ρ=0%	10%	20%	0%	10%	20%	0%	10%	20%	0%	10%	20%	0%	10%	20%
CHARM	0.536	0.493	0.453	0.492	0.461	0.425	0.445	0.426	0.397	0.519	0.492	0.459	0.485	0.459	0.430
CHARM + NR	0.575	0.549	0.514	0.543	0.524	0.491	0.476	0.456	0.432	0.551	0.540	0.521	0.510	0.503	0.480
MantisV2	0.532	0.490	0.452	0.507	0.468	0.432	0.462	0.440	0.408	0.531	0.500	0.472	0.496	0.466	0.447
MantisV2 + NR	**0.590**	0.556	0.534	**0.562**	0.533	0.509	**0.493**	0.475	0.454	**0.582**	0.560	0.540	**0.543**	0.522	0.507
Chronos-2 + NR	0.551	0.521	0.498	0.528	0.508	0.485	0.442	0.422	0.404	0.538	0.527	0.510	0.499	0.490	0.470
TiRex	0.520	0.487	0.447	0.486	0.455	0.418	0.445	0.426	0.392	0.511	0.490	0.456	0.479	0.456	0.430
TiRex + NR	0.554	0.524	0.495	0.521	0.497	0.470	0.442	0.425	0.399	0.530	0.522	0.496	0.491	0.478	0.456

A.7.6 RERANKING A FIXED POOL

Table 13 reports each reranker on each pool at ρ=0%. The label-aware GPC reranker gives the largest gain on most pools, with Majority-Vote (MajVote) ahead on the CHARM + NR and TiRex + NR pools; the label-free rerankers give little or no gain, since PRF moves the score only marginally and Purity hurts.

Table 13: Reranking a fixed pool, NDCG@10. Label-free rerankers are PRF and Purity; label-aware ones are GPC, MajVote, and QPurity. Language-model rerankers in Table 2.

Pool	Base	+GPC	+MajVote	+QPurity	+PRF	+Purity
CHARM	0.486	**0.556**	0.538	0.511	0.455	0.386
CHARM + NR	0.523	0.608	**0.612**	0.558	0.528	0.461
DTW-I	0.480	**0.573**	0.525	0.500	0.483	0.342
MantisV2	0.510	**0.604**	0.571	0.528	0.501	0.445
MantisV2 + NR	0.555	**0.645**	0.616	0.562	0.511	0.489
Chronos-2 + NR	0.474	**0.582**	0.545	0.512	0.463	0.432
TiRex	0.483	**0.528**	0.517	0.485	0.463	0.394
TiRex + NR	0.481	0.541	**0.568**	0.518	0.445	0.426

A.7.7 COMPLETE RERANKER GRID

Table 14 extends the reranker comparison to every pool (ED and NR) at all three pollution levels, confirming that the GPC gain holds across embedders and pollution and is not specific to the CHARM-based NR pool. NR's advantage also survives reranking, since MantisV2 + NR + GPC (0.645) exceeds the MantisV2 + ED + GPC reference (0.604), so the strongest reranked pool is itself an NR variant.

Table 14: Complete reranker grid. Each reranker on each pool, NDCG@10 at pollution levels ρ=0/≈10/≈20%, macro-averaged over the 12 datasets. Pools include both the ED and NR variant of each embedder, so the reranking gains are shown to hold across embedders and are not specific to the CHARM-based NR. The label-free rerankers are PRF and Purity; the label-aware ones are GPC, MajVote, and QPurity. Marks are on the clean (ρ=0%) sub-column of each reranker.

	+GPC			+MajVote			+QPurity			+PRF			+Purity		
Pool	ρ=0%	10%	20%	0%	10%	20%	0%	10%	20%	0%	10%	20%	0%	10%	20%
CHARM	0.556	0.562	0.517	0.538	0.536	0.500	0.511	0.513	0.481	0.455	0.443	0.407	0.386	0.383	0.332
CHARM + NR	0.608	0.608	0.578	0.612	0.614	0.579	0.558	0.566	0.535	**0.528**	0.524	0.499	0.461	0.464	0.427
MantisV2	0.604	0.556	0.532	0.571	0.548	0.522	0.528	0.507	0.486	0.501	0.470	0.450	0.445	0.431	0.413
MantisV2 + NR	**0.645**	0.613	0.582	**0.616**	0.610	0.589	**0.562**	0.549	0.538	0.511	0.498	0.490	**0.489**	0.478	0.458
Chronos-2 + NR	0.582	0.573	0.563	0.545	0.546	0.534	0.512	0.508	0.492	0.463	0.464	0.451	0.432	0.440	0.417
TiRex	0.528	0.514	0.486	0.517	0.513	0.456	0.485	0.488	0.440	0.463	0.450	0.411	0.394	0.400	0.359
TiRex + NR	0.541	0.525	0.509	0.568	0.566	0.551	0.518	0.523	0.498	0.445	0.447	0.424	0.426	0.427	0.392
DTW-I	0.573	0.517	0.464	0.525	0.468	0.423	0.500	0.503	0.416	0.483	0.427	0.392	0.342	0.324	0.286

A.7.8 COMPOSING FUSION AND RERANKING

Table 15 composes fusion and reranking, reporting the fused pool and the same pool reranked with GPC across pollution levels. The two additions stack close to additively, GPC adds about the same margin to a fused pool as it does to an unfused one, and CHARM + NR | DTW-I + GPC is the strongest system at every pollution level. Normal-residual scoring is what lifts these stacks above their plain-embedding counterparts (e.g. CHARM | DTW-I + GPC).

Table 15: Composing fusion and reranking. NDCG@10 of the fused pool (fusion only, at ρ=0%) and of the same pool reranked with GPC across pollution levels ρ, macro-averaged over the 12 datasets. In each column the best value is bold and the second best underlined.

	Fusion	Fusion + GPC		
Fusion	ρ=0%	ρ=0%	ρ=10%	ρ=20%
CHARM + NR | DTW-I	<u>0.575</u>	**0.687**	**0.640**	**0.619**
MantisV2 + NR | DTW-I	**0.590**	<u>0.665</u>	<u>0.641</u>	<u>0.604</u>
CHARM | DTW-I	0.536	0.631	0.582	0.535
MantisV2 | DTW-I	0.532	0.609	0.564	0.552

A.8 Complete Per-Dataset Results

This appendix reports the full metric grid for every non-fusion method in READ-Bench, with one table per method listing all 12 datasets (and their macro-average) across the three pollution levels. Metrics are precision (P), hit rate (HR), and NDCG at cutoffs $\{1, 5, 10, 20\}$, plus macro-F1. In every table the column headers abbreviate N=NDCG and mF1=macro-F1, and the *macro* row is the macro-average over the 12 datasets. Fusion combinations are reported in Appendix A.7.

Table 16: Full metric grid for **Random**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.000	0.008	0.012	0.000	0.040	0.120	0.000	0.005	0.009	0.014	0.000
DAMADICS	0.333	0.300	0.317	0.333	0.611	0.667	0.333	0.299	0.309	0.325	0.170
Exathlon	0.180	0.224	0.219	0.180	0.740	0.890	0.180	0.221	0.219	0.211	0.134
HAI	0.179	0.214	0.225	0.179	0.679	0.893	0.179	0.203	0.215	0.199	0.163
MIT-BIH	0.300	0.340	0.364	0.300	0.840	0.960	0.300	0.334	0.353	0.351	0.214
Petrobras 3W	0.120	0.140	0.151	0.120	0.470	0.730	0.120	0.138	0.146	0.149	0.130
RATS40K	0.190	0.152	0.148	0.190	0.450	0.550	0.190	0.156	0.152	0.155	0.048
RCAEval	0.160	0.212	0.206	0.160	0.700	0.840	0.160	0.206	0.204	0.202	0.231
ROAD	0.280	0.280	0.328	0.280	0.680	0.960	0.280	0.295	0.350	0.385	0.147
TelecomTS	0.110	0.112	0.114	0.110	0.420	0.650	0.110	0.113	0.114	0.115	0.110
Tennessee Eastman	0.040	0.058	0.051	0.040	0.250	0.400	0.040	0.054	0.050	0.052	0.068
Voraus	0.170	0.152	0.146	0.170	0.530	0.730	0.170	0.159	0.152	0.141	0.080
macro	*0.172*	*0.183*	*0.190*	*0.172*	*0.534*	*0.699*	*0.172*	*0.182*	*0.189*	*0.192*	*0.125*
$\rho\approx$10%											
CTSR	0.000	0.010	0.014	0.000	0.050	0.140	0.000	0.008	0.012	0.017	0.000
DAMADICS	0.389	0.289	0.283	0.389	0.667	0.667	0.389	0.298	0.289	0.289	0.177
Exathlon	0.250	0.172	0.175	0.250	0.620	0.840	0.250	0.186	0.183	0.186	0.135
HAI	0.143	0.164	0.171	0.143	0.571	0.786	0.143	0.160	0.167	0.168	0.140
MIT-BIH	0.350	0.318	0.329	0.350	0.810	0.890	0.350	0.322	0.329	0.320	0.237
Petrobras 3W	0.120	0.148	0.152	0.120	0.570	0.810	0.120	0.145	0.149	0.148	0.122
RATS40K	0.190	0.138	0.132	0.190	0.390	0.510	0.190	0.150	0.142	0.147	0.033
RCAEval	0.220	0.184	0.180	0.220	0.660	0.880	0.220	0.191	0.185	0.191	0.225
ROAD	0.360	0.280	0.304	0.360	0.640	0.880	0.360	0.294	0.313	0.321	0.220
TelecomTS	0.080	0.092	0.098	0.080	0.390	0.660	0.080	0.089	0.094	0.100	0.059
Tennessee Eastman	0.030	0.040	0.049	0.030	0.200	0.400	0.030	0.039	0.046	0.047	0.017
Voraus	0.100	0.122	0.117	0.100	0.450	0.690	0.100	0.120	0.117	0.118	0.079
macro	*0.186*	*0.163*	*0.167*	*0.186*	*0.502*	*0.679*	*0.186*	*0.167*	*0.169*	*0.171*	*0.120*
$\rho\approx$20%											
CTSR	0.010	0.012	0.010	0.010	0.060	0.100	0.010	0.011	0.010	0.013	0.011
DAMADICS	0.222	0.267	0.239	0.222	0.667	0.667	0.222	0.260	0.242	0.258	0.105
Exathlon	0.220	0.200	0.179	0.220	0.680	0.880	0.220	0.203	0.188	0.180	0.139
HAI	0.143	0.193	0.189	0.143	0.536	0.821	0.143	0.181	0.183	0.189	0.230
MIT-BIH	0.370	0.264	0.279	0.370	0.780	0.970	0.370	0.281	0.285	0.289	0.233
Petrobras 3W	0.110	0.112	0.114	0.110	0.420	0.650	0.110	0.108	0.111	0.120	0.077
RATS40K	0.150	0.144	0.133	0.150	0.450	0.580	0.150	0.144	0.136	0.128	0.046
RCAEval	0.120	0.144	0.142	0.120	0.600	0.840	0.120	0.140	0.139	0.150	0.110
ROAD	0.320	0.272	0.292	0.320	0.640	0.840	0.320	0.273	0.293	0.311	0.150
TelecomTS	0.120	0.100	0.101	0.120	0.360	0.570	0.120	0.104	0.103	0.100	0.106
Tennessee Eastman	0.060	0.042	0.046	0.060	0.190	0.380	0.060	0.046	0.047	0.045	0.082
Voraus	0.120	0.098	0.114	0.120	0.380	0.670	0.120	0.105	0.114	0.108	0.071
macro	*0.164*	*0.154*	*0.153*	*0.164*	*0.480*	*0.664*	*0.164*	*0.155*	*0.154*	*0.158*	*0.113*

Table 17: Full metric grid for **Majority**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.014	0.000
DAMADICS	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.091
Exathlon	0.420	0.420	0.420	0.420	0.420	0.420	0.420	0.420	0.420	0.420	0.099
HAI	0.214	0.214	0.214	0.214	0.214	0.214	0.214	0.214	0.214	0.214	0.059
MIT-BIH	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.135
Petrobras 3W	0.240	0.240	0.240	0.240	0.240	0.240	0.240	0.240	0.240	0.240	0.043
RATS40K	0.290	0.290	0.290	0.290	0.290	0.290	0.290	0.290	0.290	0.290	0.024
RCAEval	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.067
ROAD	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.130
TelecomTS	0.260	0.260	0.260	0.260	0.260	0.260	0.260	0.260	0.260	0.260	0.038
Tennessee Eastman	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.005
Voraus	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.028
macro	*0.246*	*0.246*	*0.246*	*0.246*	*0.246*	*0.246*	*0.246*	*0.246*	*0.246*	*0.247*	*0.060*
$\rho\approx$10%											
CTSR	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.014	0.000
DAMADICS	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.091
Exathlon	0.420	0.420	0.420	0.420	0.420	0.420	0.420	0.420	0.420	0.420	0.099
HAI	0.214	0.214	0.214	0.214	0.214	0.214	0.214	0.214	0.214	0.214	0.059
MIT-BIH	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.135
Petrobras 3W	0.240	0.240	0.240	0.240	0.240	0.240	0.240	0.240	0.240	0.240	0.043
RATS40K	0.290	0.290	0.290	0.290	0.290	0.290	0.290	0.290	0.290	0.290	0.024
RCAEval	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.067
ROAD	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.130
TelecomTS	0.260	0.260	0.260	0.260	0.260	0.260	0.260	0.260	0.260	0.260	0.038
Tennessee Eastman	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Voraus	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.028
macro	*0.242*	*0.242*	*0.242*	*0.242*	*0.242*	*0.242*	*0.242*	*0.242*	*0.242*	*0.243*	*0.059*
$\rho\approx$20%											
CTSR	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.012	0.000
DAMADICS	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.091
Exathlon	0.420	0.420	0.420	0.420	0.420	0.420	0.420	0.420	0.420	0.420	0.099
HAI	0.214	0.214	0.214	0.214	0.214	0.214	0.214	0.214	0.214	0.214	0.059
MIT-BIH	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.135
Petrobras 3W	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
RATS40K	0.290	0.290	0.290	0.290	0.290	0.290	0.290	0.290	0.290	0.290	0.024
RCAEval	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ROAD	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.130
TelecomTS	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Tennessee Eastman	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Voraus	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
macro	*0.167*	*0.167*	*0.167*	*0.167*	*0.167*	*0.167*	*0.167*	*0.167*	*0.167*	*0.167*	*0.045*

Table 18: Full metric grid for **ED**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.650	0.516	0.430	0.650	0.850	0.890	0.650	0.540	0.472	0.552	0.637
DAMADICS	0.389	0.411	0.333	0.389	0.722	0.889	0.389	0.411	0.370	0.406	0.302
Exathlon	0.860	0.760	0.732	0.860	0.920	0.950	0.860	0.782	0.755	0.720	0.744
HAI	0.286	0.350	0.354	0.286	0.679	0.786	0.286	0.337	0.349	0.337	0.514
MIT-BIH	0.430	0.496	0.543	0.430	0.710	0.790	0.430	0.484	0.521	0.548	0.224
Petrobras 3W	0.680	0.564	0.507	0.680	0.880	0.940	0.680	0.588	0.539	0.484	0.530
RATS40K	0.410	0.418	0.391	0.410	0.760	0.810	0.410	0.417	0.402	0.385	0.219
RCAEval	0.600	0.508	0.476	0.600	0.880	0.980	0.600	0.524	0.498	0.461	0.489
ROAD	0.480	0.472	0.396	0.480	0.840	0.960	0.480	0.479	0.449	0.451	0.275
TelecomTS	0.830	0.620	0.528	0.830	0.940	0.970	0.830	0.670	0.589	0.510	0.756
Tennessee Eastman	0.650	0.572	0.532	0.650	0.800	0.860	0.650	0.586	0.554	0.496	0.634
Voraus	0.330	0.368	0.321	0.330	0.880	0.940	0.330	0.359	0.331	0.299	0.454
macro	*0.550*	*0.505*	*0.462*	*0.550*	*0.822*	*0.897*	*0.550*	*0.515*	*0.486*	*0.471*	*0.481*
$\rho\approx$10%											
CTSR	0.650	0.512	0.428	0.650	0.850	0.890	0.650	0.537	0.470	0.550	0.637
DAMADICS	0.389	0.378	0.294	0.389	0.722	0.722	0.389	0.381	0.336	0.361	0.417
Exathlon	0.840	0.748	0.710	0.840	0.920	0.950	0.840	0.769	0.736	0.691	0.744
HAI	0.250	0.336	0.321	0.250	0.643	0.750	0.250	0.317	0.317	0.317	0.566
MIT-BIH	0.500	0.498	0.483	0.500	0.610	0.690	0.500	0.494	0.485	0.484	0.257
Petrobras 3W	0.610	0.534	0.472	0.610	0.860	0.910	0.610	0.551	0.501	0.450	0.501
RATS40K	0.380	0.374	0.361	0.380	0.740	0.800	0.380	0.375	0.369	0.350	0.202
RCAEval	0.540	0.456	0.444	0.540	0.820	0.980	0.540	0.474	0.460	0.426	0.464
ROAD	0.480	0.456	0.384	0.480	0.840	0.960	0.480	0.464	0.436	0.427	0.275
TelecomTS	0.820	0.614	0.515	0.820	0.940	0.950	0.820	0.662	0.576	0.498	0.761
Tennessee Eastman	0.630	0.548	0.503	0.630	0.740	0.830	0.630	0.564	0.527	0.472	0.627
Voraus	0.330	0.348	0.309	0.330	0.840	0.940	0.330	0.342	0.319	0.287	0.411
macro	*0.535*	*0.483*	*0.435*	*0.535*	*0.794*	*0.864*	*0.535*	*0.494*	*0.461*	*0.443*	*0.488*
$\rho\approx$20%											
CTSR	0.640	0.508	0.424	0.640	0.840	0.890	0.640	0.533	0.466	0.546	0.627
DAMADICS	0.389	0.344	0.283	0.389	0.722	0.722	0.389	0.348	0.318	0.326	0.384
Exathlon	0.830	0.738	0.692	0.830	0.910	0.950	0.830	0.760	0.720	0.669	0.744
HAI	0.250	0.279	0.289	0.250	0.571	0.714	0.250	0.273	0.287	0.282	0.426
MIT-BIH	0.400	0.428	0.437	0.400	0.580	0.650	0.400	0.422	0.430	0.450	0.208
Petrobras 3W	0.570	0.496	0.434	0.570	0.840	0.910	0.570	0.513	0.463	0.421	0.447
RATS40K	0.370	0.344	0.334	0.370	0.710	0.780	0.370	0.350	0.341	0.324	0.206
RCAEval	0.480	0.416	0.412	0.480	0.800	0.960	0.480	0.433	0.423	0.387	0.436
ROAD	0.480	0.432	0.360	0.480	0.840	0.920	0.480	0.447	0.416	0.404	0.275
TelecomTS	0.820	0.608	0.505	0.820	0.940	0.950	0.820	0.656	0.568	0.488	0.767
Tennessee Eastman	0.600	0.530	0.489	0.600	0.710	0.800	0.600	0.545	0.512	0.457	0.614
Voraus	0.330	0.322	0.289	0.330	0.830	0.920	0.330	0.321	0.300	0.273	0.364
macro	*0.513*	*0.454*	*0.412*	*0.513*	*0.774*	*0.847*	*0.513*	*0.467*	*0.437*	*0.419*	*0.458*

Table 19: Full metric grid for **SBD-D**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.010	0.454	0.403	0.010	0.830	0.840	0.010	0.386	0.375	0.454	0.572
DAMADICS	0.389	0.233	0.217	0.389	0.444	0.556	0.389	0.263	0.247	0.255	0.095
Exathlon	0.370	0.362	0.306	0.370	0.450	0.500	0.370	0.367	0.327	0.275	0.427
HAI	0.286	0.286	0.279	0.286	0.643	0.750	0.286	0.279	0.280	0.272	0.269
MIT-BIH	0.830	0.832	0.808	0.830	0.950	0.950	0.830	0.832	0.816	0.777	0.673
Petrobras 3W	0.430	0.388	0.352	0.430	0.590	0.720	0.430	0.399	0.370	0.333	0.350
RATS40K	0.040	0.138	0.177	0.040	0.430	0.570	0.040	0.109	0.147	0.169	0.004
RCAEval	0.460	0.384	0.350	0.460	0.760	0.840	0.460	0.404	0.373	0.320	0.424
ROAD	0.400	0.184	0.192	0.400	0.680	0.760	0.400	0.227	0.236	0.278	0.183
TelecomTS	0.530	0.384	0.329	0.530	0.860	0.920	0.530	0.417	0.367	0.313	0.475
Tennessee Eastman	0.220	0.188	0.154	0.220	0.500	0.630	0.220	0.198	0.171	0.148	0.192
Voraus	0.270	0.246	0.214	0.270	0.620	0.830	0.270	0.253	0.230	0.210	0.297
macro	*0.353*	*0.340*	*0.315*	*0.353*	*0.646*	*0.739*	*0.353*	*0.345*	*0.328*	*0.317*	*0.330*
$\rho\approx$10%											
CTSR	0.010	0.452	0.400	0.010	0.820	0.840	0.010	0.385	0.373	0.451	0.575
DAMADICS	0.111	0.044	0.100	0.111	0.222	0.444	0.111	0.052	0.091	0.142	0.000
Exathlon	0.360	0.358	0.304	0.360	0.450	0.500	0.360	0.361	0.322	0.268	0.439
HAI	0.214	0.264	0.225	0.214	0.571	0.679	0.214	0.247	0.226	0.224	0.277
MIT-BIH	0.720	0.668	0.675	0.720	0.890	0.930	0.720	0.675	0.678	0.667	0.482
Petrobras 3W	0.230	0.212	0.202	0.230	0.490	0.570	0.230	0.213	0.206	0.196	0.169
RATS40K	0.040	0.132	0.173	0.040	0.420	0.580	0.040	0.104	0.143	0.160	0.004
RCAEval	0.460	0.364	0.336	0.460	0.760	0.820	0.460	0.389	0.361	0.308	0.441
ROAD	0.120	0.176	0.164	0.120	0.680	0.720	0.120	0.171	0.177	0.223	0.178
TelecomTS	0.410	0.276	0.255	0.410	0.720	0.890	0.410	0.307	0.282	0.252	0.365
Tennessee Eastman	0.200	0.174	0.141	0.200	0.460	0.560	0.200	0.183	0.157	0.135	0.213
Voraus	0.230	0.226	0.196	0.230	0.600	0.770	0.230	0.228	0.209	0.195	0.279
macro	*0.259*	*0.279*	*0.264*	*0.259*	*0.590*	*0.692*	*0.259*	*0.276*	*0.269*	*0.269*	*0.285*
$\rho\approx$20%											
CTSR	0.010	0.446	0.394	0.010	0.820	0.840	0.010	0.380	0.368	0.445	0.566
DAMADICS	0.056	0.022	0.022	0.056	0.111	0.222	0.056	0.028	0.028	0.084	0.000
Exathlon	0.360	0.356	0.304	0.360	0.450	0.500	0.360	0.359	0.322	0.266	0.440
HAI	0.179	0.200	0.200	0.179	0.429	0.679	0.179	0.196	0.198	0.194	0.196
MIT-BIH	0.700	0.670	0.658	0.700	0.860	0.920	0.700	0.676	0.666	0.652	0.509
Petrobras 3W	0.190	0.162	0.164	0.190	0.440	0.540	0.190	0.166	0.166	0.163	0.093
RATS40K	0.040	0.116	0.166	0.040	0.360	0.560	0.040	0.093	0.137	0.151	0.004
RCAEval	0.420	0.336	0.300	0.420	0.740	0.800	0.420	0.359	0.326	0.286	0.419
ROAD	0.120	0.152	0.128	0.120	0.600	0.720	0.120	0.150	0.145	0.190	0.314
TelecomTS	0.360	0.218	0.184	0.360	0.550	0.730	0.360	0.251	0.216	0.201	0.321
Tennessee Eastman	0.180	0.156	0.132	0.180	0.430	0.530	0.180	0.164	0.145	0.123	0.169
Voraus	0.210	0.218	0.182	0.210	0.590	0.740	0.210	0.218	0.195	0.181	0.245
macro	*0.235*	*0.254*	*0.236*	*0.235*	*0.532*	*0.648*	*0.235*	*0.253*	*0.243*	*0.245*	*0.273*

Table 20: Full metric grid for **DTW-I**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.730	0.572	0.473	0.730	0.840	0.890	0.730	0.605	0.526	0.583	0.610
DAMADICS	0.667	0.622	0.583	0.667	0.722	0.833	0.667	0.641	0.635	0.651	0.506
Exathlon	0.350	0.318	0.307	0.350	0.400	0.500	0.350	0.325	0.315	0.329	0.372
HAI	0.107	0.121	0.154	0.107	0.464	0.786	0.107	0.115	0.141	0.171	0.032
MIT-BIH	0.560	0.510	0.522	0.560	0.800	0.900	0.560	0.523	0.527	0.529	0.445
Petrobras 3W	0.930	0.876	0.847	0.930	0.940	0.960	0.930	0.888	0.864	0.823	0.879
RATS40K	0.440	0.412	0.402	0.440	0.740	0.820	0.440	0.420	0.414	0.399	0.170
RCAEval	0.340	0.284	0.294	0.340	0.760	0.820	0.340	0.293	0.297	0.290	0.223
ROAD	0.680	0.504	0.340	0.680	0.920	0.960	0.680	0.541	0.438	0.457	0.450
TelecomTS	0.620	0.510	0.476	0.620	0.820	0.890	0.620	0.534	0.502	0.447	0.524
Tennessee Eastman	0.590	0.550	0.480	0.590	0.900	0.940	0.590	0.564	0.510	0.458	0.661
Voraus	0.800	0.628	0.535	0.800	0.980	0.990	0.800	0.666	0.591	0.505	0.692
macro	*0.568*	*0.492*	*0.451*	*0.568*	*0.774*	*0.857*	*0.568*	*0.510*	*0.480*	*0.470*	*0.464*
$\rho\approx$10%											
CTSR	0.730	0.570	0.468	0.730	0.840	0.890	0.730	0.603	0.521	0.578	0.622
DAMADICS	0.222	0.122	0.122	0.222	0.278	0.500	0.222	0.132	0.153	0.200	0.083
Exathlon	0.310	0.280	0.258	0.310	0.370	0.460	0.310	0.288	0.269	0.286	0.359
HAI	0.071	0.107	0.129	0.071	0.464	0.714	0.071	0.097	0.117	0.145	0.037
MIT-BIH	0.580	0.564	0.544	0.580	0.870	0.940	0.580	0.572	0.555	0.550	0.604
Petrobras 3W	0.820	0.836	0.801	0.820	0.940	0.960	0.820	0.838	0.813	0.780	0.879
RATS40K	0.410	0.370	0.365	0.410	0.700	0.810	0.410	0.383	0.378	0.367	0.161
RCAEval	0.180	0.260	0.262	0.180	0.560	0.920	0.180	0.254	0.258	0.264	0.254
ROAD	0.680	0.496	0.336	0.680	0.920	0.960	0.680	0.534	0.433	0.443	0.475
TelecomTS	0.600	0.478	0.434	0.600	0.810	0.870	0.600	0.504	0.464	0.400	0.550
Tennessee Eastman	0.550	0.510	0.450	0.550	0.850	0.920	0.550	0.523	0.477	0.422	0.618
Voraus	0.790	0.606	0.516	0.790	0.970	0.980	0.790	0.646	0.573	0.489	0.762
macro	*0.495*	*0.433*	*0.390*	*0.495*	*0.714*	*0.827*	*0.495*	*0.448*	*0.418*	*0.410*	*0.450*
$\rho\approx$20%											
CTSR	0.730	0.568	0.467	0.730	0.840	0.890	0.730	0.601	0.520	0.577	0.622
DAMADICS	0.222	0.078	0.094	0.222	0.222	0.278	0.222	0.100	0.127	0.137	0.100
Exathlon	0.280	0.264	0.233	0.280	0.370	0.410	0.280	0.271	0.247	0.258	0.354
HAI	0.071	0.064	0.096	0.071	0.286	0.607	0.071	0.065	0.089	0.126	0.024
MIT-BIH	0.530	0.576	0.558	0.530	0.800	0.880	0.530	0.565	0.556	0.562	0.440
Petrobras 3W	0.790	0.798	0.770	0.790	0.940	0.950	0.790	0.797	0.778	0.756	0.864
RATS40K	0.380	0.350	0.338	0.380	0.700	0.800	0.380	0.358	0.348	0.332	0.161
RCAEval	0.220	0.248	0.228	0.220	0.760	0.920	0.220	0.247	0.232	0.232	0.231
ROAD	0.680	0.496	0.336	0.680	0.920	0.960	0.680	0.534	0.433	0.438	0.489
TelecomTS	0.570	0.470	0.412	0.570	0.790	0.860	0.570	0.493	0.444	0.374	0.523
Tennessee Eastman	0.520	0.486	0.426	0.520	0.760	0.890	0.520	0.495	0.450	0.398	0.609
Voraus	0.790	0.590	0.497	0.790	0.970	0.980	0.790	0.633	0.556	0.472	0.752
macro	*0.482*	*0.416*	*0.371*	*0.482*	*0.696*	*0.785*	*0.482*	*0.430*	*0.398*	*0.388*	*0.431*

Table 21: Full metric grid for **DTW-D**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.730	0.572	0.473	0.730	0.840	0.890	0.730	0.605	0.526	0.583	0.610
DAMADICS	0.500	0.356	0.300	0.500	0.667	0.667	0.500	0.384	0.350	0.318	0.295
Exathlon	0.400	0.388	0.357	0.400	0.460	0.500	0.400	0.394	0.370	0.361	0.408
HAI	0.107	0.114	0.136	0.107	0.393	0.571	0.107	0.114	0.131	0.156	0.123
MIT-BIH	0.470	0.442	0.426	0.470	0.630	0.720	0.470	0.450	0.436	0.430	0.389
Petrobras 3W	0.930	0.888	0.859	0.930	0.950	0.950	0.930	0.900	0.875	0.827	0.894
RATS40K	0.440	0.408	0.400	0.440	0.740	0.820	0.440	0.418	0.412	0.399	0.168
RCAEval	0.280	0.276	0.288	0.280	0.780	0.820	0.280	0.270	0.280	0.279	0.165
ROAD	0.400	0.256	0.224	0.400	0.720	0.840	0.400	0.299	0.291	0.371	0.338
TelecomTS	0.610	0.504	0.464	0.610	0.820	0.890	0.610	0.526	0.490	0.440	0.548
Tennessee Eastman	0.390	0.392	0.355	0.390	0.680	0.780	0.390	0.396	0.368	0.338	0.442
Voraus	0.830	0.608	0.454	0.830	0.970	0.990	0.830	0.661	0.538	0.439	0.847
macro	*0.507*	*0.434*	*0.395*	*0.507*	*0.721*	*0.787*	*0.507*	*0.451*	*0.422*	*0.412*	*0.435*
$\rho\approx$10%											
CTSR	0.730	0.570	0.468	0.730	0.840	0.890	0.730	0.603	0.521	0.578	0.622
DAMADICS	0.167	0.033	0.022	0.167	0.167	0.222	0.167	0.057	0.049	0.094	0.000
Exathlon	0.270	0.320	0.302	0.270	0.450	0.480	0.270	0.311	0.302	0.305	0.383
HAI	0.107	0.093	0.121	0.107	0.393	0.571	0.107	0.095	0.116	0.129	0.039
MIT-BIH	0.500	0.478	0.463	0.500	0.740	0.820	0.500	0.481	0.470	0.468	0.441
Petrobras 3W	0.840	0.846	0.814	0.840	0.950	0.950	0.840	0.851	0.827	0.787	0.894
RATS40K	0.400	0.368	0.363	0.400	0.700	0.810	0.400	0.380	0.376	0.366	0.154
RCAEval	0.120	0.256	0.248	0.120	0.580	0.920	0.120	0.233	0.235	0.250	0.192
ROAD	0.320	0.232	0.204	0.320	0.680	0.800	0.320	0.260	0.257	0.334	0.350
TelecomTS	0.580	0.472	0.415	0.580	0.820	0.870	0.580	0.493	0.445	0.386	0.573
Tennessee Eastman	0.370	0.368	0.338	0.370	0.610	0.720	0.370	0.374	0.351	0.321	0.405
Voraus	0.820	0.600	0.437	0.820	0.970	0.990	0.820	0.652	0.524	0.428	0.868
macro	*0.435*	*0.386*	*0.350*	*0.435*	*0.658*	*0.754*	*0.435*	*0.399*	*0.373*	*0.371*	*0.410*
$\rho\approx$20%											
CTSR	0.730	0.568	0.467	0.730	0.840	0.890	0.730	0.601	0.520	0.577	0.622
DAMADICS	0.167	0.033	0.017	0.167	0.167	0.167	0.167	0.057	0.046	0.046	0.000
Exathlon	0.260	0.274	0.266	0.260	0.400	0.450	0.260	0.273	0.268	0.276	0.359
HAI	0.071	0.086	0.104	0.071	0.357	0.536	0.071	0.080	0.095	0.110	0.056
MIT-BIH	0.410	0.408	0.418	0.410	0.570	0.720	0.410	0.408	0.415	0.434	0.351
Petrobras 3W	0.780	0.802	0.781	0.780	0.950	0.950	0.780	0.800	0.787	0.758	0.850
RATS40K	0.380	0.348	0.339	0.380	0.700	0.800	0.380	0.357	0.349	0.332	0.154
RCAEval	0.180	0.232	0.212	0.180	0.760	0.920	0.180	0.221	0.210	0.217	0.128
ROAD	0.320	0.232	0.200	0.320	0.680	0.800	0.320	0.260	0.254	0.317	0.350
TelecomTS	0.560	0.464	0.389	0.560	0.790	0.860	0.560	0.483	0.424	0.359	0.566
Tennessee Eastman	0.360	0.352	0.326	0.360	0.560	0.680	0.360	0.358	0.338	0.309	0.367
Voraus	0.820	0.578	0.420	0.820	0.960	0.990	0.820	0.634	0.508	0.411	0.814
macro	*0.420*	*0.365*	*0.328*	*0.420*	*0.644*	*0.730*	*0.420*	*0.378*	*0.351*	*0.345*	*0.385*

Table 22: Full metric grid for **SAX-BM25**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.410	0.302	0.257	0.410	0.730	0.810	0.410	0.324	0.285	0.333	0.268
DAMADICS	0.333	0.311	0.289	0.333	0.722	0.889	0.333	0.322	0.315	0.333	0.231
Exathlon	0.870	0.856	0.849	0.870	0.970	1.000	0.870	0.860	0.854	0.822	0.825
HAI	0.429	0.421	0.393	0.429	0.750	0.857	0.429	0.425	0.403	0.371	0.360
MIT-BIH	0.680	0.644	0.639	0.680	0.900	0.950	0.680	0.654	0.647	0.621	0.697
Petrobras 3W	0.670	0.614	0.559	0.670	0.940	0.970	0.670	0.628	0.586	0.557	0.647
RATS40K	0.310	0.322	0.306	0.310	0.640	0.750	0.310	0.323	0.313	0.287	0.168
RCAEval	0.720	0.688	0.656	0.720	0.920	0.940	0.720	0.698	0.673	0.609	0.672
ROAD	0.800	0.696	0.644	0.800	1.000	1.000	0.800	0.731	0.732	0.721	0.751
TelecomTS	0.800	0.562	0.471	0.800	0.930	0.960	0.800	0.614	0.532	0.457	0.708
Tennessee Eastman	0.330	0.324	0.297	0.330	0.650	0.810	0.330	0.327	0.308	0.280	0.382
Voraus	0.290	0.230	0.196	0.290	0.690	0.830	0.290	0.241	0.215	0.204	0.328
macro	*0.553*	*0.498*	*0.463*	*0.553*	*0.820*	*0.897*	*0.553*	*0.512*	*0.489*	*0.466*	*0.503*
$\rho\approx$10%											
CTSR	0.390	0.292	0.249	0.390	0.720	0.820	0.390	0.315	0.277	0.325	0.248
DAMADICS	0.333	0.300	0.233	0.333	0.722	0.778	0.333	0.307	0.271	0.297	0.295
Exathlon	0.800	0.840	0.824	0.800	0.970	1.000	0.800	0.834	0.825	0.794	0.797
HAI	0.357	0.364	0.339	0.357	0.714	0.786	0.357	0.366	0.348	0.327	0.349
MIT-BIH	0.680	0.622	0.604	0.680	0.890	0.940	0.680	0.631	0.615	0.616	0.654
Petrobras 3W	0.670	0.586	0.540	0.670	0.920	0.980	0.670	0.603	0.566	0.533	0.627
RATS40K	0.320	0.302	0.285	0.320	0.630	0.740	0.320	0.308	0.296	0.272	0.171
RCAEval	0.660	0.648	0.612	0.660	0.920	0.960	0.660	0.655	0.626	0.568	0.690
ROAD	0.800	0.688	0.628	0.800	0.960	1.000	0.800	0.723	0.715	0.699	0.751
TelecomTS	0.790	0.538	0.442	0.790	0.920	0.960	0.790	0.591	0.506	0.437	0.736
Tennessee Eastman	0.290	0.304	0.278	0.290	0.600	0.740	0.290	0.304	0.287	0.264	0.360
Voraus	0.280	0.214	0.180	0.280	0.670	0.800	0.280	0.226	0.200	0.187	0.315
macro	*0.531*	*0.475*	*0.435*	*0.531*	*0.803*	*0.875*	*0.531*	*0.489*	*0.461*	*0.443*	*0.499*
$\rho\approx$20%											
CTSR	0.370	0.282	0.242	0.370	0.690	0.820	0.370	0.304	0.269	0.317	0.252
DAMADICS	0.222	0.233	0.206	0.222	0.611	0.778	0.222	0.238	0.230	0.249	0.217
Exathlon	0.770	0.826	0.807	0.770	0.980	1.000	0.770	0.816	0.807	0.773	0.843
HAI	0.321	0.314	0.264	0.321	0.571	0.750	0.321	0.318	0.282	0.284	0.310
MIT-BIH	0.660	0.608	0.595	0.660	0.900	0.950	0.660	0.615	0.605	0.587	0.476
Petrobras 3W	0.640	0.564	0.516	0.640	0.900	0.940	0.640	0.579	0.541	0.504	0.599
RATS40K	0.300	0.290	0.266	0.300	0.580	0.720	0.300	0.295	0.278	0.252	0.174
RCAEval	0.560	0.612	0.566	0.560	0.920	0.960	0.560	0.607	0.576	0.533	0.690
ROAD	0.800	0.680	0.588	0.800	1.000	1.000	0.800	0.720	0.688	0.664	0.751
TelecomTS	0.770	0.508	0.421	0.770	0.900	0.960	0.770	0.564	0.483	0.412	0.734
Tennessee Eastman	0.270	0.290	0.268	0.270	0.580	0.710	0.270	0.291	0.275	0.252	0.295
Voraus	0.260	0.204	0.163	0.260	0.650	0.780	0.260	0.214	0.184	0.172	0.318
macro	*0.495*	*0.451*	*0.408*	*0.495*	*0.774*	*0.864*	*0.495*	*0.464*	*0.435*	*0.417*	*0.472*

Table 23: Full metric grid for **SFA-BM25**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.360	0.308	0.262	0.360	0.680	0.750	0.360	0.318	0.283	0.335	0.313
DAMADICS	0.444	0.333	0.306	0.444	0.667	0.833	0.444	0.362	0.345	0.346	0.300
Exathlon	0.790	0.732	0.695	0.790	0.960	0.980	0.790	0.744	0.714	0.658	0.800
HAI	0.357	0.364	0.325	0.357	0.750	0.786	0.357	0.368	0.340	0.320	0.282
MIT-BIH	0.480	0.468	0.433	0.480	0.800	0.940	0.480	0.468	0.444	0.429	0.379
Petrobras 3W	0.660	0.538	0.510	0.660	0.870	0.940	0.660	0.564	0.536	0.478	0.576
RATS40K	0.310	0.230	0.167	0.310	0.600	0.810	0.310	0.252	0.199	0.186	0.100
RCAEval	0.660	0.584	0.526	0.660	0.880	0.960	0.660	0.600	0.555	0.504	0.679
ROAD	0.920	0.776	0.652	0.920	0.960	1.000	0.920	0.832	0.783	0.765	0.862
TelecomTS	0.260	0.220	0.209	0.260	0.600	0.790	0.260	0.229	0.218	0.206	0.218
Tennessee Eastman	0.370	0.316	0.274	0.370	0.600	0.710	0.370	0.328	0.295	0.269	0.339
Voraus	0.270	0.196	0.178	0.270	0.650	0.840	0.270	0.211	0.195	0.183	0.288
macro	*0.490*	*0.422*	*0.378*	*0.490*	*0.751*	*0.862*	*0.490*	*0.440*	*0.409*	*0.390*	*0.428*
$\rho\approx$10%											
CTSR	0.380	0.292	0.246	0.380	0.660	0.760	0.380	0.314	0.274	0.334	0.322
DAMADICS	0.389	0.289	0.256	0.389	0.611	0.778	0.389	0.316	0.295	0.304	0.292
Exathlon	0.690	0.708	0.674	0.690	0.950	1.000	0.690	0.706	0.684	0.643	0.784
HAI	0.286	0.314	0.286	0.286	0.679	0.750	0.286	0.315	0.296	0.258	0.292
MIT-BIH	0.450	0.444	0.436	0.450	0.670	0.810	0.450	0.446	0.440	0.435	0.374
Petrobras 3W	0.600	0.544	0.486	0.600	0.860	0.960	0.600	0.559	0.514	0.459	0.559
RATS40K	0.280	0.192	0.183	0.280	0.690	0.710	0.280	0.213	0.201	0.174	0.150
RCAEval	0.680	0.568	0.478	0.680	0.880	0.940	0.680	0.588	0.520	0.471	0.739
ROAD	0.920	0.736	0.616	0.920	0.960	1.000	0.920	0.802	0.750	0.736	0.862
TelecomTS	0.270	0.212	0.212	0.270	0.610	0.830	0.270	0.223	0.219	0.211	0.200
Tennessee Eastman	0.370	0.310	0.280	0.370	0.600	0.710	0.370	0.322	0.296	0.262	0.296
Voraus	0.230	0.182	0.164	0.230	0.620	0.810	0.230	0.195	0.180	0.167	0.271
macro	*0.462*	*0.399*	*0.360*	*0.462*	*0.732*	*0.838*	*0.462*	*0.417*	*0.389*	*0.371*	*0.428*
$\rho\approx$20%											
CTSR	0.370	0.298	0.250	0.370	0.630	0.760	0.370	0.314	0.274	0.322	0.318
DAMADICS	0.333	0.300	0.278	0.333	0.667	0.722	0.333	0.314	0.306	0.279	0.308
Exathlon	0.690	0.692	0.676	0.690	0.940	0.990	0.690	0.695	0.683	0.629	0.781
HAI	0.321	0.293	0.293	0.321	0.607	0.750	0.321	0.302	0.300	0.248	0.212
MIT-BIH	0.420	0.398	0.413	0.420	0.600	0.770	0.420	0.400	0.408	0.412	0.290
Petrobras 3W	0.640	0.520	0.458	0.640	0.870	0.910	0.640	0.538	0.488	0.445	0.547
RATS40K	0.370	0.192	0.177	0.370	0.610	0.630	0.370	0.231	0.205	0.179	0.082
RCAEval	0.640	0.544	0.490	0.640	0.860	0.940	0.640	0.567	0.521	0.459	0.728
ROAD	0.880	0.720	0.580	0.880	0.960	1.000	0.880	0.780	0.711	0.690	0.862
TelecomTS	0.250	0.216	0.190	0.250	0.610	0.780	0.250	0.225	0.203	0.184	0.278
Tennessee Eastman	0.370	0.300	0.267	0.370	0.580	0.650	0.370	0.314	0.286	0.255	0.341
Voraus	0.210	0.162	0.148	0.210	0.590	0.770	0.210	0.175	0.162	0.147	0.257
macro	*0.458*	*0.386*	*0.352*	*0.458*	*0.710*	*0.806*	*0.458*	*0.404*	*0.379*	*0.354*	*0.417*

Table 24: Full metric grid for **CHARM**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.650	0.516	0.430	0.650	0.850	0.890	0.650	0.540	0.472	0.552	0.637
DAMADICS	0.389	0.411	0.333	0.389	0.722	0.889	0.389	0.411	0.370	0.406	0.302
Exathlon	0.860	0.760	0.732	0.860	0.920	0.950	0.860	0.782	0.755	0.720	0.744
HAI	0.286	0.350	0.354	0.286	0.679	0.786	0.286	0.337	0.349	0.337	0.514
MIT-BIH	0.430	0.496	0.543	0.430	0.710	0.790	0.430	0.484	0.521	0.548	0.224
Petrobras 3W	0.680	0.564	0.507	0.680	0.880	0.940	0.680	0.588	0.539	0.484	0.530
RATS40K	0.410	0.418	0.391	0.410	0.760	0.810	0.410	0.417	0.402	0.385	0.219
RCAEval	0.600	0.508	0.476	0.600	0.880	0.980	0.600	0.524	0.498	0.461	0.489
ROAD	0.480	0.472	0.396	0.480	0.840	0.960	0.480	0.479	0.449	0.451	0.275
TelecomTS	0.830	0.620	0.528	0.830	0.940	0.970	0.830	0.670	0.589	0.510	0.756
Tennessee Eastman	0.650	0.572	0.532	0.650	0.800	0.860	0.650	0.586	0.554	0.496	0.634
Voraus	0.330	0.368	0.321	0.330	0.880	0.940	0.330	0.359	0.331	0.299	0.454
macro	*0.550*	*0.505*	*0.462*	*0.550*	*0.822*	*0.897*	*0.550*	*0.515*	*0.486*	*0.471*	*0.481*
$\rho\approx$10%											
CTSR	0.650	0.512	0.428	0.650	0.850	0.890	0.650	0.537	0.470	0.550	0.637
DAMADICS	0.389	0.378	0.294	0.389	0.722	0.722	0.389	0.381	0.336	0.361	0.417
Exathlon	0.840	0.748	0.710	0.840	0.920	0.950	0.840	0.769	0.736	0.691	0.744
HAI	0.250	0.336	0.321	0.250	0.643	0.750	0.250	0.317	0.317	0.317	0.566
MIT-BIH	0.500	0.498	0.483	0.500	0.610	0.690	0.500	0.494	0.485	0.484	0.257
Petrobras 3W	0.610	0.534	0.472	0.610	0.860	0.910	0.610	0.551	0.501	0.450	0.501
RATS40K	0.380	0.374	0.361	0.380	0.740	0.800	0.380	0.375	0.369	0.350	0.202
RCAEval	0.540	0.456	0.444	0.540	0.820	0.980	0.540	0.474	0.460	0.426	0.464
ROAD	0.480	0.456	0.384	0.480	0.840	0.960	0.480	0.464	0.436	0.427	0.275
TelecomTS	0.820	0.614	0.515	0.820	0.940	0.950	0.820	0.662	0.576	0.498	0.761
Tennessee Eastman	0.630	0.548	0.503	0.630	0.740	0.830	0.630	0.564	0.527	0.472	0.627
Voraus	0.330	0.348	0.309	0.330	0.840	0.940	0.330	0.342	0.319	0.287	0.411
macro	*0.535*	*0.483*	*0.435*	*0.535*	*0.794*	*0.864*	*0.535*	*0.494*	*0.461*	*0.443*	*0.488*
$\rho\approx$20%											
CTSR	0.640	0.508	0.424	0.640	0.840	0.890	0.640	0.533	0.466	0.546	0.627
DAMADICS	0.389	0.344	0.283	0.389	0.722	0.722	0.389	0.348	0.318	0.326	0.384
Exathlon	0.830	0.738	0.692	0.830	0.910	0.950	0.830	0.760	0.720	0.669	0.744
HAI	0.250	0.279	0.289	0.250	0.571	0.714	0.250	0.273	0.287	0.282	0.426
MIT-BIH	0.400	0.428	0.437	0.400	0.580	0.650	0.400	0.422	0.430	0.450	0.208
Petrobras 3W	0.570	0.496	0.434	0.570	0.840	0.910	0.570	0.513	0.463	0.421	0.447
RATS40K	0.370	0.344	0.334	0.370	0.710	0.780	0.370	0.350	0.341	0.324	0.206
RCAEval	0.480	0.416	0.412	0.480	0.800	0.960	0.480	0.433	0.423	0.387	0.436
ROAD	0.480	0.432	0.360	0.480	0.840	0.920	0.480	0.447	0.416	0.404	0.275
TelecomTS	0.820	0.608	0.505	0.820	0.940	0.950	0.820	0.656	0.568	0.488	0.767
Tennessee Eastman	0.600	0.530	0.489	0.600	0.710	0.800	0.600	0.545	0.512	0.457	0.614
Voraus	0.330	0.322	0.289	0.330	0.830	0.920	0.330	0.321	0.300	0.273	0.364
macro	*0.513*	*0.454*	*0.412*	*0.513*	*0.774*	*0.847*	*0.513*	*0.467*	*0.437*	*0.419*	*0.458*

Table 25: Full metric grid for **CHARM + NR**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.630	0.512	0.424	0.630	0.800	0.870	0.630	0.541	0.469	0.544	0.629
DAMADICS	0.556	0.489	0.478	0.556	0.778	1.000	0.556	0.499	0.518	0.521	0.354
Exathlon	0.920	0.850	0.818	0.920	0.990	1.000	0.920	0.863	0.837	0.782	0.886
HAI	0.393	0.321	0.332	0.393	0.643	0.786	0.393	0.331	0.338	0.326	0.462
MIT-BIH	0.460	0.588	0.638	0.460	0.790	0.880	0.460	0.565	0.608	0.650	0.328
Petrobras 3W	0.680	0.526	0.481	0.680	0.800	0.890	0.680	0.553	0.511	0.462	0.530
RATS40K	0.440	0.408	0.377	0.440	0.740	0.790	0.440	0.417	0.393	0.363	0.238
RCAEval	0.500	0.528	0.468	0.500	0.880	0.940	0.500	0.519	0.481	0.442	0.637
ROAD	0.520	0.512	0.448	0.520	0.840	0.880	0.520	0.522	0.496	0.495	0.360
TelecomTS	0.880	0.694	0.585	0.880	0.990	0.990	0.880	0.735	0.645	0.574	0.777
Tennessee Eastman	0.560	0.548	0.531	0.560	0.710	0.800	0.560	0.554	0.540	0.506	0.486
Voraus	0.530	0.416	0.373	0.530	0.870	0.940	0.530	0.443	0.408	0.362	0.529
macro	*0.589*	*0.533*	*0.496*	*0.589*	*0.819*	*0.897*	*0.589*	*0.545*	*0.520*	*0.502*	*0.518*
$\rho\approx$10%											
CTSR	0.630	0.512	0.424	0.630	0.800	0.870	0.630	0.541	0.469	0.542	0.629
DAMADICS	0.500	0.389	0.400	0.500	0.778	1.000	0.500	0.401	0.433	0.454	0.303
Exathlon	0.910	0.844	0.812	0.910	0.990	1.000	0.910	0.856	0.830	0.776	0.886
HAI	0.393	0.321	0.329	0.393	0.643	0.786	0.393	0.331	0.335	0.324	0.462
MIT-BIH	0.590	0.682	0.710	0.590	0.820	0.880	0.590	0.664	0.690	0.709	0.369
Petrobras 3W	0.670	0.504	0.458	0.670	0.800	0.890	0.670	0.533	0.490	0.443	0.508
RATS40K	0.430	0.410	0.371	0.430	0.740	0.780	0.430	0.415	0.387	0.355	0.241
RCAEval	0.480	0.520	0.466	0.480	0.880	0.940	0.480	0.508	0.475	0.431	0.642
ROAD	0.520	0.504	0.440	0.520	0.840	0.880	0.520	0.515	0.489	0.480	0.296
TelecomTS	0.880	0.688	0.576	0.880	0.990	0.990	0.880	0.730	0.638	0.567	0.796
Tennessee Eastman	0.550	0.546	0.525	0.550	0.700	0.790	0.550	0.551	0.535	0.500	0.492
Voraus	0.530	0.416	0.371	0.530	0.870	0.940	0.530	0.443	0.407	0.361	0.529
macro	*0.590*	*0.528*	*0.490*	*0.590*	*0.821*	*0.895*	*0.590*	*0.541*	*0.515*	*0.495*	*0.513*
$\rho\approx$20%											
CTSR	0.620	0.512	0.422	0.620	0.800	0.860	0.620	0.539	0.467	0.540	0.629
DAMADICS	0.500	0.367	0.367	0.500	0.722	1.000	0.500	0.384	0.406	0.426	0.303
Exathlon	0.880	0.836	0.799	0.880	0.990	1.000	0.880	0.845	0.817	0.763	0.886
HAI	0.393	0.321	0.318	0.393	0.643	0.786	0.393	0.330	0.327	0.319	0.462
MIT-BIH	0.490	0.486	0.521	0.490	0.650	0.780	0.490	0.484	0.509	0.543	0.254
Petrobras 3W	0.650	0.488	0.441	0.650	0.800	0.890	0.650	0.515	0.472	0.427	0.460
RATS40K	0.410	0.402	0.360	0.410	0.740	0.780	0.410	0.404	0.375	0.345	0.243
RCAEval	0.480	0.504	0.460	0.480	0.880	0.940	0.480	0.493	0.465	0.421	0.607
ROAD	0.520	0.480	0.428	0.520	0.840	0.880	0.520	0.499	0.478	0.466	0.296
TelecomTS	0.870	0.682	0.572	0.870	0.990	0.990	0.870	0.723	0.633	0.558	0.763
Tennessee Eastman	0.550	0.546	0.520	0.550	0.700	0.750	0.550	0.550	0.531	0.496	0.501
Voraus	0.530	0.414	0.370	0.530	0.870	0.940	0.530	0.441	0.406	0.359	0.529
macro	*0.574*	*0.503*	*0.465*	*0.574*	*0.802*	*0.883*	*0.574*	*0.517*	*0.490*	*0.472*	*0.494*

Table 26: Full metric grid for **Chronos-2**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.510	0.452	0.365	0.510	0.810	0.830	0.510	0.465	0.401	0.453	0.496
DAMADICS	0.111	0.189	0.200	0.111	0.611	0.611	0.111	0.176	0.191	0.210	0.029
Exathlon	0.960	0.912	0.881	0.960	0.990	0.990	0.960	0.921	0.896	0.858	0.897
HAI	0.321	0.321	0.311	0.321	0.750	0.964	0.321	0.325	0.317	0.322	0.346
MIT-BIH	0.720	0.708	0.719	0.720	0.860	0.890	0.720	0.708	0.716	0.705	0.393
Petrobras 3W	0.680	0.534	0.476	0.680	0.840	0.880	0.680	0.560	0.510	0.458	0.594
RATS40K	0.400	0.398	0.373	0.400	0.750	0.840	0.400	0.401	0.387	0.372	0.219
RCAEval	0.540	0.492	0.458	0.540	0.800	0.900	0.540	0.501	0.475	0.450	0.450
ROAD	0.520	0.472	0.424	0.520	0.920	1.000	0.520	0.502	0.479	0.465	0.413
TelecomTS	0.700	0.570	0.498	0.700	0.940	0.980	0.700	0.606	0.543	0.470	0.614
Tennessee Eastman	0.450	0.420	0.397	0.450	0.720	0.840	0.450	0.425	0.407	0.371	0.469
Voraus	0.450	0.364	0.306	0.450	0.840	0.890	0.450	0.377	0.334	0.308	0.422
macro	*0.530*	*0.486*	*0.451*	*0.530*	*0.819*	*0.885*	*0.530*	*0.497*	*0.471*	*0.454*	*0.445*
$\rho\approx$10%											
CTSR	0.500	0.450	0.363	0.500	0.810	0.830	0.500	0.462	0.398	0.451	0.498
DAMADICS	0.111	0.178	0.183	0.111	0.556	0.611	0.111	0.163	0.171	0.179	0.056
Exathlon	0.960	0.912	0.879	0.960	0.990	0.990	0.960	0.921	0.894	0.850	0.897
HAI	0.321	0.293	0.271	0.321	0.714	0.929	0.321	0.302	0.284	0.287	0.364
MIT-BIH	0.760	0.752	0.746	0.760	0.900	0.950	0.760	0.752	0.747	0.738	0.420
Petrobras 3W	0.660	0.514	0.451	0.660	0.830	0.870	0.660	0.540	0.487	0.435	0.588
RATS40K	0.390	0.372	0.360	0.390	0.750	0.820	0.390	0.378	0.372	0.354	0.238
RCAEval	0.460	0.464	0.436	0.460	0.800	0.900	0.460	0.461	0.443	0.419	0.483
ROAD	0.520	0.440	0.388	0.520	0.920	1.000	0.520	0.480	0.451	0.438	0.413
TelecomTS	0.690	0.556	0.472	0.690	0.930	0.980	0.690	0.591	0.521	0.451	0.680
Tennessee Eastman	0.430	0.406	0.372	0.430	0.690	0.810	0.430	0.408	0.383	0.357	0.469
Voraus	0.430	0.328	0.288	0.430	0.810	0.890	0.430	0.346	0.314	0.287	0.375
macro	*0.519*	*0.472*	*0.434*	*0.519*	*0.808*	*0.882*	*0.519*	*0.484*	*0.455*	*0.437*	*0.457*
$\rho\approx$20%											
CTSR	0.500	0.446	0.360	0.500	0.810	0.830	0.500	0.458	0.394	0.448	0.505
DAMADICS	0.056	0.144	0.161	0.056	0.444	0.611	0.056	0.126	0.144	0.161	0.031
Exathlon	0.960	0.900	0.866	0.960	0.990	0.990	0.960	0.912	0.884	0.841	0.928
HAI	0.250	0.257	0.243	0.250	0.643	0.857	0.250	0.258	0.248	0.251	0.360
MIT-BIH	0.710	0.716	0.700	0.710	0.840	0.890	0.710	0.714	0.704	0.692	0.459
Petrobras 3W	0.590	0.478	0.427	0.590	0.810	0.860	0.590	0.500	0.457	0.406	0.543
RATS40K	0.380	0.358	0.344	0.380	0.740	0.820	0.380	0.365	0.356	0.336	0.251
RCAEval	0.420	0.412	0.390	0.420	0.800	0.900	0.420	0.412	0.396	0.373	0.454
ROAD	0.520	0.432	0.376	0.520	0.920	1.000	0.520	0.473	0.440	0.423	0.413
TelecomTS	0.680	0.542	0.459	0.680	0.920	0.970	0.680	0.577	0.507	0.437	0.668
Tennessee Eastman	0.390	0.380	0.353	0.390	0.620	0.760	0.390	0.382	0.362	0.338	0.447
Voraus	0.410	0.306	0.269	0.410	0.800	0.880	0.410	0.325	0.294	0.267	0.360
macro	*0.489*	*0.448*	*0.412*	*0.489*	*0.778*	*0.864*	*0.489*	*0.458*	*0.432*	*0.414*	*0.452*

Table 27: Full metric grid for **Chronos-2 + NR**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.600	0.444	0.357	0.600	0.770	0.820	0.600	0.474	0.403	0.463	0.535
DAMADICS	0.167	0.233	0.267	0.167	0.611	0.889	0.167	0.230	0.261	0.292	0.163
Exathlon	0.910	0.852	0.810	0.910	0.950	0.980	0.910	0.860	0.828	0.781	0.748
HAI	0.357	0.364	0.354	0.357	0.714	0.786	0.357	0.363	0.357	0.359	0.369
MIT-BIH	0.680	0.752	0.759	0.680	0.870	0.920	0.680	0.739	0.748	0.745	0.416
Petrobras 3W	0.580	0.484	0.439	0.580	0.850	0.910	0.580	0.510	0.470	0.421	0.549
RATS40K	0.410	0.372	0.370	0.410	0.740	0.820	0.410	0.384	0.381	0.363	0.198
RCAEval	0.500	0.420	0.400	0.500	0.840	0.900	0.500	0.431	0.414	0.391	0.395
ROAD	0.480	0.456	0.376	0.480	0.760	0.840	0.480	0.463	0.413	0.459	0.252
TelecomTS	0.770	0.646	0.568	0.770	0.970	0.980	0.770	0.675	0.611	0.532	0.705
Tennessee Eastman	0.510	0.446	0.414	0.510	0.710	0.750	0.510	0.459	0.432	0.402	0.469
Voraus	0.430	0.400	0.354	0.430	0.870	0.940	0.430	0.407	0.375	0.354	0.542
macro	*0.533*	*0.489*	*0.456*	*0.533*	*0.805*	*0.878*	*0.533*	*0.500*	*0.474*	*0.464*	*0.445*
$\rho\approx$10%											
CTSR	0.600	0.442	0.356	0.600	0.770	0.820	0.600	0.472	0.402	0.461	0.546
DAMADICS	0.167	0.189	0.211	0.167	0.611	0.833	0.167	0.184	0.208	0.238	0.127
Exathlon	0.910	0.852	0.810	0.910	0.950	0.980	0.910	0.860	0.828	0.779	0.748
HAI	0.321	0.350	0.343	0.321	0.714	0.786	0.321	0.345	0.341	0.347	0.332
MIT-BIH	0.800	0.810	0.818	0.800	0.870	0.910	0.800	0.809	0.815	0.811	0.428
Petrobras 3W	0.550	0.472	0.418	0.550	0.850	0.900	0.550	0.492	0.448	0.402	0.554
RATS40K	0.410	0.366	0.364	0.410	0.750	0.820	0.410	0.378	0.376	0.359	0.197
RCAEval	0.460	0.396	0.374	0.460	0.820	0.900	0.460	0.408	0.390	0.367	0.362
ROAD	0.480	0.432	0.364	0.480	0.680	0.800	0.480	0.445	0.401	0.443	0.252
TelecomTS	0.750	0.632	0.558	0.750	0.970	0.980	0.750	0.661	0.600	0.522	0.704
Tennessee Eastman	0.510	0.444	0.413	0.510	0.700	0.750	0.510	0.456	0.430	0.397	0.480
Voraus	0.430	0.400	0.349	0.430	0.870	0.940	0.430	0.407	0.371	0.350	0.545
macro	*0.532*	*0.482*	*0.448*	*0.532*	*0.796*	*0.868*	*0.532*	*0.493*	*0.467*	*0.456*	*0.440*
$\rho\approx$20%											
CTSR	0.600	0.442	0.356	0.600	0.770	0.820	0.600	0.472	0.402	0.461	0.546
DAMADICS	0.167	0.189	0.189	0.167	0.611	0.833	0.167	0.175	0.186	0.204	0.077
Exathlon	0.910	0.844	0.805	0.910	0.950	0.980	0.910	0.854	0.824	0.775	0.746
HAI	0.321	0.343	0.336	0.321	0.714	0.786	0.321	0.337	0.334	0.334	0.314
MIT-BIH	0.760	0.780	0.792	0.760	0.860	0.920	0.760	0.775	0.785	0.790	0.416
Petrobras 3W	0.550	0.460	0.406	0.550	0.830	0.900	0.550	0.481	0.437	0.390	0.515
RATS40K	0.400	0.358	0.357	0.400	0.740	0.810	0.400	0.370	0.369	0.353	0.193
RCAEval	0.420	0.364	0.348	0.420	0.800	0.900	0.420	0.379	0.363	0.344	0.332
ROAD	0.440	0.384	0.328	0.440	0.640	0.760	0.440	0.393	0.359	0.385	0.196
TelecomTS	0.750	0.620	0.543	0.750	0.970	0.980	0.750	0.650	0.586	0.510	0.712
Tennessee Eastman	0.500	0.436	0.408	0.500	0.690	0.750	0.500	0.448	0.424	0.393	0.462
Voraus	0.430	0.384	0.335	0.430	0.850	0.930	0.430	0.395	0.360	0.340	0.548
macro	*0.521*	*0.467*	*0.434*	*0.521*	*0.785*	*0.864*	*0.521*	*0.477*	*0.452*	*0.440*	*0.421*

Table 28: Full metric grid for **MantisV2**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.590	0.522	0.411	0.590	0.780	0.850	0.590	0.539	0.456	0.528	0.596
DAMADICS	0.500	0.578	0.611	0.500	0.944	1.000	0.500	0.555	0.610	0.622	0.359
Exathlon	0.950	0.934	0.908	0.950	1.000	1.000	0.950	0.939	0.919	0.897	0.964
HAI	0.536	0.429	0.396	0.536	0.714	0.964	0.536	0.460	0.432	0.397	0.539
MIT-BIH	0.310	0.328	0.333	0.310	0.400	0.450	0.310	0.325	0.329	0.324	0.355
Petrobras 3W	0.860	0.816	0.766	0.860	0.980	0.980	0.860	0.827	0.788	0.709	0.776
RATS40K	0.440	0.436	0.407	0.440	0.750	0.840	0.440	0.439	0.421	0.414	0.268
RCAEval	0.340	0.336	0.338	0.340	0.880	0.880	0.340	0.332	0.335	0.319	0.285
ROAD	0.640	0.456	0.388	0.640	0.880	0.880	0.640	0.494	0.446	0.437	0.395
TelecomTS	0.860	0.662	0.544	0.860	0.920	0.960	0.860	0.706	0.610	0.523	0.779
Tennessee Eastman	0.660	0.546	0.498	0.660	0.810	0.870	0.660	0.568	0.526	0.472	0.638
Voraus	0.310	0.258	0.228	0.310	0.690	0.860	0.310	0.267	0.245	0.221	0.266
macro	*0.583*	*0.525*	*0.486*	*0.583*	*0.812*	*0.878*	*0.583*	*0.538*	*0.510*	*0.489*	*0.518*
$\rho\approx$10%											
CTSR	0.590	0.518	0.411	0.590	0.770	0.850	0.590	0.535	0.455	0.525	0.598
DAMADICS	0.278	0.389	0.439	0.278	0.889	1.000	0.278	0.363	0.430	0.489	0.252
Exathlon	0.930	0.924	0.902	0.930	1.000	1.000	0.930	0.926	0.910	0.885	0.965
HAI	0.536	0.407	0.357	0.536	0.714	0.857	0.536	0.437	0.397	0.370	0.586
MIT-BIH	0.320	0.300	0.287	0.320	0.330	0.390	0.320	0.304	0.294	0.292	0.140
Petrobras 3W	0.830	0.782	0.734	0.830	0.980	0.980	0.830	0.791	0.755	0.684	0.763
RATS40K	0.430	0.412	0.395	0.430	0.740	0.830	0.430	0.420	0.409	0.393	0.249
RCAEval	0.160	0.304	0.306	0.160	0.640	1.000	0.160	0.285	0.295	0.294	0.332
ROAD	0.640	0.408	0.360	0.640	0.840	0.880	0.640	0.459	0.420	0.409	0.388
TelecomTS	0.860	0.656	0.535	0.860	0.920	0.960	0.860	0.699	0.601	0.514	0.786
Tennessee Eastman	0.630	0.518	0.464	0.630	0.770	0.820	0.630	0.541	0.496	0.446	0.601
Voraus	0.280	0.236	0.212	0.280	0.660	0.830	0.280	0.243	0.226	0.206	0.245
macro	*0.540*	*0.488*	*0.450*	*0.540*	*0.771*	*0.866*	*0.540*	*0.500*	*0.474*	*0.459*	*0.492*
$\rho\approx$20%											
CTSR	0.590	0.518	0.409	0.590	0.770	0.840	0.590	0.535	0.454	0.523	0.601
DAMADICS	0.222	0.344	0.400	0.222	0.833	1.000	0.222	0.319	0.390	0.423	0.220
Exathlon	0.930	0.918	0.894	0.930	1.000	1.000	0.930	0.920	0.903	0.878	0.967
HAI	0.464	0.379	0.332	0.464	0.643	0.821	0.464	0.404	0.367	0.339	0.582
MIT-BIH	0.270	0.274	0.287	0.270	0.400	0.440	0.270	0.271	0.282	0.291	0.125
Petrobras 3W	0.780	0.728	0.688	0.780	0.960	0.980	0.780	0.740	0.708	0.651	0.685
RATS40K	0.420	0.404	0.388	0.420	0.730	0.830	0.420	0.412	0.401	0.380	0.205
RCAEval	0.240	0.288	0.266	0.240	0.840	1.000	0.240	0.286	0.270	0.262	0.278
ROAD	0.640	0.376	0.316	0.640	0.840	0.880	0.640	0.437	0.385	0.389	0.333
TelecomTS	0.860	0.642	0.526	0.860	0.910	0.950	0.860	0.687	0.592	0.502	0.784
Tennessee Eastman	0.620	0.504	0.451	0.620	0.750	0.800	0.620	0.527	0.482	0.427	0.612
Voraus	0.260	0.214	0.197	0.260	0.620	0.800	0.260	0.223	0.210	0.189	0.245
macro	*0.525*	*0.466*	*0.430*	*0.525*	*0.775*	*0.862*	*0.525*	*0.480*	*0.454*	*0.438*	*0.470*

Table 29: Full metric grid for **MantisV2 + NR**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.610	0.508	0.410	0.610	0.770	0.850	0.610	0.534	0.458	0.527	0.580
DAMADICS	0.611	0.578	0.472	0.611	0.778	0.889	0.611	0.596	0.544	0.540	0.389
Exathlon	0.940	0.938	0.924	0.940	0.970	0.980	0.940	0.939	0.929	0.914	0.958
HAI	0.357	0.307	0.357	0.357	0.464	0.786	0.357	0.312	0.350	0.354	0.351
MIT-BIH	0.510	0.538	0.546	0.510	0.700	0.830	0.510	0.528	0.537	0.550	0.356
Petrobras 3W	0.780	0.728	0.676	0.780	0.930	0.990	0.780	0.744	0.703	0.639	0.639
RATS40K	0.450	0.432	0.406	0.450	0.760	0.890	0.450	0.438	0.421	0.403	0.231
RCAEval	0.460	0.400	0.404	0.460	0.880	0.880	0.460	0.408	0.408	0.379	0.340
ROAD	0.600	0.576	0.504	0.600	0.920	0.960	0.600	0.588	0.556	0.548	0.542
TelecomTS	0.920	0.762	0.659	0.920	0.980	0.990	0.920	0.799	0.714	0.633	0.874
Tennessee Eastman	0.660	0.646	0.615	0.660	0.820	0.910	0.660	0.651	0.628	0.597	0.577
Voraus	0.480	0.434	0.393	0.480	0.880	0.950	0.480	0.442	0.413	0.373	0.499
macro	*0.615*	*0.571*	*0.531*	*0.615*	*0.821*	*0.909*	*0.615*	*0.581*	*0.555*	*0.538*	*0.528*
$\rho\approx$10%											
CTSR	0.610	0.508	0.407	0.610	0.770	0.840	0.610	0.534	0.456	0.526	0.583
DAMADICS	0.500	0.456	0.394	0.500	0.778	0.889	0.500	0.478	0.456	0.472	0.354
Exathlon	0.940	0.932	0.917	0.940	0.970	0.980	0.940	0.934	0.923	0.908	0.958
HAI	0.321	0.300	0.357	0.321	0.464	0.786	0.321	0.301	0.344	0.346	0.351
MIT-BIH	0.500	0.526	0.525	0.500	0.750	0.790	0.500	0.519	0.521	0.523	0.361
Petrobras 3W	0.760	0.708	0.663	0.760	0.920	0.990	0.760	0.724	0.688	0.624	0.609
RATS40K	0.450	0.428	0.399	0.450	0.750	0.890	0.450	0.434	0.415	0.393	0.231
RCAEval	0.300	0.384	0.384	0.300	0.680	1.000	0.300	0.376	0.379	0.363	0.423
ROAD	0.600	0.576	0.496	0.600	0.920	0.960	0.600	0.588	0.550	0.534	0.542
TelecomTS	0.920	0.756	0.653	0.920	0.980	0.980	0.920	0.793	0.709	0.629	0.874
Tennessee Eastman	0.660	0.642	0.611	0.660	0.800	0.880	0.660	0.648	0.624	0.594	0.588
Voraus	0.460	0.428	0.384	0.460	0.880	0.950	0.460	0.435	0.404	0.367	0.500
macro	*0.585*	*0.554*	*0.516*	*0.585*	*0.805*	*0.911*	*0.585*	*0.564*	*0.539*	*0.523*	*0.531*
$\rho\approx$20%											
CTSR	0.610	0.508	0.406	0.610	0.770	0.840	0.610	0.534	0.455	0.525	0.583
DAMADICS	0.500	0.422	0.372	0.500	0.778	0.889	0.500	0.449	0.433	0.442	0.333
Exathlon	0.940	0.930	0.913	0.940	0.970	0.980	0.940	0.933	0.920	0.904	0.954
HAI	0.321	0.300	0.350	0.321	0.464	0.786	0.321	0.301	0.339	0.342	0.356
MIT-BIH	0.510	0.480	0.485	0.510	0.630	0.680	0.510	0.484	0.486	0.480	0.271
Petrobras 3W	0.730	0.698	0.646	0.730	0.920	0.980	0.730	0.709	0.669	0.606	0.613
RATS40K	0.450	0.418	0.393	0.450	0.720	0.890	0.450	0.426	0.409	0.387	0.227
RCAEval	0.380	0.384	0.356	0.380	0.880	1.000	0.380	0.389	0.366	0.342	0.364
ROAD	0.600	0.552	0.464	0.600	0.920	0.960	0.600	0.570	0.525	0.514	0.542
TelecomTS	0.920	0.752	0.646	0.920	0.980	0.980	0.920	0.789	0.702	0.621	0.878
Tennessee Eastman	0.650	0.638	0.607	0.650	0.780	0.850	0.650	0.644	0.620	0.590	0.588
Voraus	0.460	0.424	0.379	0.460	0.880	0.950	0.460	0.432	0.400	0.363	0.500
macro	*0.589*	*0.542*	*0.501*	*0.589*	*0.808*	*0.899*	*0.589*	*0.555*	*0.527*	*0.510*	*0.517*

Table 30: Full metric grid for **TiRex**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.670	0.518	0.456	0.670	0.840	0.880	0.670	0.548	0.494	0.565	0.543
DAMADICS	0.444	0.322	0.333	0.444	0.611	0.833	0.444	0.345	0.353	0.361	0.208
Exathlon	0.860	0.840	0.821	0.860	0.990	1.000	0.860	0.849	0.832	0.786	0.860
HAI	0.429	0.321	0.314	0.429	0.786	0.964	0.429	0.344	0.335	0.333	0.350
MIT-BIH	0.580	0.614	0.633	0.580	0.820	0.880	0.580	0.607	0.622	0.641	0.458
Petrobras 3W	0.740	0.582	0.521	0.740	0.890	0.920	0.740	0.616	0.562	0.508	0.586
RATS40K	0.460	0.454	0.431	0.460	0.770	0.830	0.460	0.455	0.441	0.412	0.267
RCAEval	0.540	0.480	0.472	0.540	0.840	0.940	0.540	0.488	0.480	0.451	0.453
ROAD	0.600	0.464	0.404	0.600	0.920	0.960	0.600	0.495	0.468	0.478	0.371
TelecomTS	0.870	0.648	0.529	0.870	0.950	0.970	0.870	0.697	0.597	0.512	0.762
Tennessee Eastman	0.490	0.418	0.371	0.490	0.720	0.810	0.490	0.434	0.396	0.351	0.542
Voraus	0.270	0.226	0.197	0.270	0.730	0.870	0.270	0.241	0.217	0.199	0.281
macro	*0.579*	*0.491*	*0.457*	*0.579*	*0.822*	*0.905*	*0.579*	*0.510*	*0.483*	*0.466*	*0.473*
$\rho\approx$10%											
CTSR	0.670	0.516	0.454	0.670	0.830	0.880	0.670	0.547	0.493	0.563	0.543
DAMADICS	0.333	0.289	0.278	0.333	0.611	0.833	0.333	0.297	0.296	0.325	0.167
Exathlon	0.860	0.838	0.816	0.860	0.990	1.000	0.860	0.847	0.828	0.777	0.860
HAI	0.286	0.307	0.257	0.286	0.714	0.893	0.286	0.309	0.275	0.277	0.175
MIT-BIH	0.520	0.528	0.543	0.520	0.710	0.890	0.520	0.522	0.534	0.540	0.363
Petrobras 3W	0.690	0.554	0.485	0.690	0.890	0.920	0.690	0.583	0.525	0.474	0.544
RATS40K	0.460	0.418	0.397	0.460	0.750	0.820	0.460	0.426	0.411	0.378	0.232
RCAEval	0.500	0.412	0.418	0.500	0.840	0.920	0.500	0.429	0.428	0.408	0.418
ROAD	0.600	0.448	0.388	0.600	0.920	0.960	0.600	0.482	0.453	0.454	0.371
TelecomTS	0.860	0.632	0.505	0.860	0.950	0.970	0.860	0.679	0.574	0.492	0.760
Tennessee Eastman	0.460	0.396	0.350	0.460	0.670	0.750	0.460	0.409	0.373	0.330	0.537
Voraus	0.250	0.208	0.177	0.250	0.700	0.840	0.250	0.221	0.196	0.185	0.241
macro	*0.541*	*0.462*	*0.422*	*0.541*	*0.798*	*0.890*	*0.541*	*0.479*	*0.449*	*0.434*	*0.434*
$\rho\approx$20%											
CTSR	0.670	0.516	0.453	0.670	0.830	0.880	0.670	0.546	0.492	0.562	0.543
DAMADICS	0.333	0.267	0.228	0.333	0.611	0.722	0.333	0.279	0.259	0.294	0.165
Exathlon	0.860	0.830	0.798	0.860	0.990	1.000	0.860	0.839	0.813	0.760	0.862
HAI	0.250	0.293	0.232	0.250	0.714	0.821	0.250	0.288	0.248	0.247	0.129
MIT-BIH	0.620	0.588	0.595	0.620	0.670	0.730	0.620	0.593	0.596	0.609	0.413
Petrobras 3W	0.660	0.504	0.449	0.660	0.850	0.910	0.660	0.536	0.486	0.441	0.500
RATS40K	0.430	0.402	0.370	0.430	0.740	0.810	0.430	0.408	0.386	0.359	0.234
RCAEval	0.440	0.388	0.380	0.440	0.800	0.880	0.440	0.404	0.394	0.383	0.435
ROAD	0.560	0.416	0.368	0.560	0.880	0.960	0.560	0.450	0.429	0.424	0.350
TelecomTS	0.860	0.614	0.487	0.860	0.950	0.970	0.860	0.665	0.559	0.474	0.743
Tennessee Eastman	0.420	0.380	0.336	0.420	0.630	0.710	0.420	0.389	0.355	0.312	0.473
Voraus	0.240	0.196	0.163	0.240	0.660	0.810	0.240	0.207	0.181	0.168	0.238
macro	*0.529*	*0.449*	*0.405*	*0.529*	*0.777*	*0.850*	*0.529*	*0.467*	*0.433*	*0.419*	*0.424*

Table 31: Full metric grid for **TiRex + NR**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.670	0.516	0.435	0.670	0.850	0.870	0.670	0.552	0.483	0.550	0.596
DAMADICS	0.444	0.322	0.344	0.444	0.778	0.889	0.444	0.349	0.373	0.407	0.215
Exathlon	0.880	0.828	0.809	0.880	0.990	0.990	0.880	0.838	0.822	0.758	0.820
HAI	0.321	0.329	0.339	0.321	0.679	0.786	0.321	0.328	0.335	0.333	0.374
MIT-BIH	0.650	0.662	0.656	0.650	0.800	0.850	0.650	0.661	0.657	0.663	0.367
Petrobras 3W	0.610	0.484	0.441	0.610	0.800	0.910	0.610	0.513	0.474	0.420	0.537
RATS40K	0.410	0.404	0.380	0.410	0.730	0.800	0.410	0.408	0.392	0.369	0.216
RCAEval	0.360	0.388	0.368	0.360	0.740	0.860	0.360	0.383	0.371	0.364	0.438
ROAD	0.600	0.456	0.400	0.600	0.960	1.000	0.600	0.498	0.472	0.499	0.202
TelecomTS	0.880	0.640	0.546	0.880	0.970	0.970	0.880	0.694	0.610	0.517	0.746
Tennessee Eastman	0.480	0.426	0.417	0.480	0.640	0.740	0.480	0.439	0.428	0.396	0.415
Voraus	0.420	0.368	0.330	0.420	0.860	0.940	0.420	0.382	0.352	0.316	0.402
macro	*0.560*	*0.485*	*0.455*	*0.560*	*0.816*	*0.884*	*0.560*	*0.504*	*0.481*	*0.466*	*0.444*
$\rho\approx$10%											
CTSR	0.670	0.516	0.435	0.670	0.850	0.870	0.670	0.552	0.483	0.550	0.596
DAMADICS	0.278	0.300	0.283	0.278	0.778	0.889	0.278	0.304	0.308	0.366	0.215
Exathlon	0.880	0.828	0.809	0.880	0.990	0.990	0.880	0.838	0.822	0.757	0.820
HAI	0.321	0.329	0.339	0.321	0.679	0.786	0.321	0.328	0.335	0.328	0.374
MIT-BIH	0.750	0.726	0.718	0.750	0.890	0.930	0.750	0.728	0.721	0.713	0.407
Petrobras 3W	0.580	0.470	0.426	0.580	0.780	0.910	0.580	0.496	0.457	0.404	0.507
RATS40K	0.410	0.404	0.374	0.410	0.730	0.800	0.410	0.408	0.388	0.365	0.212
RCAEval	0.360	0.380	0.360	0.360	0.740	0.860	0.360	0.377	0.364	0.356	0.437
ROAD	0.600	0.448	0.392	0.600	0.920	1.000	0.600	0.492	0.466	0.478	0.259
TelecomTS	0.870	0.632	0.541	0.870	0.970	0.970	0.870	0.686	0.604	0.509	0.750
Tennessee Eastman	0.480	0.418	0.406	0.480	0.620	0.700	0.480	0.432	0.419	0.391	0.458
Voraus	0.410	0.362	0.323	0.410	0.850	0.940	0.410	0.376	0.346	0.311	0.391
macro	*0.551*	*0.484*	*0.451*	*0.551*	*0.816*	*0.887*	*0.551*	*0.501*	*0.476*	*0.461*	*0.452*
$\rho\approx$20%											
CTSR	0.670	0.516	0.435	0.670	0.850	0.870	0.670	0.552	0.483	0.550	0.596
DAMADICS	0.278	0.278	0.250	0.278	0.722	0.889	0.278	0.285	0.281	0.333	0.215
Exathlon	0.860	0.818	0.799	0.860	0.990	0.990	0.860	0.827	0.812	0.750	0.820
HAI	0.321	0.329	0.336	0.321	0.679	0.786	0.321	0.326	0.331	0.321	0.397
MIT-BIH	0.490	0.556	0.571	0.490	0.750	0.790	0.490	0.542	0.557	0.567	0.416
Petrobras 3W	0.580	0.456	0.414	0.580	0.780	0.890	0.580	0.484	0.445	0.394	0.431
RATS40K	0.390	0.396	0.363	0.390	0.730	0.800	0.390	0.397	0.377	0.356	0.206
RCAEval	0.360	0.368	0.344	0.360	0.720	0.860	0.360	0.368	0.351	0.342	0.427
ROAD	0.520	0.448	0.372	0.520	0.920	1.000	0.520	0.477	0.438	0.447	0.370
TelecomTS	0.870	0.622	0.529	0.870	0.970	0.970	0.870	0.676	0.593	0.498	0.761
Tennessee Eastman	0.470	0.414	0.401	0.470	0.620	0.670	0.470	0.427	0.414	0.385	0.455
Voraus	0.410	0.356	0.311	0.410	0.850	0.940	0.410	0.370	0.336	0.300	0.392
macro	*0.518*	*0.463*	*0.427*	*0.518*	*0.798*	*0.871*	*0.518*	*0.478*	*0.451*	*0.437*	*0.457*

Table 32: Full metric grid for **CHARM + GPC**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.711	0.572
DAMADICS	0.278	0.278	0.278	0.278	0.278	0.333	0.278	0.278	0.289	0.305	0.217
Exathlon	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.677
HAI	0.429	0.429	0.421	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.388
MIT-BIH	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.243
Petrobras 3W	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.562
RATS40K	0.540	0.540	0.541	0.540	0.540	0.550	0.540	0.540	0.541	0.544	0.215
RCAEval	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.528
ROAD	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.483	0.200
TelecomTS	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.625
Tennessee Eastman	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.642
Voraus	0.540	0.540	0.538	0.540	0.540	0.550	0.540	0.540	0.541	0.544	0.474
macro	*0.555*	*0.555*	*0.554*	*0.555*	*0.555*	*0.561*	*0.555*	*0.555*	*0.556*	*0.565*	*0.445*
$\rho\approx$10%											
CTSR	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.688	0.541
DAMADICS	0.278	0.278	0.278	0.278	0.278	0.333	0.278	0.278	0.289	0.305	0.217
Exathlon	0.830	0.830	0.830	0.830	0.830	0.830	0.830	0.830	0.830	0.830	0.683
HAI	0.429	0.429	0.421	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.416
MIT-BIH	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.265
Petrobras 3W	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.471
RATS40K	0.520	0.520	0.521	0.520	0.520	0.530	0.520	0.520	0.521	0.524	0.209
RCAEval	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.538
ROAD	0.560	0.560	0.552	0.560	0.560	0.560	0.560	0.560	0.560	0.563	0.304
TelecomTS	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.657
Tennessee Eastman	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.640
Voraus	0.540	0.540	0.538	0.540	0.540	0.550	0.540	0.540	0.541	0.544	0.476
macro	*0.561*	*0.561*	*0.559*	*0.561*	*0.561*	*0.567*	*0.561*	*0.561*	*0.562*	*0.571*	*0.451*
$\rho\approx$20%											
CTSR	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.672	0.520
DAMADICS	0.278	0.278	0.278	0.278	0.278	0.333	0.278	0.278	0.289	0.305	0.234
Exathlon	0.830	0.830	0.830	0.830	0.830	0.830	0.830	0.830	0.830	0.830	0.736
HAI	0.429	0.429	0.421	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.427
MIT-BIH	0.320	0.320	0.320	0.320	0.320	0.320	0.320	0.320	0.320	0.320	0.206
Petrobras 3W	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.428
RATS40K	0.440	0.440	0.441	0.440	0.440	0.450	0.440	0.440	0.441	0.444	0.189
RCAEval	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.542
ROAD	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.528	0.247
TelecomTS	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.614
Tennessee Eastman	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.585
Voraus	0.480	0.480	0.478	0.480	0.480	0.490	0.480	0.480	0.481	0.484	0.435
macro	*0.516*	*0.516*	*0.516*	*0.516*	*0.516*	*0.523*	*0.516*	*0.516*	*0.517*	*0.528*	*0.430*

Table 33: Full metric grid for **CHARM + MajVote**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.631	0.516
DAMADICS	0.278	0.278	0.272	0.278	0.278	0.333	0.278	0.278	0.285	0.289	0.217
Exathlon	0.780	0.780	0.780	0.780	0.780	0.780	0.780	0.780	0.780	0.780	0.541
HAI	0.464	0.464	0.457	0.464	0.464	0.464	0.464	0.464	0.464	0.464	0.417
MIT-BIH	0.440	0.440	0.440	0.440	0.440	0.440	0.440	0.440	0.440	0.440	0.275
Petrobras 3W	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.556
RATS40K	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.228
RCAEval	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.524
ROAD	0.480	0.480	0.484	0.480	0.480	0.520	0.480	0.480	0.483	0.492	0.209
TelecomTS	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.546
Tennessee Eastman	0.640	0.640	0.640	0.640	0.640	0.640	0.640	0.640	0.640	0.640	0.619
Voraus	0.460	0.460	0.457	0.460	0.460	0.470	0.460	0.460	0.461	0.462	0.362
macro	*0.537*	*0.537*	*0.536*	*0.537*	*0.537*	*0.546*	*0.537*	*0.537*	*0.538*	*0.544*	*0.417*
$\rho\approx$10%											
CTSR	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.630	0.520
DAMADICS	0.278	0.278	0.272	0.278	0.278	0.333	0.278	0.278	0.285	0.289	0.217
Exathlon	0.780	0.780	0.780	0.780	0.780	0.780	0.780	0.780	0.780	0.780	0.539
HAI	0.429	0.429	0.421	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.416
MIT-BIH	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.360
Petrobras 3W	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.462
RATS40K	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.225
RCAEval	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.517
ROAD	0.480	0.480	0.484	0.480	0.480	0.520	0.480	0.480	0.483	0.495	0.200
TelecomTS	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.561
Tennessee Eastman	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.627
Voraus	0.480	0.480	0.477	0.480	0.480	0.490	0.480	0.480	0.481	0.482	0.392
macro	*0.535*	*0.535*	*0.534*	*0.535*	*0.535*	*0.543*	*0.535*	*0.535*	*0.536*	*0.542*	*0.420*
$\rho\approx$20%											
CTSR	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.636	0.526
DAMADICS	0.278	0.278	0.267	0.278	0.278	0.333	0.278	0.278	0.281	0.288	0.243
Exathlon	0.760	0.760	0.760	0.760	0.760	0.760	0.760	0.760	0.760	0.760	0.523
HAI	0.429	0.429	0.421	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.427
MIT-BIH	0.340	0.340	0.340	0.340	0.340	0.340	0.340	0.340	0.340	0.340	0.175
Petrobras 3W	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.420
RATS40K	0.460	0.460	0.460	0.460	0.460	0.460	0.460	0.460	0.460	0.460	0.175
RCAEval	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.532
ROAD	0.480	0.480	0.484	0.480	0.480	0.520	0.480	0.480	0.483	0.497	0.204
TelecomTS	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.534
Tennessee Eastman	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.579
Voraus	0.440	0.440	0.437	0.440	0.440	0.450	0.440	0.440	0.441	0.442	0.370
macro	*0.500*	*0.500*	*0.498*	*0.500*	*0.500*	*0.508*	*0.500*	*0.500*	*0.500*	*0.507*	*0.392*

Table 34: Full metric grid for **CHARM + QPurity**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.570	0.510	0.417	0.570	0.730	0.840	0.570	0.523	0.454	0.502	0.525
DAMADICS	0.222	0.289	0.278	0.222	0.444	0.444	0.222	0.273	0.274	0.282	0.167
Exathlon	0.810	0.786	0.781	0.810	0.810	0.810	0.810	0.792	0.786	0.781	0.608
HAI	0.429	0.407	0.414	0.429	0.429	0.536	0.429	0.414	0.417	0.404	0.299
MIT-BIH	0.440	0.434	0.434	0.440	0.440	0.490	0.440	0.435	0.435	0.435	0.275
Petrobras 3W	0.690	0.672	0.648	0.690	0.730	0.730	0.690	0.675	0.658	0.609	0.571
RATS40K	0.530	0.496	0.474	0.530	0.650	0.730	0.530	0.503	0.488	0.462	0.192
RCAEval	0.540	0.552	0.540	0.540	0.580	0.580	0.540	0.551	0.543	0.543	0.527
ROAD	0.480	0.464	0.464	0.480	0.520	0.560	0.480	0.467	0.466	0.448	0.209
TelecomTS	0.570	0.542	0.539	0.570	0.630	0.720	0.570	0.549	0.545	0.521	0.462
Tennessee Eastman	0.630	0.612	0.601	0.630	0.700	0.730	0.630	0.617	0.608	0.584	0.581
Voraus	0.460	0.462	0.458	0.460	0.490	0.530	0.460	0.462	0.461	0.454	0.373
macro	*0.531*	*0.519*	*0.504*	*0.531*	*0.596*	*0.642*	*0.531*	*0.522*	*0.511*	*0.502*	*0.399*
$\rho\approx$10%											
CTSR	0.590	0.550	0.464	0.590	0.700	0.850	0.590	0.559	0.497	0.542	0.526
DAMADICS	0.222	0.267	0.267	0.222	0.278	0.444	0.222	0.259	0.263	0.267	0.167
Exathlon	0.790	0.786	0.780	0.790	0.800	0.800	0.790	0.788	0.783	0.779	0.608
HAI	0.393	0.393	0.407	0.393	0.500	0.536	0.393	0.394	0.404	0.394	0.276
MIT-BIH	0.530	0.532	0.530	0.530	0.540	0.540	0.530	0.531	0.530	0.531	0.277
Petrobras 3W	0.580	0.572	0.575	0.580	0.630	0.680	0.580	0.573	0.575	0.552	0.442
RATS40K	0.480	0.476	0.470	0.480	0.620	0.710	0.480	0.479	0.477	0.449	0.180
RCAEval	0.540	0.532	0.542	0.540	0.580	0.620	0.540	0.532	0.539	0.538	0.504
ROAD	0.480	0.464	0.464	0.480	0.560	0.600	0.480	0.471	0.468	0.469	0.129
TelecomTS	0.600	0.564	0.550	0.600	0.660	0.750	0.600	0.572	0.560	0.531	0.498
Tennessee Eastman	0.590	0.594	0.585	0.590	0.670	0.750	0.590	0.595	0.588	0.576	0.602
Voraus	0.480	0.476	0.459	0.480	0.540	0.540	0.480	0.478	0.467	0.458	0.404
macro	*0.523*	*0.517*	*0.508*	*0.523*	*0.590*	*0.652*	*0.523*	*0.519*	*0.513*	*0.507*	*0.384*
$\rho\approx$20%											
CTSR	0.590	0.548	0.457	0.590	0.700	0.840	0.590	0.557	0.491	0.537	0.520
DAMADICS	0.278	0.322	0.300	0.278	0.444	0.444	0.278	0.316	0.303	0.322	0.196
Exathlon	0.780	0.768	0.760	0.780	0.780	0.780	0.780	0.771	0.764	0.759	0.592
HAI	0.429	0.386	0.396	0.429	0.500	0.536	0.429	0.391	0.396	0.389	0.294
MIT-BIH	0.360	0.358	0.357	0.360	0.390	0.390	0.360	0.359	0.357	0.359	0.233
Petrobras 3W	0.510	0.512	0.502	0.510	0.520	0.540	0.510	0.511	0.505	0.480	0.406
RATS40K	0.430	0.418	0.405	0.430	0.530	0.600	0.430	0.423	0.411	0.385	0.160
RCAEval	0.520	0.520	0.526	0.520	0.520	0.560	0.520	0.520	0.524	0.521	0.532
ROAD	0.480	0.464	0.468	0.480	0.560	0.600	0.480	0.471	0.472	0.460	0.133
TelecomTS	0.630	0.586	0.561	0.630	0.710	0.770	0.630	0.594	0.574	0.544	0.532
Tennessee Eastman	0.570	0.556	0.536	0.570	0.590	0.590	0.570	0.560	0.545	0.529	0.589
Voraus	0.440	0.430	0.423	0.440	0.480	0.490	0.440	0.433	0.429	0.417	0.355
macro	*0.501*	*0.489*	*0.474*	*0.501*	*0.560*	*0.595*	*0.501*	*0.492*	*0.481*	*0.475*	*0.378*

Table 35: Full metric grid for **CHARM + PRF**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.480	0.444	0.412	0.480	0.790	0.880	0.480	0.452	0.428	0.508	0.428
DAMADICS	0.389	0.311	0.289	0.389	0.611	0.778	0.389	0.327	0.315	0.360	0.274
Exathlon	0.740	0.728	0.703	0.740	0.860	0.920	0.740	0.731	0.713	0.681	0.513
HAI	0.429	0.379	0.364	0.429	0.571	0.786	0.429	0.395	0.383	0.367	0.331
MIT-BIH	0.440	0.450	0.454	0.440	0.480	0.520	0.440	0.447	0.451	0.464	0.275
Petrobras 3W	0.660	0.560	0.495	0.660	0.790	0.870	0.660	0.575	0.524	0.473	0.561
RATS40K	0.450	0.416	0.376	0.450	0.760	0.800	0.450	0.427	0.397	0.394	0.221
RCAEval	0.500	0.492	0.460	0.500	0.760	0.880	0.500	0.494	0.470	0.451	0.509
ROAD	0.480	0.376	0.380	0.480	0.800	0.880	0.480	0.398	0.411	0.412	0.214
TelecomTS	0.650	0.564	0.513	0.650	0.860	0.930	0.650	0.584	0.541	0.486	0.535
Tennessee Eastman	0.540	0.546	0.526	0.540	0.790	0.850	0.540	0.550	0.535	0.482	0.597
Voraus	0.280	0.302	0.287	0.280	0.740	0.870	0.280	0.297	0.291	0.268	0.249
macro	*0.503*	*0.464*	*0.438*	*0.503*	*0.734*	*0.830*	*0.503*	*0.473*	*0.455*	*0.446*	*0.392*
$\rho\approx$10%											
CTSR	0.490	0.444	0.411	0.490	0.790	0.880	0.490	0.453	0.428	0.508	0.428
DAMADICS	0.389	0.278	0.267	0.389	0.556	0.722	0.389	0.306	0.299	0.341	0.239
Exathlon	0.740	0.722	0.682	0.740	0.870	0.920	0.740	0.725	0.696	0.657	0.513
HAI	0.393	0.350	0.332	0.393	0.536	0.714	0.393	0.355	0.345	0.331	0.317
MIT-BIH	0.520	0.520	0.535	0.520	0.560	0.630	0.520	0.522	0.531	0.535	0.277
Petrobras 3W	0.630	0.506	0.458	0.630	0.760	0.850	0.630	0.530	0.488	0.441	0.498
RATS40K	0.410	0.368	0.347	0.410	0.720	0.790	0.410	0.383	0.366	0.355	0.211
RCAEval	0.460	0.452	0.438	0.460	0.700	0.900	0.460	0.454	0.443	0.414	0.529
ROAD	0.480	0.376	0.376	0.480	0.720	0.840	0.480	0.401	0.407	0.397	0.197
TelecomTS	0.630	0.558	0.499	0.630	0.860	0.920	0.630	0.574	0.527	0.473	0.563
Tennessee Eastman	0.460	0.506	0.492	0.460	0.740	0.820	0.460	0.501	0.493	0.447	0.535
Voraus	0.300	0.300	0.284	0.300	0.720	0.870	0.300	0.297	0.289	0.261	0.262
macro	*0.492*	*0.448*	*0.427*	*0.492*	*0.711*	*0.821*	*0.492*	*0.458*	*0.443*	*0.430*	*0.381*
$\rho\approx$20%											
CTSR	0.490	0.438	0.407	0.490	0.780	0.870	0.490	0.449	0.425	0.506	0.439
DAMADICS	0.389	0.256	0.250	0.389	0.556	0.667	0.389	0.287	0.282	0.310	0.253
Exathlon	0.730	0.700	0.662	0.730	0.840	0.910	0.730	0.706	0.677	0.631	0.514
HAI	0.357	0.314	0.304	0.357	0.536	0.714	0.357	0.324	0.317	0.305	0.387
MIT-BIH	0.370	0.354	0.364	0.370	0.390	0.480	0.370	0.357	0.362	0.364	0.310
Petrobras 3W	0.540	0.450	0.420	0.540	0.710	0.840	0.540	0.468	0.441	0.408	0.421
RATS40K	0.350	0.344	0.318	0.350	0.660	0.780	0.350	0.345	0.329	0.324	0.203
RCAEval	0.400	0.408	0.390	0.400	0.720	0.860	0.400	0.409	0.397	0.384	0.502
ROAD	0.520	0.384	0.348	0.520	0.760	0.840	0.520	0.416	0.397	0.381	0.325
TelecomTS	0.640	0.554	0.488	0.640	0.840	0.910	0.640	0.572	0.520	0.466	0.563
Tennessee Eastman	0.440	0.498	0.471	0.440	0.730	0.790	0.440	0.487	0.472	0.429	0.511
Voraus	0.240	0.272	0.265	0.240	0.700	0.850	0.240	0.263	0.263	0.242	0.242
macro	*0.456*	*0.414*	*0.391*	*0.456*	*0.685*	*0.793*	*0.456*	*0.423*	*0.407*	*0.396*	*0.389*

Table 36: Full metric grid for **CHARM + Purity**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.130	0.106	0.106	0.130	0.170	0.220	0.130	0.110	0.109	0.138	0.061
DAMADICS	0.056	0.156	0.189	0.056	0.278	0.333	0.056	0.133	0.164	0.191	0.075
Exathlon	0.750	0.728	0.719	0.750	0.850	0.910	0.750	0.735	0.727	0.688	0.512
HAI	0.250	0.271	0.246	0.250	0.429	0.571	0.250	0.267	0.252	0.254	0.143
MIT-BIH	0.380	0.446	0.469	0.380	0.580	0.660	0.380	0.431	0.453	0.476	0.275
Petrobras 3W	0.580	0.532	0.504	0.580	0.630	0.680	0.580	0.544	0.520	0.488	0.448
RATS40K	0.580	0.526	0.482	0.580	0.650	0.700	0.580	0.539	0.506	0.476	0.180
RCAEval	0.420	0.416	0.420	0.420	0.680	0.740	0.420	0.419	0.420	0.406	0.362
ROAD	0.480	0.480	0.484	0.480	0.480	0.520	0.480	0.480	0.483	0.460	0.130
TelecomTS	0.500	0.462	0.432	0.500	0.590	0.600	0.500	0.468	0.445	0.412	0.320
Tennessee Eastman	0.390	0.368	0.340	0.390	0.420	0.420	0.390	0.374	0.352	0.319	0.304
Voraus	0.200	0.200	0.193	0.200	0.410	0.480	0.200	0.198	0.195	0.189	0.110
macro	*0.393*	*0.391*	*0.382*	*0.393*	*0.514*	*0.570*	*0.393*	*0.392*	*0.386*	*0.375*	*0.243*
$\rho\approx$10%											
CTSR	0.130	0.100	0.106	0.130	0.170	0.220	0.130	0.105	0.108	0.136	0.049
DAMADICS	0.111	0.178	0.206	0.111	0.222	0.278	0.111	0.163	0.188	0.195	0.095
Exathlon	0.740	0.724	0.697	0.740	0.810	0.850	0.740	0.728	0.709	0.660	0.517
HAI	0.357	0.279	0.268	0.357	0.500	0.571	0.357	0.291	0.281	0.270	0.139
MIT-BIH	0.450	0.488	0.521	0.450	0.610	0.710	0.450	0.480	0.505	0.521	0.348
Petrobras 3W	0.520	0.484	0.465	0.520	0.570	0.610	0.520	0.492	0.476	0.454	0.329
RATS40K	0.530	0.492	0.453	0.530	0.650	0.690	0.530	0.502	0.474	0.444	0.167
RCAEval	0.440	0.416	0.412	0.440	0.620	0.740	0.440	0.421	0.419	0.400	0.332
ROAD	0.480	0.480	0.488	0.480	0.480	0.600	0.480	0.480	0.486	0.470	0.130
TelecomTS	0.500	0.468	0.437	0.500	0.570	0.590	0.500	0.476	0.451	0.414	0.325
Tennessee Eastman	0.390	0.366	0.340	0.390	0.420	0.420	0.390	0.372	0.352	0.318	0.285
Voraus	0.190	0.124	0.139	0.190	0.200	0.330	0.190	0.144	0.149	0.138	0.110
macro	*0.403*	*0.383*	*0.378*	*0.403*	*0.485*	*0.551*	*0.403*	*0.388*	*0.383*	*0.368*	*0.236*
$\rho\approx$20%											
CTSR	0.120	0.102	0.102	0.120	0.170	0.220	0.120	0.105	0.104	0.134	0.049
DAMADICS	0.167	0.178	0.189	0.167	0.222	0.278	0.167	0.170	0.182	0.196	0.095
Exathlon	0.730	0.674	0.631	0.730	0.770	0.820	0.730	0.687	0.653	0.574	0.528
HAI	0.250	0.271	0.250	0.250	0.536	0.536	0.250	0.268	0.254	0.250	0.106
MIT-BIH	0.380	0.380	0.377	0.380	0.490	0.570	0.380	0.380	0.378	0.387	0.200
Petrobras 3W	0.370	0.360	0.362	0.370	0.440	0.510	0.370	0.361	0.362	0.345	0.247
RATS40K	0.440	0.378	0.378	0.440	0.530	0.640	0.440	0.392	0.387	0.369	0.141
RCAEval	0.420	0.416	0.394	0.420	0.660	0.760	0.420	0.417	0.402	0.397	0.326
ROAD	0.480	0.472	0.472	0.480	0.480	0.600	0.480	0.475	0.474	0.436	0.130
TelecomTS	0.480	0.438	0.412	0.480	0.540	0.560	0.480	0.449	0.427	0.398	0.300
Tennessee Eastman	0.390	0.366	0.339	0.390	0.420	0.420	0.390	0.372	0.351	0.317	0.285
Voraus	0.010	0.010	0.005	0.010	0.010	0.010	0.010	0.010	0.007	0.009	0.083
macro	*0.353*	*0.337*	*0.326*	*0.353*	*0.439*	*0.494*	*0.353*	*0.341*	*0.332*	*0.318*	*0.207*

Table 37: Full metric grid for **CHARM + NR + GPC**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.661	0.499
DAMADICS	0.444	0.444	0.444	0.444	0.444	0.444	0.444	0.444	0.444	0.444	0.224
Exathlon	0.900	0.900	0.900	0.900	0.900	0.900	0.900	0.900	0.900	0.900	0.776
HAI	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.339
MIT-BIH	0.710	0.710	0.710	0.710	0.710	0.710	0.710	0.710	0.710	0.710	0.444
Petrobras 3W	0.650	0.650	0.650	0.650	0.650	0.650	0.650	0.650	0.650	0.650	0.544
RATS40K	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.208
RCAEval	0.640	0.640	0.640	0.640	0.640	0.640	0.640	0.640	0.640	0.640	0.604
ROAD	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.250
TelecomTS	0.730	0.730	0.730	0.730	0.730	0.730	0.730	0.730	0.730	0.730	0.624
Tennessee Eastman	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.528
Voraus	0.580	0.580	0.576	0.580	0.580	0.590	0.580	0.580	0.581	0.584	0.487
macro	*0.608*	*0.608*	*0.608*	*0.608*	*0.608*	*0.609*	*0.608*	*0.608*	*0.608*	*0.615*	*0.461*
$\rho\approx$10%											
CTSR	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.661	0.499
DAMADICS	0.444	0.444	0.444	0.444	0.444	0.444	0.444	0.444	0.444	0.444	0.224
Exathlon	0.900	0.900	0.900	0.900	0.900	0.900	0.900	0.900	0.900	0.900	0.776
HAI	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.339
MIT-BIH	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.393
Petrobras 3W	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.439
RATS40K	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.211
RCAEval	0.640	0.640	0.640	0.640	0.640	0.640	0.640	0.640	0.640	0.640	0.604
ROAD	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.568	0.250
TelecomTS	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.654
Tennessee Eastman	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.532
Voraus	0.580	0.580	0.576	0.580	0.580	0.590	0.580	0.580	0.581	0.584	0.487
macro	*0.608*	*0.608*	*0.608*	*0.608*	*0.608*	*0.609*	*0.608*	*0.608*	*0.608*	*0.616*	*0.451*
$\rho\approx$20%											
CTSR	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.656	0.497
DAMADICS	0.389	0.389	0.389	0.389	0.389	0.389	0.389	0.389	0.389	0.389	0.226
Exathlon	0.900	0.900	0.900	0.900	0.900	0.900	0.900	0.900	0.900	0.900	0.776
HAI	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.339
MIT-BIH	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.285
Petrobras 3W	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.397
RATS40K	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.215
RCAEval	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.587
ROAD	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.568	0.250
TelecomTS	0.760	0.760	0.760	0.760	0.760	0.760	0.760	0.760	0.760	0.760	0.673
Tennessee Eastman	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.535
Voraus	0.580	0.580	0.576	0.580	0.580	0.590	0.580	0.580	0.581	0.584	0.490
macro	*0.578*	*0.578*	*0.577*	*0.578*	*0.578*	*0.578*	*0.578*	*0.578*	*0.578*	*0.585*	*0.439*

Table 38: Full metric grid for **CHARM + NR + MajVote**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.632	0.520
DAMADICS	0.556	0.556	0.522	0.556	0.556	0.722	0.556	0.556	0.566	0.596	0.342
Exathlon	0.920	0.920	0.920	0.920	0.920	0.920	0.920	0.920	0.920	0.920	0.858
HAI	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.374
MIT-BIH	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.355
Petrobras 3W	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.505
RATS40K	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.182
RCAEval	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.657
ROAD	0.520	0.520	0.528	0.520	0.520	0.560	0.520	0.520	0.526	0.549	0.236
TelecomTS	0.770	0.770	0.770	0.770	0.770	0.770	0.770	0.770	0.770	0.770	0.686
Tennessee Eastman	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.488
Voraus	0.540	0.540	0.537	0.540	0.540	0.570	0.540	0.540	0.542	0.545	0.472
macro	*0.610*	*0.610*	*0.608*	*0.610*	*0.610*	*0.630*	*0.610*	*0.610*	*0.612*	*0.621*	*0.473*
$\rho\approx$10%											
CTSR	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.632	0.520
DAMADICS	0.556	0.556	0.517	0.556	0.556	0.667	0.556	0.556	0.563	0.596	0.342
Exathlon	0.920	0.920	0.920	0.920	0.920	0.920	0.920	0.920	0.920	0.920	0.858
HAI	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.374
MIT-BIH	0.760	0.760	0.760	0.760	0.760	0.760	0.760	0.760	0.760	0.760	0.398
Petrobras 3W	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.442
RATS40K	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.176
RCAEval	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.661
ROAD	0.520	0.520	0.528	0.520	0.520	0.560	0.520	0.520	0.526	0.549	0.236
TelecomTS	0.790	0.790	0.790	0.790	0.790	0.790	0.790	0.790	0.790	0.790	0.701
Tennessee Eastman	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.479
Voraus	0.550	0.550	0.547	0.550	0.550	0.580	0.550	0.550	0.552	0.555	0.477
macro	*0.613*	*0.613*	*0.610*	*0.613*	*0.613*	*0.628*	*0.613*	*0.613*	*0.614*	*0.623*	*0.472*
$\rho\approx$20%											
CTSR	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.631	0.520
DAMADICS	0.444	0.444	0.406	0.444	0.444	0.556	0.444	0.444	0.452	0.478	0.303
Exathlon	0.910	0.910	0.910	0.910	0.910	0.910	0.910	0.910	0.910	0.910	0.853
HAI	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.374
MIT-BIH	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.293
Petrobras 3W	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.391
RATS40K	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.182
RCAEval	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.624
ROAD	0.520	0.520	0.528	0.520	0.520	0.560	0.520	0.520	0.526	0.549	0.236
TelecomTS	0.780	0.780	0.780	0.780	0.780	0.780	0.780	0.780	0.780	0.780	0.694
Tennessee Eastman	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.498
Voraus	0.550	0.550	0.547	0.550	0.550	0.580	0.550	0.550	0.552	0.555	0.477
macro	*0.578*	*0.578*	*0.575*	*0.578*	*0.578*	*0.593*	*0.578*	*0.578*	*0.579*	*0.588*	*0.454*

Table 39: Full metric grid for **CHARM + NR + QPurity**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.600	0.522	0.430	0.600	0.730	0.860	0.600	0.540	0.470	0.519	0.529
DAMADICS	0.389	0.444	0.472	0.389	0.778	0.944	0.389	0.429	0.461	0.481	0.205
Exathlon	0.890	0.854	0.838	0.890	0.930	0.960	0.890	0.861	0.847	0.829	0.624
HAI	0.464	0.436	0.414	0.464	0.607	0.750	0.464	0.443	0.427	0.391	0.394
MIT-BIH	0.570	0.616	0.653	0.570	0.720	0.830	0.570	0.609	0.637	0.674	0.348
Petrobras 3W	0.550	0.538	0.515	0.550	0.670	0.740	0.550	0.541	0.524	0.498	0.458
RATS40K	0.540	0.478	0.462	0.540	0.660	0.730	0.540	0.488	0.473	0.454	0.172
RCAEval	0.640	0.636	0.588	0.640	0.760	0.860	0.640	0.639	0.604	0.567	0.635
ROAD	0.520	0.552	0.532	0.520	0.800	0.840	0.520	0.543	0.545	0.540	0.264
TelecomTS	0.830	0.716	0.669	0.830	0.930	0.950	0.830	0.741	0.700	0.650	0.672
Tennessee Eastman	0.520	0.526	0.519	0.520	0.610	0.680	0.520	0.525	0.521	0.510	0.469
Voraus	0.550	0.512	0.460	0.550	0.740	0.840	0.550	0.521	0.485	0.439	0.407
macro	*0.589*	*0.569*	*0.546*	*0.589*	*0.745*	*0.832*	*0.589*	*0.573*	*0.558*	*0.546*	*0.431*
$\rho\approx$10%											
CTSR	0.600	0.552	0.466	0.600	0.710	0.870	0.600	0.563	0.500	0.548	0.544
DAMADICS	0.389	0.444	0.439	0.389	0.889	0.944	0.389	0.439	0.452	0.471	0.363
Exathlon	0.880	0.866	0.846	0.880	0.940	0.950	0.880	0.869	0.854	0.834	0.630
HAI	0.429	0.414	0.393	0.429	0.536	0.643	0.429	0.423	0.407	0.402	0.374
MIT-BIH	0.700	0.738	0.740	0.700	0.760	0.770	0.700	0.732	0.736	0.743	0.398
Petrobras 3W	0.520	0.500	0.488	0.520	0.670	0.730	0.520	0.505	0.495	0.471	0.384
RATS40K	0.480	0.470	0.452	0.480	0.640	0.740	0.480	0.474	0.461	0.442	0.171
RCAEval	0.640	0.648	0.612	0.640	0.740	0.840	0.640	0.646	0.622	0.576	0.630
ROAD	0.560	0.512	0.540	0.560	0.640	0.840	0.560	0.522	0.553	0.546	0.264
TelecomTS	0.830	0.718	0.680	0.830	0.940	0.940	0.830	0.743	0.708	0.664	0.658
Tennessee Eastman	0.500	0.520	0.519	0.500	0.600	0.680	0.500	0.516	0.517	0.508	0.450
Voraus	0.550	0.524	0.463	0.550	0.740	0.840	0.550	0.530	0.489	0.439	0.424
macro	*0.590*	*0.576*	*0.553*	*0.590*	*0.734*	*0.816*	*0.590*	*0.580*	*0.566*	*0.554*	*0.441*
$\rho\approx$20%											
CTSR	0.600	0.556	0.466	0.600	0.720	0.860	0.600	0.566	0.502	0.549	0.559
DAMADICS	0.389	0.356	0.400	0.389	0.722	0.944	0.389	0.369	0.415	0.441	0.282
Exathlon	0.860	0.836	0.823	0.860	0.920	0.950	0.860	0.842	0.831	0.815	0.614
HAI	0.429	0.414	0.400	0.429	0.536	0.643	0.429	0.424	0.411	0.397	0.378
MIT-BIH	0.550	0.550	0.543	0.550	0.590	0.590	0.550	0.550	0.545	0.552	0.296
Petrobras 3W	0.500	0.462	0.451	0.500	0.660	0.730	0.500	0.470	0.460	0.436	0.339
RATS40K	0.460	0.448	0.438	0.460	0.630	0.740	0.460	0.454	0.446	0.426	0.174
RCAEval	0.600	0.628	0.610	0.600	0.720	0.820	0.600	0.622	0.612	0.566	0.654
ROAD	0.560	0.520	0.532	0.560	0.640	0.840	0.560	0.527	0.547	0.532	0.264
TelecomTS	0.830	0.712	0.680	0.830	0.950	0.960	0.830	0.739	0.707	0.663	0.625
Tennessee Eastman	0.500	0.508	0.508	0.500	0.580	0.630	0.500	0.508	0.508	0.501	0.451
Voraus	0.560	0.472	0.405	0.560	0.800	0.900	0.560	0.490	0.440	0.389	0.467
macro	*0.570*	*0.538*	*0.521*	*0.570*	*0.706*	*0.801*	*0.570*	*0.547*	*0.535*	*0.522*	*0.425*

Table 40: Full metric grid for **CHARM + NR + PRF**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.530	0.480	0.418	0.530	0.800	0.850	0.530	0.495	0.447	0.523	0.433
DAMADICS	0.556	0.511	0.461	0.556	0.833	1.000	0.556	0.518	0.510	0.530	0.354
Exathlon	0.870	0.846	0.838	0.870	0.970	1.000	0.870	0.854	0.845	0.806	0.749
HAI	0.393	0.371	0.329	0.393	0.643	0.786	0.393	0.372	0.343	0.343	0.446
MIT-BIH	0.700	0.674	0.691	0.700	0.760	0.830	0.700	0.679	0.689	0.704	0.360
Petrobras 3W	0.640	0.526	0.487	0.640	0.760	0.830	0.640	0.551	0.516	0.473	0.502
RATS40K	0.430	0.402	0.373	0.430	0.730	0.790	0.430	0.408	0.386	0.380	0.242
RCAEval	0.540	0.536	0.500	0.540	0.880	0.940	0.540	0.547	0.519	0.473	0.635
ROAD	0.480	0.504	0.448	0.480	0.840	0.880	0.480	0.519	0.497	0.495	0.246
TelecomTS	0.850	0.682	0.584	0.850	0.950	0.980	0.850	0.720	0.638	0.576	0.788
Tennessee Eastman	0.550	0.540	0.536	0.550	0.730	0.770	0.550	0.543	0.539	0.513	0.488
Voraus	0.550	0.432	0.371	0.550	0.880	0.940	0.550	0.457	0.410	0.363	0.548
macro	*0.591*	*0.542*	*0.503*	*0.591*	*0.815*	*0.883*	*0.591*	*0.555*	*0.528*	*0.515*	*0.483*
$\rho\approx$10%											
CTSR	0.530	0.482	0.421	0.530	0.800	0.850	0.530	0.496	0.450	0.524	0.447
DAMADICS	0.500	0.467	0.417	0.500	0.833	0.944	0.500	0.475	0.466	0.491	0.333
Exathlon	0.860	0.838	0.829	0.860	0.970	1.000	0.860	0.845	0.836	0.797	0.749
HAI	0.393	0.371	0.329	0.393	0.643	0.786	0.393	0.372	0.343	0.341	0.446
MIT-BIH	0.740	0.748	0.747	0.740	0.770	0.780	0.740	0.747	0.747	0.749	0.398
Petrobras 3W	0.620	0.510	0.468	0.620	0.760	0.830	0.620	0.534	0.497	0.456	0.507
RATS40K	0.420	0.400	0.369	0.420	0.730	0.800	0.420	0.406	0.382	0.365	0.234
RCAEval	0.560	0.532	0.488	0.560	0.860	0.940	0.560	0.541	0.509	0.463	0.618
ROAD	0.480	0.496	0.444	0.480	0.840	0.880	0.480	0.510	0.491	0.481	0.246
TelecomTS	0.870	0.676	0.575	0.870	0.940	0.970	0.870	0.716	0.631	0.571	0.765
Tennessee Eastman	0.540	0.534	0.529	0.540	0.710	0.760	0.540	0.536	0.532	0.508	0.498
Voraus	0.550	0.432	0.369	0.550	0.880	0.940	0.550	0.457	0.408	0.362	0.548
macro	*0.589*	*0.541*	*0.499*	*0.589*	*0.811*	*0.873*	*0.589*	*0.553*	*0.524*	*0.509*	*0.482*
$\rho\approx$20%											
CTSR	0.530	0.482	0.420	0.530	0.800	0.850	0.530	0.496	0.449	0.523	0.447
DAMADICS	0.500	0.456	0.394	0.500	0.778	0.889	0.500	0.465	0.447	0.464	0.333
Exathlon	0.840	0.836	0.813	0.840	0.970	1.000	0.840	0.839	0.822	0.786	0.749
HAI	0.357	0.350	0.325	0.357	0.607	0.786	0.357	0.350	0.333	0.327	0.331
MIT-BIH	0.540	0.542	0.538	0.540	0.600	0.620	0.540	0.541	0.538	0.553	0.292
Petrobras 3W	0.590	0.480	0.448	0.590	0.740	0.830	0.590	0.506	0.475	0.435	0.471
RATS40K	0.420	0.390	0.357	0.420	0.750	0.800	0.420	0.399	0.373	0.358	0.242
RCAEval	0.540	0.520	0.484	0.540	0.860	0.920	0.540	0.529	0.502	0.451	0.603
ROAD	0.520	0.488	0.432	0.520	0.840	0.880	0.520	0.509	0.487	0.472	0.246
TelecomTS	0.860	0.664	0.574	0.860	0.950	0.970	0.860	0.706	0.629	0.567	0.746
Tennessee Eastman	0.540	0.530	0.527	0.540	0.710	0.740	0.540	0.533	0.530	0.502	0.500
Voraus	0.550	0.430	0.369	0.550	0.880	0.940	0.550	0.455	0.408	0.361	0.548
macro	*0.566*	*0.514*	*0.473*	*0.566*	*0.790*	*0.852*	*0.566*	*0.527*	*0.499*	*0.483*	*0.459*

Table 41: Full metric grid for **CHARM + NR + Purity**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.280	0.230	0.206	0.280	0.460	0.490	0.280	0.236	0.218	0.259	0.148
DAMADICS	0.167	0.344	0.372	0.167	0.611	0.611	0.167	0.314	0.346	0.344	0.222
Exathlon	0.890	0.796	0.755	0.890	0.910	0.940	0.890	0.814	0.779	0.733	0.576
HAI	0.286	0.329	0.318	0.286	0.571	0.750	0.286	0.320	0.315	0.313	0.211
MIT-BIH	0.410	0.578	0.638	0.410	0.790	0.870	0.410	0.547	0.600	0.649	0.291
Petrobras 3W	0.560	0.528	0.512	0.560	0.660	0.710	0.560	0.535	0.522	0.478	0.431
RATS40K	0.510	0.494	0.453	0.510	0.670	0.720	0.510	0.499	0.470	0.426	0.186
RCAEval	0.440	0.416	0.372	0.440	0.600	0.740	0.440	0.424	0.390	0.365	0.462
ROAD	0.560	0.512	0.504	0.560	0.560	0.600	0.560	0.521	0.512	0.497	0.133
TelecomTS	0.740	0.630	0.581	0.740	0.840	0.880	0.740	0.651	0.609	0.568	0.518
Tennessee Eastman	0.450	0.414	0.409	0.450	0.550	0.570	0.450	0.422	0.416	0.391	0.307
Voraus	0.330	0.370	0.354	0.330	0.660	0.750	0.330	0.363	0.356	0.330	0.290
macro	*0.469*	*0.470*	*0.456*	*0.469*	*0.657*	*0.719*	*0.469*	*0.471*	*0.461*	*0.446*	*0.315*
$\rho\approx$10%											
CTSR	0.280	0.234	0.206	0.280	0.460	0.490	0.280	0.239	0.218	0.260	0.136
DAMADICS	0.222	0.322	0.389	0.222	0.611	0.833	0.222	0.309	0.359	0.398	0.153
Exathlon	0.860	0.772	0.742	0.860	0.880	0.930	0.860	0.791	0.764	0.718	0.562
HAI	0.286	0.314	0.304	0.286	0.607	0.714	0.286	0.306	0.302	0.305	0.250
MIT-BIH	0.590	0.684	0.705	0.590	0.820	0.870	0.590	0.665	0.687	0.703	0.369
Petrobras 3W	0.530	0.504	0.485	0.530	0.640	0.700	0.530	0.511	0.496	0.455	0.419
RATS40K	0.510	0.488	0.443	0.510	0.680	0.710	0.510	0.495	0.461	0.418	0.166
RCAEval	0.440	0.424	0.400	0.440	0.720	0.840	0.440	0.437	0.415	0.376	0.481
ROAD	0.560	0.512	0.508	0.560	0.560	0.600	0.560	0.521	0.515	0.500	0.213
TelecomTS	0.720	0.626	0.577	0.720	0.840	0.870	0.720	0.645	0.605	0.563	0.477
Tennessee Eastman	0.450	0.412	0.406	0.450	0.540	0.570	0.450	0.422	0.414	0.387	0.305
Voraus	0.330	0.350	0.331	0.330	0.630	0.710	0.330	0.349	0.338	0.317	0.294
macro	*0.481*	*0.470*	*0.458*	*0.481*	*0.666*	*0.736*	*0.481*	*0.474*	*0.464*	*0.450*	*0.319*
$\rho\approx$20%											
CTSR	0.280	0.232	0.205	0.280	0.450	0.480	0.280	0.237	0.217	0.257	0.148
DAMADICS	0.222	0.289	0.306	0.222	0.667	0.833	0.222	0.273	0.294	0.342	0.220
Exathlon	0.810	0.754	0.731	0.810	0.880	0.930	0.810	0.768	0.748	0.703	0.547
HAI	0.321	0.314	0.311	0.321	0.571	0.714	0.321	0.319	0.315	0.296	0.224
MIT-BIH	0.490	0.502	0.535	0.490	0.630	0.730	0.490	0.499	0.522	0.550	0.255
Petrobras 3W	0.450	0.442	0.424	0.450	0.560	0.640	0.450	0.444	0.431	0.404	0.343
RATS40K	0.470	0.440	0.404	0.470	0.640	0.690	0.470	0.449	0.421	0.386	0.164
RCAEval	0.460	0.420	0.394	0.460	0.720	0.800	0.460	0.431	0.411	0.388	0.429
ROAD	0.520	0.488	0.480	0.520	0.560	0.560	0.520	0.498	0.489	0.471	0.137
TelecomTS	0.710	0.614	0.572	0.710	0.820	0.850	0.710	0.634	0.598	0.555	0.507
Tennessee Eastman	0.450	0.410	0.403	0.450	0.530	0.560	0.450	0.420	0.411	0.383	0.320
Voraus	0.280	0.284	0.269	0.280	0.580	0.670	0.280	0.281	0.273	0.263	0.244
macro	*0.455*	*0.432*	*0.419*	*0.455*	*0.634*	*0.705*	*0.455*	*0.438*	*0.427*	*0.416*	*0.295*

Table 42: Full metric grid for **DTW-I + GPC**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.690	0.690	0.690	0.690	0.690	0.690	0.690	0.690	0.690	0.765	0.624
DAMADICS	0.611	0.611	0.594	0.611	0.611	0.667	0.611	0.611	0.622	0.638	0.416
Exathlon	0.330	0.330	0.330	0.330	0.330	0.330	0.330	0.330	0.330	0.330	0.322
HAI	0.214	0.214	0.200	0.214	0.214	0.214	0.214	0.214	0.214	0.214	0.230
MIT-BIH	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.544
Petrobras 3W	0.900	0.900	0.900	0.900	0.900	0.900	0.900	0.900	0.900	0.904	0.879
RATS40K	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.218
RCAEval	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.358
ROAD	0.480	0.480	0.492	0.480	0.480	0.600	0.480	0.480	0.501	0.576	0.319
TelecomTS	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.629
Tennessee Eastman	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.727
Voraus	0.760	0.760	0.756	0.760	0.760	0.760	0.760	0.760	0.760	0.760	0.657
macro	*0.570*	*0.570*	*0.569*	*0.570*	*0.570*	*0.585*	*0.570*	*0.570*	*0.573*	*0.587*	*0.494*
$\rho\approx$10%											
CTSR	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.741	0.591
DAMADICS	0.111	0.111	0.094	0.111	0.111	0.111	0.111	0.111	0.111	0.111	0.183
Exathlon	0.330	0.330	0.330	0.330	0.330	0.330	0.330	0.330	0.330	0.330	0.386
HAI	0.143	0.143	0.136	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.159
MIT-BIH	0.720	0.720	0.720	0.720	0.720	0.720	0.720	0.720	0.720	0.720	0.617
Petrobras 3W	0.880	0.880	0.880	0.880	0.880	0.880	0.880	0.880	0.880	0.884	0.873
RATS40K	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.206
RCAEval	0.360	0.360	0.360	0.360	0.360	0.360	0.360	0.360	0.360	0.360	0.337
ROAD	0.520	0.520	0.528	0.520	0.520	0.640	0.520	0.520	0.538	0.612	0.335
TelecomTS	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.576
Tennessee Eastman	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.626
Voraus	0.770	0.770	0.766	0.770	0.770	0.770	0.770	0.770	0.770	0.770	0.663
macro	*0.515*	*0.515*	*0.514*	*0.515*	*0.515*	*0.525*	*0.515*	*0.515*	*0.517*	*0.529*	*0.463*
$\rho\approx$20%											
CTSR	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.741	0.589
DAMADICS	0.056	0.056	0.039	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.100
Exathlon	0.330	0.330	0.330	0.330	0.330	0.330	0.330	0.330	0.330	0.330	0.341
HAI	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.022
MIT-BIH	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.292
Petrobras 3W	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.866
RATS40K	0.430	0.430	0.430	0.430	0.430	0.430	0.430	0.430	0.430	0.430	0.218
RCAEval	0.240	0.240	0.240	0.240	0.240	0.240	0.240	0.240	0.240	0.240	0.305
ROAD	0.520	0.520	0.528	0.520	0.520	0.640	0.520	0.520	0.538	0.609	0.335
TelecomTS	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.604
Tennessee Eastman	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.543
Voraus	0.740	0.740	0.736	0.740	0.740	0.740	0.740	0.740	0.740	0.740	0.651
macro	*0.463*	*0.463*	*0.462*	*0.463*	*0.463*	*0.473*	*0.463*	*0.463*	*0.464*	*0.476*	*0.406*

Table 43: Full metric grid for **DTW-I + MajVote**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.662	0.539
DAMADICS	0.556	0.556	0.544	0.556	0.556	0.611	0.556	0.556	0.559	0.572	0.346
Exathlon	0.350	0.318	0.307	0.350	0.400	0.500	0.350	0.325	0.315	0.329	0.372
HAI	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.033
MIT-BIH	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.535
Petrobras 3W	0.930	0.874	0.850	0.930	0.940	0.960	0.930	0.886	0.865	0.823	0.873
RATS40K	0.470	0.470	0.470	0.470	0.470	0.470	0.470	0.470	0.470	0.470	0.163
RCAEval	0.340	0.284	0.294	0.340	0.760	0.820	0.340	0.293	0.297	0.290	0.223
ROAD	0.440	0.440	0.468	0.440	0.440	0.760	0.440	0.440	0.480	0.528	0.290
TelecomTS	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.618	0.554
Tennessee Eastman	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.691
Voraus	0.720	0.720	0.716	0.720	0.720	0.720	0.720	0.720	0.720	0.720	0.641
macro	*0.533*	*0.521*	*0.520*	*0.533*	*0.573*	*0.619*	*0.533*	*0.523*	*0.525*	*0.531*	*0.438*
$\rho\approx$10%											
CTSR	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.661	0.539
DAMADICS	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.100
Exathlon	0.310	0.280	0.258	0.310	0.370	0.460	0.310	0.288	0.269	0.286	0.359
HAI	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.035
MIT-BIH	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.527
Petrobras 3W	0.850	0.836	0.805	0.850	0.940	0.960	0.850	0.842	0.818	0.782	0.879
RATS40K	0.450	0.450	0.450	0.450	0.450	0.450	0.450	0.450	0.450	0.450	0.161
RCAEval	0.180	0.260	0.260	0.180	0.560	0.920	0.180	0.253	0.257	0.264	0.254
ROAD	0.440	0.440	0.460	0.440	0.440	0.760	0.440	0.440	0.475	0.525	0.290
TelecomTS	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.598	0.556
Tennessee Eastman	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.622
Voraus	0.740	0.740	0.736	0.740	0.740	0.740	0.740	0.740	0.740	0.740	0.654
macro	*0.465*	*0.468*	*0.465*	*0.465*	*0.509*	*0.575*	*0.465*	*0.468*	*0.468*	*0.475*	*0.415*
$\rho\approx$20%											
CTSR	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.661	0.539
DAMADICS	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Exathlon	0.280	0.264	0.233	0.280	0.370	0.410	0.280	0.271	0.247	0.258	0.354
HAI	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.020
MIT-BIH	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.301
Petrobras 3W	0.820	0.810	0.778	0.820	0.940	0.950	0.820	0.813	0.790	0.762	0.882
RATS40K	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.138
RCAEval	0.220	0.244	0.228	0.220	0.760	0.920	0.220	0.245	0.232	0.231	0.223
ROAD	0.440	0.440	0.460	0.440	0.440	0.760	0.440	0.440	0.475	0.522	0.290
TelecomTS	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.569	0.544
Tennessee Eastman	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.534
Voraus	0.710	0.710	0.706	0.710	0.710	0.710	0.710	0.710	0.710	0.710	0.638
macro	*0.425*	*0.424*	*0.419*	*0.425*	*0.487*	*0.531*	*0.425*	*0.425*	*0.423*	*0.430*	*0.372*

Table 44: Full metric grid for **DTW-I + QPurity**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.580	0.534	0.435	0.580	0.710	0.810	0.580	0.547	0.475	0.507	0.532
DAMADICS	0.500	0.533	0.533	0.500	0.556	0.556	0.500	0.525	0.528	0.530	0.300
Exathlon	0.340	0.348	0.413	0.340	0.350	0.550	0.340	0.347	0.392	0.431	0.212
HAI	0.071	0.100	0.136	0.071	0.179	0.464	0.071	0.093	0.121	0.149	0.032
MIT-BIH	0.610	0.616	0.633	0.610	0.630	0.700	0.610	0.614	0.626	0.633	0.535
Petrobras 3W	0.840	0.826	0.819	0.840	0.850	0.890	0.840	0.827	0.822	0.782	0.731
RATS40K	0.480	0.474	0.460	0.480	0.480	0.500	0.480	0.475	0.465	0.457	0.169
RCAEval	0.200	0.200	0.200	0.200	0.800	0.800	0.200	0.200	0.200	0.200	0.067
ROAD	0.640	0.528	0.472	0.640	0.840	0.880	0.640	0.564	0.531	0.516	0.480
TelecomTS	0.600	0.586	0.567	0.600	0.630	0.700	0.600	0.589	0.575	0.547	0.486
Tennessee Eastman	0.680	0.646	0.634	0.680	0.800	0.840	0.680	0.654	0.643	0.625	0.689
Voraus	0.690	0.642	0.601	0.690	0.730	0.760	0.690	0.654	0.623	0.581	0.504
macro	*0.519*	*0.503*	*0.492*	*0.519*	*0.630*	*0.704*	*0.519*	*0.507*	*0.500*	*0.497*	*0.395*
$\rho\approx$10%											
CTSR	0.600	0.580	0.490	0.600	0.680	0.790	0.600	0.586	0.523	0.556	0.537
DAMADICS	0.722	0.633	0.600	0.722	0.722	0.833	0.722	0.660	0.655	0.668	0.511
Exathlon	0.290	0.300	0.326	0.290	0.310	0.430	0.290	0.298	0.316	0.354	0.208
HAI	0.071	0.100	0.121	0.071	0.214	0.429	0.071	0.097	0.114	0.137	0.033
MIT-BIH	0.700	0.702	0.715	0.700	0.710	0.780	0.700	0.702	0.711	0.716	0.525
Petrobras 3W	0.840	0.824	0.814	0.840	0.880	0.900	0.840	0.828	0.820	0.791	0.753
RATS40K	0.450	0.450	0.436	0.450	0.450	0.460	0.450	0.450	0.440	0.436	0.161
RCAEval	0.000	0.120	0.160	0.000	0.400	1.000	0.000	0.103	0.137	0.159	0.000
ROAD	0.680	0.512	0.452	0.680	0.840	0.960	0.680	0.561	0.523	0.514	0.345
TelecomTS	0.600	0.572	0.565	0.600	0.620	0.700	0.600	0.577	0.571	0.539	0.492
Tennessee Eastman	0.580	0.598	0.593	0.580	0.790	0.860	0.580	0.593	0.591	0.580	0.644
Voraus	0.710	0.642	0.607	0.710	0.760	0.780	0.710	0.658	0.630	0.585	0.531
macro	*0.520*	*0.503*	*0.490*	*0.520*	*0.615*	*0.743*	*0.520*	*0.509*	*0.503*	*0.503*	*0.395*
$\rho\approx$20%											
CTSR	0.610	0.578	0.491	0.610	0.680	0.790	0.610	0.586	0.525	0.557	0.537
DAMADICS	0.000	0.044	0.167	0.000	0.222	0.722	0.000	0.029	0.133	0.294	0.000
Exathlon	0.300	0.302	0.309	0.300	0.310	0.380	0.300	0.301	0.306	0.356	0.216
HAI	0.071	0.093	0.100	0.071	0.250	0.357	0.071	0.084	0.092	0.112	0.029
MIT-BIH	0.520	0.534	0.548	0.520	0.560	0.610	0.520	0.530	0.541	0.550	0.301
Petrobras 3W	0.820	0.788	0.777	0.820	0.840	0.850	0.820	0.795	0.784	0.726	0.658
RATS40K	0.370	0.372	0.361	0.370	0.380	0.380	0.370	0.372	0.364	0.357	0.138
RCAEval	0.200	0.160	0.140	0.200	0.800	1.000	0.200	0.171	0.151	0.168	0.067
ROAD	0.680	0.520	0.444	0.680	0.840	0.920	0.680	0.567	0.518	0.508	0.358
TelecomTS	0.560	0.536	0.521	0.560	0.560	0.620	0.560	0.541	0.529	0.496	0.474
Tennessee Eastman	0.490	0.492	0.488	0.490	0.520	0.530	0.490	0.493	0.490	0.482	0.545
Voraus	0.620	0.580	0.531	0.620	0.690	0.770	0.620	0.592	0.555	0.495	0.472
macro	*0.437*	*0.417*	*0.406*	*0.437*	*0.554*	*0.661*	*0.437*	*0.422*	*0.416*	*0.425*	*0.316*

Table 45: Full metric grid for **DTW-I + PRF**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.730	0.576	0.474	0.730	0.840	0.890	0.730	0.607	0.526	0.584	0.610
DAMADICS	0.667	0.622	0.583	0.667	0.722	0.833	0.667	0.641	0.635	0.649	0.506
Exathlon	0.350	0.318	0.307	0.350	0.400	0.500	0.350	0.325	0.315	0.329	0.372
HAI	0.107	0.121	0.154	0.107	0.464	0.786	0.107	0.115	0.141	0.171	0.032
MIT-BIH	0.500	0.576	0.574	0.500	0.850	0.910	0.500	0.562	0.566	0.564	0.458
Petrobras 3W	0.930	0.876	0.847	0.930	0.940	0.960	0.930	0.888	0.864	0.823	0.879
RATS40K	0.410	0.418	0.403	0.410	0.730	0.820	0.410	0.420	0.412	0.398	0.171
RCAEval	0.340	0.284	0.294	0.340	0.760	0.820	0.340	0.293	0.297	0.290	0.223
ROAD	0.680	0.496	0.340	0.680	0.920	0.960	0.680	0.536	0.437	0.455	0.416
TelecomTS	0.620	0.510	0.476	0.620	0.820	0.890	0.620	0.534	0.502	0.447	0.524
Tennessee Eastman	0.590	0.548	0.480	0.590	0.890	0.940	0.590	0.563	0.510	0.458	0.661
Voraus	0.800	0.630	0.535	0.800	0.980	0.990	0.800	0.668	0.591	0.505	0.769
macro	*0.560*	*0.498*	*0.456*	*0.560*	*0.776*	*0.858*	*0.560*	*0.513*	*0.483*	*0.473*	*0.468*
$\rho\approx$10%											
CTSR	0.730	0.574	0.469	0.730	0.840	0.890	0.730	0.605	0.522	0.578	0.622
DAMADICS	0.222	0.122	0.122	0.222	0.278	0.500	0.222	0.132	0.153	0.200	0.083
Exathlon	0.310	0.280	0.258	0.310	0.370	0.460	0.310	0.288	0.269	0.286	0.359
HAI	0.071	0.107	0.129	0.071	0.464	0.714	0.071	0.097	0.117	0.145	0.037
MIT-BIH	0.680	0.686	0.659	0.680	0.900	0.940	0.680	0.685	0.666	0.665	0.543
Petrobras 3W	0.820	0.836	0.801	0.820	0.940	0.960	0.820	0.838	0.813	0.780	0.879
RATS40K	0.380	0.376	0.369	0.380	0.700	0.810	0.380	0.382	0.379	0.367	0.157
RCAEval	0.180	0.260	0.262	0.180	0.560	0.920	0.180	0.254	0.258	0.264	0.254
ROAD	0.680	0.488	0.336	0.680	0.920	0.960	0.680	0.529	0.433	0.440	0.440
TelecomTS	0.600	0.478	0.434	0.600	0.810	0.870	0.600	0.504	0.464	0.400	0.550
Tennessee Eastman	0.550	0.508	0.450	0.550	0.840	0.920	0.550	0.521	0.477	0.422	0.618
Voraus	0.790	0.610	0.516	0.790	0.970	0.980	0.790	0.649	0.573	0.489	0.762
macro	*0.501*	*0.444*	*0.400*	*0.501*	*0.716*	*0.827*	*0.501*	*0.457*	*0.427*	*0.420*	*0.442*
$\rho\approx$20%											
CTSR	0.730	0.572	0.468	0.730	0.840	0.890	0.730	0.603	0.521	0.577	0.622
DAMADICS	0.222	0.078	0.094	0.222	0.222	0.278	0.222	0.100	0.127	0.137	0.100
Exathlon	0.280	0.264	0.233	0.280	0.370	0.410	0.280	0.271	0.247	0.258	0.354
HAI	0.071	0.064	0.096	0.071	0.286	0.607	0.071	0.065	0.089	0.126	0.024
MIT-BIH	0.350	0.478	0.516	0.350	0.820	0.920	0.350	0.443	0.482	0.514	0.377
Petrobras 3W	0.790	0.798	0.770	0.790	0.940	0.950	0.790	0.797	0.778	0.756	0.864
RATS40K	0.350	0.352	0.344	0.350	0.700	0.800	0.350	0.356	0.351	0.333	0.139
RCAEval	0.220	0.248	0.228	0.220	0.760	0.920	0.220	0.247	0.232	0.232	0.231
ROAD	0.680	0.488	0.332	0.680	0.920	0.920	0.680	0.529	0.430	0.435	0.451
TelecomTS	0.570	0.470	0.412	0.570	0.790	0.860	0.570	0.493	0.444	0.374	0.523
Tennessee Eastman	0.520	0.486	0.426	0.520	0.760	0.890	0.520	0.495	0.450	0.398	0.609
Voraus	0.790	0.592	0.497	0.790	0.970	0.980	0.790	0.634	0.557	0.472	0.752
macro	*0.464*	*0.408*	*0.368*	*0.464*	*0.698*	*0.785*	*0.464*	*0.419*	*0.392*	*0.384*	*0.420*

Table 46: Full metric grid for **DTW-I + Purity**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.140	0.100	0.105	0.140	0.200	0.310	0.140	0.110	0.110	0.139	0.058
DAMADICS	0.500	0.500	0.494	0.500	0.556	0.611	0.500	0.499	0.496	0.476	0.295
Exathlon	0.330	0.386	0.376	0.330	0.600	0.710	0.330	0.383	0.378	0.377	0.288
HAI	0.107	0.150	0.182	0.107	0.286	0.571	0.107	0.139	0.165	0.168	0.046
MIT-BIH	0.500	0.574	0.570	0.500	0.830	0.930	0.500	0.560	0.563	0.563	0.403
Petrobras 3W	0.740	0.538	0.508	0.740	0.780	0.810	0.740	0.581	0.546	0.481	0.462
RATS40K	0.210	0.210	0.193	0.210	0.250	0.420	0.210	0.211	0.198	0.217	0.032
RCAEval	0.200	0.200	0.200	0.200	0.800	0.800	0.200	0.200	0.200	0.200	0.067
ROAD	0.600	0.536	0.516	0.600	0.680	0.680	0.600	0.547	0.531	0.489	0.130
TelecomTS	0.370	0.326	0.307	0.370	0.400	0.500	0.370	0.338	0.320	0.296	0.134
Tennessee Eastman	0.460	0.372	0.326	0.460	0.520	0.540	0.460	0.391	0.353	0.320	0.330
Voraus	0.300	0.260	0.237	0.300	0.500	0.600	0.300	0.265	0.248	0.238	0.125
macro	*0.371*	*0.346*	*0.335*	*0.371*	*0.533*	*0.624*	*0.371*	*0.352*	*0.342*	*0.330*	*0.197*
$\rho\approx$10%											
CTSR	0.150	0.098	0.103	0.150	0.200	0.310	0.150	0.110	0.109	0.140	0.058
DAMADICS	0.500	0.444	0.439	0.500	0.556	0.611	0.500	0.454	0.447	0.386	0.233
Exathlon	0.410	0.366	0.337	0.410	0.590	0.640	0.410	0.375	0.352	0.350	0.245
HAI	0.107	0.150	0.193	0.107	0.321	0.607	0.107	0.136	0.171	0.166	0.048
MIT-BIH	0.570	0.642	0.656	0.570	0.810	0.880	0.570	0.629	0.643	0.656	0.519
Petrobras 3W	0.670	0.516	0.457	0.670	0.740	0.760	0.670	0.546	0.495	0.458	0.392
RATS40K	0.210	0.166	0.162	0.210	0.290	0.420	0.210	0.179	0.170	0.196	0.012
RCAEval	0.000	0.120	0.160	0.000	0.400	1.000	0.000	0.103	0.137	0.159	0.000
ROAD	0.600	0.528	0.492	0.600	0.680	0.680	0.600	0.541	0.515	0.487	0.130
TelecomTS	0.370	0.328	0.302	0.370	0.400	0.430	0.370	0.340	0.317	0.296	0.134
Tennessee Eastman	0.470	0.366	0.322	0.470	0.520	0.530	0.470	0.387	0.349	0.318	0.313
Voraus	0.170	0.184	0.184	0.170	0.300	0.440	0.170	0.178	0.181	0.191	0.114
macro	*0.352*	*0.326*	*0.317*	*0.352*	*0.484*	*0.609*	*0.352*	*0.331*	*0.324*	*0.317*	*0.183*
$\rho\approx$20%											
CTSR	0.150	0.096	0.103	0.150	0.200	0.310	0.150	0.108	0.108	0.139	0.046
DAMADICS	0.500	0.444	0.439	0.500	0.556	0.667	0.500	0.455	0.448	0.382	0.233
Exathlon	0.270	0.298	0.302	0.270	0.530	0.610	0.270	0.290	0.296	0.286	0.270
HAI	0.107	0.179	0.204	0.107	0.321	0.607	0.107	0.163	0.186	0.179	0.054
MIT-BIH	0.360	0.466	0.518	0.360	0.740	0.850	0.360	0.439	0.485	0.522	0.293
Petrobras 3W	0.490	0.416	0.399	0.490	0.550	0.570	0.490	0.429	0.413	0.395	0.323
RATS40K	0.200	0.162	0.141	0.200	0.290	0.330	0.200	0.174	0.154	0.170	0.012
RCAEval	0.200	0.160	0.140	0.200	0.800	1.000	0.200	0.171	0.151	0.168	0.067
ROAD	0.600	0.520	0.488	0.600	0.680	0.680	0.600	0.538	0.513	0.480	0.130
TelecomTS	0.350	0.296	0.265	0.350	0.400	0.440	0.350	0.310	0.284	0.268	0.146
Tennessee Eastman	0.470	0.368	0.323	0.470	0.520	0.540	0.470	0.389	0.350	0.317	0.316
Voraus	0.070	0.044	0.042	0.070	0.100	0.170	0.070	0.049	0.047	0.045	0.099
macro	*0.314*	*0.287*	*0.280*	*0.314*	*0.474*	*0.564*	*0.314*	*0.293*	*0.286*	*0.279*	*0.166*

Table 47: Full metric grid for **MantisV2 + GPC**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.667	0.573
DAMADICS	0.611	0.611	0.628	0.611	0.611	0.833	0.611	0.611	0.655	0.719	0.359
Exathlon	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.964
HAI	0.429	0.429	0.421	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.494
MIT-BIH	0.340	0.340	0.340	0.340	0.340	0.340	0.340	0.340	0.340	0.340	0.295
Petrobras 3W	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.756
RATS40K	0.490	0.488	0.487	0.490	0.490	0.510	0.490	0.490	0.493	0.499	0.241
RCAEval	0.400	0.368	0.380	0.400	0.800	0.800	0.400	0.370	0.378	0.379	0.267
ROAD	0.640	0.640	0.628	0.640	0.640	0.640	0.640	0.640	0.640	0.640	0.488
TelecomTS	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.856
Tennessee Eastman	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.660
Voraus	0.360	0.360	0.358	0.360	0.360	0.360	0.360	0.360	0.360	0.360	0.295
macro	*0.602*	*0.599*	*0.599*	*0.602*	*0.635*	*0.655*	*0.602*	*0.599*	*0.604*	*0.613*	*0.521*
$\rho\approx$10%											
CTSR	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.650	0.524
DAMADICS	0.333	0.333	0.350	0.333	0.333	0.556	0.333	0.333	0.378	0.441	0.252
Exathlon	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.965
HAI	0.464	0.464	0.457	0.464	0.464	0.464	0.464	0.464	0.464	0.464	0.565
MIT-BIH	0.260	0.260	0.260	0.260	0.260	0.260	0.260	0.260	0.260	0.260	0.117
Petrobras 3W	0.830	0.830	0.830	0.830	0.830	0.830	0.830	0.830	0.830	0.830	0.720
RATS40K	0.470	0.468	0.464	0.470	0.470	0.480	0.470	0.470	0.471	0.474	0.253
RCAEval	0.240	0.352	0.364	0.240	0.600	0.920	0.240	0.338	0.352	0.365	0.315
ROAD	0.600	0.600	0.588	0.600	0.600	0.600	0.600	0.600	0.600	0.607	0.467
TelecomTS	0.850	0.850	0.850	0.850	0.850	0.850	0.850	0.850	0.850	0.850	0.850
Tennessee Eastman	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.612
Voraus	0.350	0.350	0.348	0.350	0.350	0.350	0.350	0.350	0.350	0.350	0.292
macro	*0.543*	*0.552*	*0.553*	*0.543*	*0.573*	*0.619*	*0.543*	*0.551*	*0.556*	*0.568*	*0.494*
$\rho\approx$20%											
CTSR	0.600	0.594	0.593	0.600	0.600	0.600	0.600	0.595	0.594	0.650	0.527
DAMADICS	0.222	0.222	0.239	0.222	0.222	0.444	0.222	0.222	0.267	0.330	0.182
Exathlon	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.965
HAI	0.429	0.429	0.421	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.551
MIT-BIH	0.210	0.210	0.210	0.210	0.210	0.210	0.210	0.210	0.210	0.210	0.102
Petrobras 3W	0.800	0.800	0.800	0.800	0.800	0.800	0.800	0.800	0.800	0.800	0.682
RATS40K	0.450	0.448	0.444	0.450	0.450	0.460	0.450	0.450	0.451	0.454	0.242
RCAEval	0.320	0.352	0.336	0.320	0.800	0.920	0.320	0.351	0.339	0.346	0.273
ROAD	0.600	0.600	0.588	0.600	0.600	0.600	0.600	0.600	0.600	0.607	0.467
TelecomTS	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.864
Tennessee Eastman	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.598
Voraus	0.310	0.310	0.308	0.310	0.310	0.310	0.310	0.310	0.310	0.310	0.281
macro	*0.527*	*0.529*	*0.527*	*0.527*	*0.567*	*0.596*	*0.527*	*0.529*	*0.532*	*0.543*	*0.478*

Table 48: Full metric grid for **MantisV2 + MajVote**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.599	0.482
DAMADICS	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.389
Exathlon	0.920	0.920	0.920	0.920	0.920	0.920	0.920	0.920	0.920	0.920	0.792
HAI	0.393	0.393	0.386	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.465
MIT-BIH	0.340	0.340	0.340	0.340	0.340	0.340	0.340	0.340	0.340	0.340	0.227
Petrobras 3W	0.850	0.850	0.850	0.850	0.850	0.850	0.850	0.850	0.850	0.850	0.743
RATS40K	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.209
RCAEval	0.400	0.368	0.380	0.400	0.800	0.800	0.400	0.370	0.378	0.379	0.267
ROAD	0.600	0.600	0.588	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.475
TelecomTS	0.640	0.640	0.640	0.640	0.640	0.640	0.640	0.640	0.640	0.640	0.602
Tennessee Eastman	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.660
Voraus	0.340	0.340	0.338	0.340	0.340	0.340	0.340	0.340	0.340	0.340	0.299
macro	*0.573*	*0.571*	*0.570*	*0.573*	*0.607*	*0.607*	*0.573*	*0.571*	*0.571*	*0.576*	*0.467*
$\rho\approx$10%											
CTSR	0.540	0.542	0.542	0.540	0.550	0.550	0.540	0.541	0.542	0.593	0.470
DAMADICS	0.556	0.556	0.522	0.556	0.556	0.556	0.556	0.556	0.556	0.556	0.510
Exathlon	0.930	0.930	0.930	0.930	0.930	0.930	0.930	0.930	0.930	0.930	0.798
HAI	0.393	0.393	0.386	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.501
MIT-BIH	0.310	0.310	0.310	0.310	0.310	0.310	0.310	0.310	0.310	0.310	0.126
Petrobras 3W	0.830	0.830	0.830	0.830	0.830	0.830	0.830	0.830	0.830	0.830	0.719
RATS40K	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.200
RCAEval	0.240	0.352	0.364	0.240	0.600	0.920	0.240	0.338	0.352	0.365	0.315
ROAD	0.600	0.600	0.588	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.475
TelecomTS	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.660	0.629
Tennessee Eastman	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.620
Voraus	0.300	0.300	0.298	0.300	0.300	0.300	0.300	0.300	0.300	0.300	0.267
macro	*0.539*	*0.549*	*0.545*	*0.539*	*0.570*	*0.597*	*0.539*	*0.547*	*0.548*	*0.554*	*0.469*
$\rho\approx$20%											
CTSR	0.540	0.540	0.542	0.540	0.540	0.550	0.540	0.540	0.541	0.591	0.470
DAMADICS	0.500	0.500	0.467	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.482
Exathlon	0.910	0.910	0.910	0.910	0.910	0.910	0.910	0.910	0.910	0.910	0.790
HAI	0.429	0.429	0.421	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.545
MIT-BIH	0.280	0.280	0.280	0.280	0.280	0.280	0.280	0.280	0.280	0.280	0.126
Petrobras 3W	0.800	0.800	0.800	0.800	0.800	0.800	0.800	0.800	0.800	0.800	0.684
RATS40K	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.204
RCAEval	0.300	0.332	0.316	0.300	0.780	0.900	0.300	0.331	0.319	0.326	0.253
ROAD	0.520	0.520	0.508	0.520	0.520	0.520	0.520	0.520	0.520	0.525	0.458
TelecomTS	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.606
Tennessee Eastman	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.593
Voraus	0.270	0.270	0.268	0.270	0.270	0.270	0.270	0.270	0.270	0.270	0.259
macro	*0.521*	*0.523*	*0.518*	*0.521*	*0.561*	*0.572*	*0.521*	*0.523*	*0.522*	*0.528*	*0.456*

Table 49: Full metric grid for **MantisV2 + QPurity**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.580	0.520	0.438	0.580	0.730	0.850	0.580	0.531	0.470	0.515	0.538
DAMADICS	0.667	0.700	0.656	0.667	0.944	0.944	0.667	0.688	0.680	0.704	0.750
Exathlon	0.930	0.926	0.922	0.930	0.930	0.930	0.930	0.927	0.924	0.922	0.800
HAI	0.429	0.371	0.361	0.429	0.500	0.643	0.429	0.384	0.376	0.380	0.436
MIT-BIH	0.360	0.340	0.339	0.360	0.370	0.390	0.360	0.343	0.342	0.348	0.227
Petrobras 3W	0.860	0.832	0.797	0.860	0.890	0.890	0.860	0.838	0.812	0.764	0.708
RATS40K	0.510	0.486	0.458	0.510	0.590	0.660	0.510	0.494	0.474	0.463	0.206
RCAEval	0.200	0.200	0.200	0.200	0.800	0.800	0.200	0.200	0.200	0.200	0.067
ROAD	0.520	0.512	0.496	0.520	0.520	0.520	0.520	0.514	0.508	0.512	0.341
TelecomTS	0.720	0.660	0.619	0.720	0.820	0.870	0.720	0.674	0.640	0.595	0.667
Tennessee Eastman	0.660	0.634	0.608	0.660	0.790	0.850	0.660	0.640	0.620	0.580	0.664
Voraus	0.320	0.306	0.280	0.320	0.370	0.410	0.320	0.309	0.291	0.281	0.259
macro	*0.563*	*0.541*	*0.514*	*0.563*	*0.688*	*0.730*	*0.563*	*0.545*	*0.528*	*0.522*	*0.472*
$\rho\approx$10%											
CTSR	0.570	0.530	0.455	0.570	0.710	0.850	0.570	0.539	0.484	0.528	0.550
DAMADICS	0.500	0.544	0.578	0.500	0.889	0.889	0.500	0.531	0.570	0.589	0.405
Exathlon	0.940	0.938	0.933	0.940	0.940	0.940	0.940	0.939	0.935	0.933	0.806
HAI	0.393	0.393	0.371	0.393	0.500	0.571	0.393	0.396	0.384	0.368	0.498
MIT-BIH	0.260	0.288	0.296	0.260	0.320	0.330	0.260	0.283	0.290	0.293	0.124
Petrobras 3W	0.860	0.800	0.781	0.860	0.870	0.890	0.860	0.812	0.795	0.761	0.699
RATS40K	0.480	0.476	0.453	0.480	0.590	0.650	0.480	0.480	0.464	0.449	0.195
RCAEval	0.000	0.120	0.160	0.000	0.400	1.000	0.000	0.103	0.137	0.159	0.000
ROAD	0.520	0.512	0.496	0.520	0.520	0.520	0.520	0.515	0.510	0.526	0.341
TelecomTS	0.700	0.656	0.630	0.700	0.800	0.870	0.700	0.667	0.645	0.608	0.684
Tennessee Eastman	0.660	0.626	0.602	0.660	0.760	0.850	0.660	0.632	0.613	0.574	0.653
Voraus	0.280	0.268	0.253	0.280	0.330	0.380	0.280	0.271	0.261	0.252	0.228
macro	*0.514*	*0.513*	*0.501*	*0.514*	*0.636*	*0.728*	*0.514*	*0.514*	*0.507*	*0.503*	*0.432*
$\rho\approx$20%											
CTSR	0.570	0.534	0.458	0.570	0.710	0.830	0.570	0.543	0.487	0.534	0.550
DAMADICS	0.389	0.489	0.539	0.389	0.889	0.889	0.389	0.469	0.521	0.547	0.386
Exathlon	0.920	0.914	0.912	0.920	0.920	0.920	0.920	0.916	0.914	0.914	0.790
HAI	0.429	0.421	0.389	0.429	0.464	0.571	0.429	0.425	0.405	0.385	0.538
MIT-BIH	0.290	0.284	0.294	0.290	0.320	0.330	0.290	0.284	0.291	0.298	0.128
Petrobras 3W	0.800	0.764	0.745	0.800	0.820	0.860	0.800	0.770	0.756	0.734	0.624
RATS40K	0.450	0.452	0.439	0.450	0.560	0.650	0.450	0.455	0.446	0.432	0.191
RCAEval	0.200	0.160	0.140	0.200	0.800	1.000	0.200	0.171	0.151	0.168	0.067
ROAD	0.520	0.488	0.452	0.520	0.520	0.600	0.520	0.495	0.473	0.489	0.341
TelecomTS	0.710	0.658	0.618	0.710	0.780	0.830	0.710	0.670	0.638	0.597	0.717
Tennessee Eastman	0.590	0.566	0.535	0.590	0.610	0.630	0.590	0.571	0.548	0.517	0.621
Voraus	0.240	0.224	0.196	0.240	0.320	0.430	0.240	0.229	0.209	0.194	0.219
macro	*0.509*	*0.496*	*0.476*	*0.509*	*0.643*	*0.712*	*0.509*	*0.500*	*0.486*	*0.484*	*0.431*

Table 50: Full metric grid for **MantisV2 + PRF**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.540	0.480	0.405	0.540	0.750	0.830	0.540	0.495	0.439	0.504	0.528
DAMADICS	0.556	0.678	0.600	0.556	0.944	0.944	0.556	0.648	0.624	0.624	0.776
Exathlon	0.940	0.922	0.913	0.940	0.960	0.980	0.940	0.926	0.918	0.904	0.958
HAI	0.500	0.471	0.393	0.500	0.643	0.857	0.500	0.481	0.429	0.395	0.569
MIT-BIH	0.330	0.328	0.333	0.330	0.350	0.370	0.330	0.329	0.332	0.336	0.232
Petrobras 3W	0.850	0.808	0.758	0.850	0.940	0.980	0.850	0.818	0.780	0.707	0.738
RATS40K	0.410	0.408	0.401	0.410	0.730	0.830	0.410	0.411	0.410	0.406	0.253
RCAEval	0.380	0.312	0.334	0.380	0.840	0.880	0.380	0.320	0.333	0.319	0.227
ROAD	0.520	0.472	0.388	0.520	0.840	0.880	0.520	0.479	0.430	0.423	0.404
TelecomTS	0.790	0.638	0.529	0.790	0.900	0.930	0.790	0.669	0.582	0.502	0.697
Tennessee Eastman	0.540	0.510	0.485	0.540	0.770	0.870	0.540	0.520	0.500	0.453	0.604
Voraus	0.290	0.240	0.218	0.290	0.650	0.830	0.290	0.246	0.230	0.206	0.230
macro	*0.554*	*0.522*	*0.480*	*0.554*	*0.776*	*0.848*	*0.554*	*0.528*	*0.501*	*0.481*	*0.518*
$\rho\approx$10%											
CTSR	0.540	0.480	0.406	0.540	0.750	0.830	0.540	0.495	0.439	0.505	0.528
DAMADICS	0.444	0.511	0.472	0.444	0.889	0.944	0.444	0.498	0.493	0.511	0.679
Exathlon	0.950	0.924	0.912	0.950	0.960	0.990	0.950	0.929	0.919	0.902	0.965
HAI	0.500	0.471	0.393	0.500	0.607	0.821	0.500	0.481	0.428	0.385	0.569
MIT-BIH	0.250	0.258	0.257	0.250	0.310	0.340	0.250	0.256	0.256	0.248	0.117
Petrobras 3W	0.790	0.772	0.729	0.790	0.940	0.980	0.790	0.778	0.746	0.681	0.738
RATS40K	0.390	0.394	0.389	0.390	0.730	0.810	0.390	0.396	0.397	0.386	0.167
RCAEval	0.220	0.292	0.306	0.220	0.640	1.000	0.220	0.285	0.298	0.293	0.309
ROAD	0.520	0.432	0.364	0.520	0.840	0.880	0.520	0.454	0.408	0.394	0.402
TelecomTS	0.780	0.638	0.521	0.780	0.900	0.930	0.780	0.667	0.575	0.495	0.718
Tennessee Eastman	0.520	0.486	0.453	0.520	0.730	0.810	0.520	0.496	0.470	0.425	0.566
Voraus	0.250	0.218	0.201	0.250	0.630	0.800	0.250	0.221	0.210	0.189	0.202
macro	*0.513*	*0.490*	*0.450*	*0.513*	*0.744*	*0.845*	*0.513*	*0.496*	*0.470*	*0.451*	*0.497*
$\rho\approx$20%											
CTSR	0.540	0.476	0.403	0.540	0.740	0.830	0.540	0.492	0.437	0.504	0.520
DAMADICS	0.389	0.444	0.417	0.389	0.833	0.889	0.389	0.432	0.433	0.445	0.643
Exathlon	0.940	0.922	0.909	0.940	0.960	1.000	0.940	0.926	0.916	0.898	0.961
HAI	0.500	0.421	0.357	0.500	0.536	0.750	0.500	0.443	0.396	0.360	0.601
MIT-BIH	0.270	0.258	0.269	0.270	0.310	0.320	0.270	0.259	0.266	0.266	0.156
Petrobras 3W	0.750	0.716	0.683	0.750	0.930	0.970	0.750	0.726	0.700	0.646	0.672
RATS40K	0.400	0.386	0.385	0.400	0.710	0.800	0.400	0.391	0.393	0.371	0.169
RCAEval	0.300	0.276	0.260	0.300	0.840	0.960	0.300	0.288	0.271	0.266	0.266
ROAD	0.520	0.392	0.316	0.520	0.800	0.880	0.520	0.427	0.370	0.377	0.402
TelecomTS	0.780	0.628	0.512	0.780	0.890	0.940	0.780	0.660	0.568	0.484	0.722
Tennessee Eastman	0.500	0.474	0.435	0.500	0.720	0.770	0.500	0.479	0.450	0.403	0.519
Voraus	0.230	0.194	0.186	0.230	0.580	0.750	0.230	0.200	0.194	0.172	0.194
macro	*0.510*	*0.466*	*0.428*	*0.510*	*0.737*	*0.822*	*0.510*	*0.477*	*0.450*	*0.433*	*0.485*

Table 51: Full metric grid for **MantisV2 + Purity**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.270	0.248	0.223	0.270	0.380	0.420	0.270	0.250	0.232	0.265	0.168
DAMADICS	0.611	0.600	0.583	0.611	0.667	0.667	0.611	0.603	0.592	0.564	0.360
Exathlon	0.940	0.902	0.888	0.940	0.940	0.940	0.940	0.910	0.897	0.884	0.609
HAI	0.464	0.500	0.454	0.464	0.643	0.714	0.464	0.498	0.465	0.438	0.484
MIT-BIH	0.350	0.346	0.348	0.350	0.410	0.450	0.350	0.348	0.349	0.344	0.223
Petrobras 3W	0.740	0.694	0.657	0.740	0.820	0.850	0.740	0.706	0.676	0.627	0.612
RATS40K	0.490	0.436	0.421	0.490	0.570	0.590	0.490	0.449	0.434	0.419	0.166
RCAEval	0.200	0.200	0.200	0.200	0.800	0.800	0.200	0.200	0.200	0.200	0.067
ROAD	0.480	0.480	0.484	0.480	0.480	0.520	0.480	0.480	0.483	0.496	0.130
TelecomTS	0.520	0.454	0.417	0.520	0.590	0.640	0.520	0.468	0.437	0.395	0.306
Tennessee Eastman	0.460	0.404	0.364	0.460	0.460	0.500	0.460	0.415	0.384	0.337	0.332
Voraus	0.200	0.196	0.183	0.200	0.330	0.420	0.200	0.197	0.188	0.193	0.119
macro	*0.477*	*0.455*	*0.435*	*0.477*	*0.591*	*0.626*	*0.477*	*0.460*	*0.445*	*0.430*	*0.298*
$\rho\approx$10%											
CTSR	0.250	0.260	0.229	0.250	0.390	0.430	0.250	0.257	0.236	0.269	0.182
DAMADICS	0.611	0.611	0.594	0.611	0.722	0.722	0.611	0.607	0.602	0.581	0.360
Exathlon	0.910	0.898	0.883	0.910	0.920	0.940	0.910	0.900	0.889	0.877	0.609
HAI	0.464	0.493	0.454	0.464	0.643	0.714	0.464	0.497	0.468	0.444	0.458
MIT-BIH	0.310	0.292	0.288	0.310	0.350	0.380	0.310	0.297	0.293	0.297	0.126
Petrobras 3W	0.700	0.650	0.621	0.700	0.770	0.830	0.700	0.663	0.638	0.592	0.580
RATS40K	0.450	0.430	0.414	0.450	0.560	0.590	0.450	0.435	0.422	0.411	0.148
RCAEval	0.000	0.120	0.160	0.000	0.400	1.000	0.000	0.103	0.137	0.159	0.000
ROAD	0.480	0.480	0.488	0.480	0.480	0.520	0.480	0.480	0.486	0.496	0.130
TelecomTS	0.520	0.450	0.417	0.520	0.580	0.640	0.520	0.466	0.437	0.393	0.308
Tennessee Eastman	0.460	0.386	0.348	0.460	0.460	0.500	0.460	0.403	0.371	0.333	0.336
Voraus	0.200	0.182	0.186	0.200	0.320	0.460	0.200	0.185	0.188	0.191	0.116
macro	*0.446*	*0.438*	*0.424*	*0.446*	*0.550*	*0.644*	*0.446*	*0.441*	*0.431*	*0.420*	*0.279*
$\rho\approx$20%											
CTSR	0.250	0.258	0.227	0.250	0.370	0.410	0.250	0.254	0.234	0.267	0.182
DAMADICS	0.611	0.578	0.561	0.611	0.667	0.722	0.611	0.583	0.571	0.554	0.360
Exathlon	0.910	0.896	0.878	0.910	0.920	0.940	0.910	0.898	0.885	0.869	0.609
HAI	0.464	0.464	0.436	0.464	0.607	0.679	0.464	0.468	0.447	0.413	0.462
MIT-BIH	0.330	0.320	0.331	0.330	0.350	0.450	0.330	0.322	0.329	0.327	0.139
Petrobras 3W	0.660	0.612	0.578	0.660	0.730	0.790	0.660	0.624	0.596	0.558	0.500
RATS40K	0.420	0.404	0.395	0.420	0.540	0.580	0.420	0.408	0.401	0.383	0.139
RCAEval	0.200	0.160	0.140	0.200	0.800	1.000	0.200	0.171	0.151	0.168	0.067
ROAD	0.480	0.448	0.472	0.480	0.480	0.520	0.480	0.451	0.469	0.469	0.130
TelecomTS	0.500	0.428	0.397	0.500	0.560	0.600	0.500	0.444	0.417	0.377	0.294
Tennessee Eastman	0.460	0.366	0.342	0.460	0.460	0.500	0.460	0.386	0.362	0.325	0.303
Voraus	0.070	0.094	0.089	0.070	0.180	0.320	0.070	0.090	0.089	0.085	0.116
macro	*0.446*	*0.419*	*0.404*	*0.446*	*0.555*	*0.626*	*0.446*	*0.425*	*0.413*	*0.400*	*0.275*

Table 52: Full metric grid for **MantisV2 + NR + GPC**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.600	0.600	0.602	0.600	0.600	0.610	0.600	0.600	0.601	0.660	0.546
DAMADICS	0.556	0.556	0.506	0.556	0.556	0.556	0.556	0.556	0.556	0.556	0.342
Exathlon	0.950	0.950	0.950	0.950	0.950	0.950	0.950	0.950	0.950	0.950	0.914
HAI	0.536	0.536	0.529	0.536	0.536	0.536	0.536	0.536	0.536	0.536	0.616
MIT-BIH	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.428
Petrobras 3W	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.712
RATS40K	0.460	0.464	0.462	0.460	0.480	0.510	0.460	0.465	0.467	0.477	0.251
RCAEval	0.460	0.428	0.440	0.460	0.860	0.860	0.460	0.430	0.438	0.439	0.361
ROAD	0.680	0.680	0.668	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.511
TelecomTS	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.854
Tennessee Eastman	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.645
Voraus	0.600	0.600	0.596	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.545
macro	*0.646*	*0.644*	*0.639*	*0.646*	*0.681*	*0.684*	*0.646*	*0.644*	*0.645*	*0.651*	*0.561*
$\rho\approx$10%											
CTSR	0.600	0.592	0.592	0.600	0.600	0.600	0.600	0.593	0.593	0.650	0.535
DAMADICS	0.444	0.444	0.394	0.444	0.444	0.444	0.444	0.444	0.444	0.444	0.303
Exathlon	0.930	0.930	0.930	0.930	0.930	0.930	0.930	0.930	0.930	0.930	0.906
HAI	0.536	0.536	0.529	0.536	0.536	0.536	0.536	0.536	0.536	0.536	0.616
MIT-BIH	0.450	0.450	0.450	0.450	0.450	0.450	0.450	0.450	0.450	0.450	0.309
Petrobras 3W	0.790	0.790	0.790	0.790	0.790	0.790	0.790	0.790	0.790	0.790	0.674
RATS40K	0.460	0.460	0.459	0.460	0.470	0.500	0.460	0.461	0.466	0.477	0.279
RCAEval	0.300	0.412	0.424	0.300	0.660	0.980	0.300	0.398	0.412	0.425	0.421
ROAD	0.640	0.640	0.628	0.640	0.640	0.640	0.640	0.640	0.640	0.640	0.480
TelecomTS	0.830	0.830	0.830	0.830	0.830	0.830	0.830	0.830	0.830	0.830	0.814
Tennessee Eastman	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.650
Voraus	0.590	0.590	0.586	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.541
macro	*0.603*	*0.612*	*0.607*	*0.603*	*0.634*	*0.663*	*0.603*	*0.611*	*0.613*	*0.619*	*0.544*
$\rho\approx$20%											
CTSR	0.590	0.594	0.595	0.590	0.600	0.610	0.590	0.594	0.595	0.653	0.543
DAMADICS	0.389	0.389	0.339	0.389	0.389	0.389	0.389	0.389	0.389	0.389	0.279
Exathlon	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.915
HAI	0.500	0.500	0.493	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.586
MIT-BIH	0.270	0.270	0.270	0.270	0.270	0.270	0.270	0.270	0.270	0.270	0.152
Petrobras 3W	0.760	0.760	0.760	0.760	0.760	0.760	0.760	0.760	0.760	0.760	0.638
RATS40K	0.420	0.418	0.415	0.420	0.420	0.440	0.420	0.420	0.421	0.425	0.247
RCAEval	0.380	0.412	0.396	0.380	0.860	0.980	0.380	0.411	0.399	0.406	0.361
ROAD	0.640	0.640	0.628	0.640	0.640	0.640	0.640	0.640	0.640	0.640	0.480
TelecomTS	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.807
Tennessee Eastman	0.650	0.650	0.650	0.650	0.650	0.650	0.650	0.650	0.650	0.650	0.635
Voraus	0.600	0.600	0.596	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.548
macro	*0.580*	*0.583*	*0.575*	*0.580*	*0.621*	*0.633*	*0.580*	*0.583*	*0.582*	*0.588*	*0.516*

Table 53: Full metric grid for **MantisV2 + NR + MajVote**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.632	0.526
DAMADICS	0.500	0.500	0.483	0.500	0.500	0.500	0.500	0.500	0.500	0.508	0.257
Exathlon	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.789
HAI	0.357	0.357	0.350	0.357	0.357	0.357	0.357	0.357	0.357	0.357	0.342
MIT-BIH	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.361
Petrobras 3W	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.568
RATS40K	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.229
RCAEval	0.460	0.428	0.440	0.460	0.860	0.860	0.460	0.430	0.438	0.439	0.361
ROAD	0.680	0.680	0.668	0.680	0.680	0.680	0.680	0.680	0.680	0.685	0.520
TelecomTS	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.793
Tennessee Eastman	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.550
Voraus	0.580	0.580	0.576	0.580	0.580	0.600	0.580	0.580	0.581	0.584	0.486
macro	*0.617*	*0.615*	*0.612*	*0.617*	*0.651*	*0.652*	*0.617*	*0.615*	*0.616*	*0.620*	*0.482*
$\rho\approx$10%											
CTSR	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.632	0.522
DAMADICS	0.500	0.500	0.467	0.500	0.500	0.500	0.500	0.500	0.500	0.508	0.302
Exathlon	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.789
HAI	0.357	0.357	0.350	0.357	0.357	0.357	0.357	0.357	0.357	0.357	0.342
MIT-BIH	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.379
Petrobras 3W	0.740	0.740	0.740	0.740	0.740	0.740	0.740	0.740	0.740	0.740	0.568
RATS40K	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.212
RCAEval	0.300	0.412	0.424	0.300	0.660	0.980	0.300	0.398	0.412	0.425	0.421
ROAD	0.680	0.680	0.668	0.680	0.680	0.680	0.680	0.680	0.680	0.685	0.520
TelecomTS	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.793
Tennessee Eastman	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.552
Voraus	0.550	0.550	0.546	0.550	0.550	0.570	0.550	0.550	0.551	0.553	0.462
macro	*0.601*	*0.610*	*0.606*	*0.601*	*0.631*	*0.659*	*0.601*	*0.609*	*0.610*	*0.616*	*0.489*
$\rho\approx$20%											
CTSR	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.631	0.522
DAMADICS	0.389	0.389	0.356	0.389	0.389	0.389	0.389	0.389	0.389	0.397	0.268
Exathlon	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.789
HAI	0.357	0.357	0.350	0.357	0.357	0.357	0.357	0.357	0.357	0.357	0.342
MIT-BIH	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.288
Petrobras 3W	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.541
RATS40K	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.211
RCAEval	0.380	0.412	0.396	0.380	0.860	0.980	0.380	0.411	0.399	0.406	0.361
ROAD	0.640	0.640	0.628	0.640	0.640	0.640	0.640	0.640	0.640	0.648	0.512
TelecomTS	0.810	0.810	0.810	0.810	0.810	0.810	0.810	0.810	0.810	0.810	0.789
Tennessee Eastman	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.572
Voraus	0.550	0.550	0.546	0.550	0.550	0.570	0.550	0.550	0.551	0.553	0.462
macro	*0.587*	*0.590*	*0.584*	*0.587*	*0.627*	*0.639*	*0.587*	*0.590*	*0.589*	*0.594*	*0.471*

Table 54: Full metric grid for **MantisV2 + NR + QPurity**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.570	0.508	0.440	0.570	0.680	0.820	0.570	0.521	0.470	0.512	0.477
DAMADICS	0.722	0.600	0.522	0.722	0.778	0.889	0.722	0.633	0.591	0.577	0.622
Exathlon	0.920	0.922	0.915	0.920	0.970	0.970	0.920	0.923	0.918	0.906	0.840
HAI	0.357	0.379	0.368	0.357	0.429	0.464	0.357	0.377	0.372	0.371	0.405
MIT-BIH	0.560	0.548	0.552	0.560	0.650	0.730	0.560	0.551	0.552	0.559	0.352
Petrobras 3W	0.770	0.734	0.696	0.770	0.840	0.890	0.770	0.743	0.714	0.667	0.559
RATS40K	0.500	0.486	0.471	0.500	0.690	0.750	0.500	0.492	0.481	0.470	0.209
RCAEval	0.200	0.200	0.200	0.200	0.800	0.800	0.200	0.200	0.200	0.200	0.067
ROAD	0.600	0.584	0.556	0.600	0.760	0.880	0.600	0.586	0.581	0.593	0.422
TelecomTS	0.900	0.758	0.700	0.900	0.970	0.970	0.900	0.784	0.734	0.671	0.815
Tennessee Eastman	0.620	0.622	0.608	0.620	0.690	0.730	0.620	0.621	0.612	0.601	0.534
Voraus	0.600	0.534	0.501	0.600	0.760	0.850	0.600	0.549	0.523	0.476	0.533
macro	*0.610*	*0.573*	*0.544*	*0.610*	*0.751*	*0.812*	*0.610*	*0.582*	*0.562*	*0.550*	*0.486*
$\rho\approx$10%											
CTSR	0.560	0.520	0.449	0.560	0.670	0.810	0.560	0.528	0.477	0.521	0.495
DAMADICS	0.556	0.511	0.461	0.556	0.778	0.944	0.556	0.528	0.506	0.496	0.295
Exathlon	0.940	0.922	0.910	0.940	0.970	0.970	0.940	0.926	0.916	0.902	0.840
HAI	0.321	0.350	0.361	0.321	0.429	0.464	0.321	0.347	0.359	0.370	0.342
MIT-BIH	0.540	0.552	0.557	0.540	0.580	0.650	0.540	0.551	0.555	0.555	0.379
Petrobras 3W	0.760	0.734	0.693	0.760	0.830	0.910	0.760	0.740	0.710	0.662	0.589
RATS40K	0.470	0.482	0.462	0.470	0.690	0.750	0.470	0.483	0.470	0.458	0.219
RCAEval	0.000	0.120	0.160	0.000	0.400	1.000	0.000	0.103	0.137	0.159	0.000
ROAD	0.600	0.584	0.564	0.600	0.760	0.880	0.600	0.596	0.593	0.607	0.422
TelecomTS	0.900	0.758	0.709	0.900	0.960	0.970	0.900	0.786	0.742	0.681	0.816
Tennessee Eastman	0.620	0.618	0.607	0.620	0.670	0.720	0.620	0.618	0.610	0.596	0.539
Voraus	0.570	0.528	0.486	0.570	0.720	0.840	0.570	0.539	0.507	0.464	0.523
macro	*0.570*	*0.557*	*0.535*	*0.570*	*0.705*	*0.826*	*0.570*	*0.562*	*0.549*	*0.539*	*0.455*
$\rho\approx$20%											
CTSR	0.560	0.518	0.451	0.560	0.670	0.800	0.560	0.527	0.478	0.525	0.495
DAMADICS	0.556	0.467	0.444	0.556	0.778	0.944	0.556	0.491	0.491	0.487	0.344
Exathlon	0.940	0.920	0.906	0.940	0.970	0.970	0.940	0.925	0.913	0.897	0.840
HAI	0.321	0.357	0.354	0.321	0.429	0.464	0.321	0.350	0.352	0.363	0.342
MIT-BIH	0.540	0.544	0.545	0.540	0.560	0.570	0.540	0.543	0.545	0.546	0.288
Petrobras 3W	0.760	0.704	0.661	0.760	0.810	0.900	0.760	0.717	0.682	0.636	0.581
RATS40K	0.480	0.464	0.454	0.480	0.660	0.760	0.480	0.472	0.464	0.449	0.221
RCAEval	0.200	0.160	0.140	0.200	0.800	1.000	0.200	0.171	0.151	0.168	0.067
ROAD	0.640	0.576	0.544	0.640	0.760	0.880	0.640	0.596	0.580	0.584	0.460
TelecomTS	0.880	0.744	0.704	0.880	0.960	0.970	0.880	0.772	0.735	0.674	0.763
Tennessee Eastman	0.610	0.614	0.598	0.610	0.650	0.680	0.610	0.613	0.603	0.591	0.552
Voraus	0.560	0.482	0.434	0.560	0.780	0.890	0.560	0.498	0.460	0.403	0.440
macro	*0.587*	*0.546*	*0.520*	*0.587*	*0.736*	*0.819*	*0.587*	*0.556*	*0.538*	*0.527*	*0.449*

Table 55: Full metric grid for **MantisV2 + NR + PRF**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.460	0.434	0.379	0.460	0.720	0.830	0.460	0.441	0.401	0.474	0.374
DAMADICS	0.500	0.467	0.428	0.500	0.667	0.778	0.500	0.479	0.464	0.444	0.426
Exathlon	0.870	0.886	0.877	0.870	0.930	0.940	0.870	0.883	0.878	0.876	0.664
HAI	0.321	0.343	0.354	0.321	0.464	0.643	0.321	0.334	0.349	0.326	0.477
MIT-BIH	0.550	0.534	0.544	0.550	0.560	0.600	0.550	0.537	0.543	0.548	0.358
Petrobras 3W	0.730	0.692	0.661	0.730	0.820	0.870	0.730	0.701	0.677	0.628	0.541
RATS40K	0.480	0.396	0.384	0.480	0.730	0.810	0.480	0.410	0.400	0.393	0.142
RCAEval	0.480	0.400	0.402	0.480	0.880	0.880	0.480	0.410	0.408	0.384	0.337
ROAD	0.560	0.408	0.340	0.560	0.600	0.840	0.560	0.446	0.393	0.402	0.195
TelecomTS	0.710	0.686	0.611	0.710	0.880	0.920	0.710	0.695	0.640	0.565	0.733
Tennessee Eastman	0.630	0.628	0.613	0.630	0.740	0.830	0.630	0.628	0.618	0.574	0.592
Voraus	0.410	0.384	0.344	0.410	0.810	0.950	0.410	0.390	0.362	0.324	0.440
macro	*0.558*	*0.521*	*0.495*	*0.558*	*0.733*	*0.824*	*0.558*	*0.529*	*0.511*	*0.495*	*0.440*
$\rho\approx$10%											
CTSR	0.460	0.434	0.379	0.460	0.720	0.830	0.460	0.441	0.400	0.473	0.385
DAMADICS	0.500	0.478	0.367	0.500	0.667	0.778	0.500	0.491	0.424	0.407	0.461
Exathlon	0.870	0.876	0.870	0.870	0.930	0.950	0.870	0.876	0.872	0.869	0.664
HAI	0.321	0.343	0.354	0.321	0.464	0.643	0.321	0.334	0.349	0.324	0.477
MIT-BIH	0.560	0.546	0.545	0.560	0.610	0.610	0.560	0.550	0.548	0.532	0.375
Petrobras 3W	0.700	0.680	0.662	0.700	0.820	0.880	0.700	0.685	0.671	0.617	0.582
RATS40K	0.480	0.380	0.368	0.480	0.710	0.810	0.480	0.400	0.388	0.379	0.146
RCAEval	0.320	0.384	0.382	0.320	0.680	1.000	0.320	0.377	0.379	0.369	0.397
ROAD	0.480	0.352	0.304	0.480	0.560	0.800	0.480	0.385	0.345	0.360	0.133
TelecomTS	0.710	0.686	0.610	0.710	0.890	0.910	0.710	0.695	0.640	0.563	0.753
Tennessee Eastman	0.620	0.624	0.605	0.620	0.740	0.820	0.620	0.623	0.610	0.568	0.586
Voraus	0.410	0.380	0.335	0.410	0.830	0.940	0.410	0.386	0.355	0.318	0.443
macro	*0.536*	*0.514*	*0.482*	*0.536*	*0.718*	*0.831*	*0.536*	*0.520*	*0.498*	*0.482*	*0.450*
$\rho\approx$20%											
CTSR	0.460	0.432	0.378	0.460	0.720	0.830	0.460	0.439	0.399	0.472	0.385
DAMADICS	0.556	0.444	0.356	0.556	0.667	0.778	0.556	0.483	0.424	0.395	0.444
Exathlon	0.870	0.864	0.856	0.870	0.930	0.950	0.870	0.867	0.860	0.854	0.660
HAI	0.321	0.343	0.350	0.321	0.464	0.607	0.321	0.334	0.346	0.321	0.477
MIT-BIH	0.520	0.514	0.506	0.520	0.570	0.580	0.520	0.514	0.509	0.505	0.265
Petrobras 3W	0.660	0.648	0.641	0.660	0.820	0.910	0.660	0.654	0.647	0.605	0.501
RATS40K	0.450	0.372	0.365	0.450	0.700	0.800	0.450	0.389	0.382	0.373	0.151
RCAEval	0.400	0.384	0.352	0.400	0.880	1.000	0.400	0.391	0.364	0.348	0.337
ROAD	0.440	0.400	0.328	0.440	0.680	0.800	0.440	0.413	0.362	0.368	0.141
TelecomTS	0.720	0.666	0.600	0.720	0.890	0.910	0.720	0.680	0.630	0.553	0.714
Tennessee Eastman	0.630	0.624	0.602	0.630	0.760	0.850	0.630	0.625	0.609	0.564	0.614
Voraus	0.400	0.370	0.333	0.400	0.840	0.940	0.400	0.377	0.351	0.310	0.412
macro	*0.536*	*0.505*	*0.472*	*0.536*	*0.743*	*0.830*	*0.536*	*0.514*	*0.490*	*0.472*	*0.425*

Table 56: Full metric grid for **MantisV2 + NR + Purity**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.260	0.240	0.204	0.260	0.360	0.390	0.260	0.242	0.216	0.248	0.157
DAMADICS	0.556	0.522	0.494	0.556	0.667	0.722	0.556	0.532	0.510	0.450	0.330
Exathlon	0.950	0.936	0.927	0.950	0.960	0.970	0.950	0.938	0.932	0.920	0.730
HAI	0.286	0.307	0.304	0.286	0.536	0.750	0.286	0.303	0.302	0.327	0.233
MIT-BIH	0.540	0.554	0.567	0.540	0.710	0.820	0.540	0.553	0.562	0.569	0.372
Petrobras 3W	0.740	0.696	0.636	0.740	0.810	0.860	0.740	0.707	0.661	0.609	0.580
RATS40K	0.510	0.452	0.446	0.510	0.630	0.720	0.510	0.465	0.457	0.433	0.179
RCAEval	0.200	0.200	0.200	0.200	0.800	0.800	0.200	0.200	0.200	0.200	0.067
ROAD	0.520	0.536	0.508	0.520	0.680	0.680	0.520	0.535	0.521	0.515	0.333
TelecomTS	0.880	0.706	0.625	0.880	0.940	0.970	0.880	0.742	0.673	0.606	0.683
Tennessee Eastman	0.540	0.520	0.505	0.540	0.550	0.610	0.540	0.524	0.512	0.488	0.415
Voraus	0.330	0.316	0.312	0.330	0.590	0.720	0.330	0.321	0.318	0.301	0.206
macro	*0.526*	*0.499*	*0.477*	*0.526*	*0.686*	*0.751*	*0.526*	*0.505*	*0.489*	*0.472*	*0.357*
$\rho\approx$10%											
CTSR	0.260	0.242	0.205	0.260	0.360	0.390	0.260	0.244	0.218	0.253	0.161
DAMADICS	0.556	0.511	0.439	0.556	0.722	0.778	0.556	0.531	0.486	0.496	0.511
Exathlon	0.960	0.938	0.922	0.960	0.960	0.970	0.960	0.942	0.930	0.914	0.899
HAI	0.286	0.307	0.304	0.286	0.500	0.750	0.286	0.295	0.296	0.329	0.252
MIT-BIH	0.530	0.564	0.560	0.530	0.700	0.780	0.530	0.559	0.558	0.561	0.407
Petrobras 3W	0.730	0.692	0.624	0.730	0.800	0.830	0.730	0.701	0.651	0.601	0.538
RATS40K	0.470	0.446	0.437	0.470	0.630	0.710	0.470	0.455	0.446	0.424	0.192
RCAEval	0.000	0.120	0.160	0.000	0.400	1.000	0.000	0.103	0.137	0.159	0.000
ROAD	0.520	0.536	0.504	0.520	0.680	0.680	0.520	0.535	0.518	0.517	0.333
TelecomTS	0.890	0.708	0.625	0.890	0.930	0.960	0.890	0.743	0.673	0.607	0.683
Tennessee Eastman	0.540	0.516	0.501	0.540	0.550	0.610	0.540	0.521	0.509	0.486	0.419
Voraus	0.380	0.330	0.305	0.380	0.630	0.700	0.380	0.339	0.319	0.297	0.265
macro	*0.510*	*0.493*	*0.465*	*0.510*	*0.655*	*0.763*	*0.510*	*0.497*	*0.478*	*0.470*	*0.388*
$\rho\approx$20%											
CTSR	0.260	0.242	0.203	0.260	0.360	0.390	0.260	0.244	0.217	0.252	0.165
DAMADICS	0.444	0.456	0.417	0.444	0.722	0.778	0.444	0.467	0.448	0.464	0.480
Exathlon	0.960	0.936	0.916	0.960	0.960	0.970	0.960	0.941	0.926	0.908	0.899
HAI	0.286	0.279	0.282	0.286	0.429	0.714	0.286	0.274	0.277	0.314	0.252
MIT-BIH	0.570	0.518	0.513	0.570	0.650	0.730	0.570	0.527	0.520	0.525	0.277
Petrobras 3W	0.730	0.656	0.603	0.730	0.790	0.820	0.730	0.671	0.629	0.583	0.531
RATS40K	0.470	0.442	0.423	0.470	0.630	0.700	0.470	0.451	0.434	0.416	0.193
RCAEval	0.200	0.160	0.140	0.200	0.800	1.000	0.200	0.171	0.151	0.168	0.067
ROAD	0.560	0.544	0.496	0.560	0.680	0.680	0.560	0.550	0.521	0.515	0.333
TelecomTS	0.840	0.688	0.612	0.840	0.910	0.950	0.840	0.719	0.655	0.595	0.692
Tennessee Eastman	0.530	0.510	0.499	0.530	0.550	0.610	0.530	0.514	0.505	0.481	0.404
Voraus	0.290	0.214	0.195	0.290	0.430	0.580	0.290	0.229	0.213	0.199	0.236
macro	*0.512*	*0.470*	*0.442*	*0.512*	*0.659*	*0.744*	*0.512*	*0.480*	*0.458*	*0.452*	*0.377*

Table 57: Full metric grid for **TiRex + GPC**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.660	0.662	0.661	0.660	0.670	0.670	0.660	0.662	0.661	0.706	0.592
DAMADICS	0.389	0.389	0.372	0.389	0.389	0.389	0.389	0.389	0.389	0.389	0.274
Exathlon	0.890	0.890	0.890	0.890	0.890	0.890	0.890	0.890	0.890	0.890	0.886
HAI	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.168
MIT-BIH	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.321
Petrobras 3W	0.610	0.614	0.618	0.610	0.640	0.650	0.610	0.614	0.617	0.617	0.516
RATS40K	0.300	0.370	0.381	0.300	0.630	0.690	0.300	0.362	0.375	0.377	0.176
RCAEval	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.527
ROAD	0.560	0.560	0.548	0.560	0.560	0.560	0.560	0.560	0.560	0.567	0.391
TelecomTS	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.755
Tennessee Eastman	0.490	0.490	0.490	0.490	0.490	0.490	0.490	0.490	0.490	0.490	0.497
Voraus	0.250	0.250	0.248	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.240
macro	*0.522*	*0.528*	*0.527*	*0.522*	*0.552*	*0.558*	*0.522*	*0.527*	*0.528*	*0.533*	*0.445*
$\rho\approx$10%											
CTSR	0.620	0.614	0.613	0.620	0.650	0.670	0.620	0.616	0.615	0.675	0.531
DAMADICS	0.278	0.278	0.261	0.278	0.278	0.278	0.278	0.278	0.278	0.278	0.181
Exathlon	0.890	0.890	0.890	0.890	0.890	0.890	0.890	0.890	0.890	0.890	0.886
HAI	0.286	0.286	0.286	0.286	0.286	0.286	0.286	0.286	0.286	0.286	0.207
MIT-BIH	0.740	0.740	0.740	0.740	0.740	0.740	0.740	0.740	0.740	0.740	0.507
Petrobras 3W	0.540	0.544	0.545	0.540	0.560	0.590	0.540	0.543	0.544	0.545	0.499
RATS40K	0.340	0.344	0.337	0.340	0.570	0.630	0.340	0.342	0.340	0.349	0.148
RCAEval	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.525
ROAD	0.560	0.560	0.548	0.560	0.560	0.560	0.560	0.560	0.560	0.567	0.391
TelecomTS	0.780	0.780	0.780	0.780	0.780	0.780	0.780	0.780	0.780	0.780	0.778
Tennessee Eastman	0.420	0.420	0.420	0.420	0.420	0.420	0.420	0.420	0.420	0.420	0.450
Voraus	0.200	0.200	0.198	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.207
macro	*0.514*	*0.515*	*0.511*	*0.514*	*0.538*	*0.547*	*0.514*	*0.514*	*0.514*	*0.521*	*0.443*
$\rho\approx$20%											
CTSR	0.600	0.610	0.609	0.600	0.640	0.650	0.600	0.608	0.608	0.672	0.519
DAMADICS	0.278	0.278	0.261	0.278	0.278	0.278	0.278	0.278	0.278	0.278	0.188
Exathlon	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.870
HAI	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.174
MIT-BIH	0.560	0.568	0.566	0.560	0.580	0.580	0.560	0.567	0.566	0.567	0.390
Petrobras 3W	0.540	0.536	0.540	0.540	0.540	0.580	0.540	0.537	0.539	0.539	0.495
RATS40K	0.400	0.296	0.305	0.400	0.520	0.590	0.400	0.311	0.314	0.326	0.130
RCAEval	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.518
ROAD	0.520	0.520	0.508	0.520	0.520	0.520	0.520	0.520	0.520	0.529	0.391
TelecomTS	0.810	0.810	0.810	0.810	0.810	0.810	0.810	0.810	0.810	0.810	0.799
Tennessee Eastman	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.439
Voraus	0.190	0.190	0.188	0.190	0.190	0.190	0.190	0.190	0.190	0.190	0.227
macro	*0.492*	*0.485*	*0.483*	*0.492*	*0.507*	*0.517*	*0.492*	*0.486*	*0.486*	*0.493*	*0.428*

Table 58: Full metric grid for **TiRex + MajVote**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.655	0.534
DAMADICS	0.278	0.278	0.278	0.278	0.278	0.278	0.278	0.278	0.278	0.278	0.109
Exathlon	0.870	0.870	0.870	0.870	0.870	0.870	0.870	0.870	0.870	0.870	0.818
HAI	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.161
MIT-BIH	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.334
Petrobras 3W	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.631	0.499
RATS40K	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.235
RCAEval	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.489
ROAD	0.520	0.520	0.512	0.520	0.520	0.560	0.520	0.520	0.523	0.542	0.387
TelecomTS	0.640	0.640	0.640	0.640	0.640	0.640	0.640	0.640	0.640	0.640	0.553
Tennessee Eastman	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.464
Voraus	0.240	0.240	0.238	0.240	0.240	0.240	0.240	0.240	0.240	0.240	0.206
macro	*0.516*	*0.516*	*0.516*	*0.516*	*0.516*	*0.520*	*0.516*	*0.516*	*0.517*	*0.523*	*0.399*
$\rho\approx$10%											
CTSR	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.647	0.521
DAMADICS	0.278	0.278	0.278	0.278	0.278	0.278	0.278	0.278	0.278	0.278	0.109
Exathlon	0.870	0.870	0.870	0.870	0.870	0.870	0.870	0.870	0.870	0.870	0.818
HAI	0.179	0.179	0.179	0.179	0.179	0.179	0.179	0.179	0.179	0.179	0.108
MIT-BIH	0.730	0.730	0.730	0.730	0.730	0.730	0.730	0.730	0.730	0.730	0.479
Petrobras 3W	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.451
RATS40K	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.234
RCAEval	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.477
ROAD	0.520	0.520	0.512	0.520	0.520	0.560	0.520	0.520	0.523	0.542	0.387
TelecomTS	0.650	0.650	0.650	0.650	0.650	0.650	0.650	0.650	0.650	0.650	0.583
Tennessee Eastman	0.470	0.470	0.470	0.470	0.470	0.470	0.470	0.470	0.470	0.470	0.483
Voraus	0.240	0.240	0.238	0.240	0.240	0.240	0.240	0.240	0.240	0.240	0.216
macro	*0.513*	*0.513*	*0.512*	*0.513*	*0.513*	*0.516*	*0.513*	*0.513*	*0.513*	*0.520*	*0.406*
$\rho\approx$20%											
CTSR	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.590	0.646	0.521
DAMADICS	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.095
Exathlon	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.812
HAI	0.214	0.214	0.214	0.214	0.214	0.214	0.214	0.214	0.214	0.214	0.140
MIT-BIH	0.490	0.490	0.490	0.490	0.490	0.490	0.490	0.490	0.490	0.490	0.272
Petrobras 3W	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.425
RATS40K	0.470	0.470	0.470	0.470	0.470	0.470	0.470	0.470	0.470	0.470	0.208
RCAEval	0.420	0.420	0.420	0.420	0.420	0.420	0.420	0.420	0.420	0.420	0.435
ROAD	0.440	0.440	0.432	0.440	0.440	0.480	0.440	0.440	0.443	0.467	0.367
TelecomTS	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.630	0.581
Tennessee Eastman	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.445
Voraus	0.200	0.200	0.198	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.214
macro	*0.456*	*0.456*	*0.455*	*0.456*	*0.456*	*0.459*	*0.456*	*0.456*	*0.456*	*0.463*	*0.376*

Table 59: Full metric grid for **TiRex + QPurity**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.590	0.514	0.444	0.590	0.680	0.820	0.590	0.531	0.477	0.526	0.483
DAMADICS	0.333	0.300	0.256	0.333	0.444	0.556	0.333	0.308	0.283	0.299	0.114
Exathlon	0.820	0.828	0.830	0.820	0.860	0.890	0.820	0.827	0.828	0.827	0.580
HAI	0.179	0.257	0.243	0.179	0.464	0.571	0.179	0.246	0.242	0.263	0.097
MIT-BIH	0.620	0.606	0.605	0.620	0.650	0.680	0.620	0.608	0.607	0.608	0.331
Petrobras 3W	0.670	0.614	0.587	0.670	0.700	0.740	0.670	0.627	0.604	0.566	0.529
RATS40K	0.530	0.486	0.482	0.530	0.610	0.700	0.530	0.495	0.492	0.476	0.228
RCAEval	0.520	0.520	0.510	0.520	0.540	0.640	0.520	0.520	0.514	0.513	0.489
ROAD	0.480	0.464	0.460	0.480	0.640	0.640	0.480	0.463	0.463	0.471	0.133
TelecomTS	0.710	0.632	0.576	0.710	0.780	0.850	0.710	0.647	0.603	0.560	0.552
Tennessee Eastman	0.490	0.482	0.456	0.490	0.610	0.670	0.490	0.485	0.466	0.437	0.477
Voraus	0.250	0.244	0.234	0.250	0.290	0.340	0.250	0.245	0.239	0.231	0.228
macro	*0.516*	*0.496*	*0.474*	*0.516*	*0.606*	*0.675*	*0.516*	*0.500*	*0.485*	*0.481*	*0.353*
$\rho\approx$10%											
CTSR	0.570	0.538	0.471	0.570	0.700	0.820	0.570	0.547	0.498	0.548	0.477
DAMADICS	0.278	0.289	0.256	0.278	0.389	0.444	0.278	0.286	0.272	0.274	0.095
Exathlon	0.840	0.832	0.832	0.840	0.880	0.900	0.840	0.833	0.833	0.830	0.580
HAI	0.179	0.193	0.204	0.179	0.357	0.536	0.179	0.193	0.203	0.229	0.090
MIT-BIH	0.730	0.720	0.716	0.730	0.730	0.730	0.730	0.720	0.717	0.713	0.479
Petrobras 3W	0.610	0.574	0.556	0.610	0.670	0.730	0.610	0.583	0.567	0.542	0.448
RATS40K	0.510	0.490	0.483	0.510	0.610	0.710	0.510	0.494	0.490	0.470	0.261
RCAEval	0.500	0.500	0.500	0.500	0.520	0.600	0.500	0.500	0.500	0.493	0.477
ROAD	0.520	0.480	0.460	0.520	0.600	0.640	0.520	0.484	0.475	0.484	0.412
TelecomTS	0.710	0.632	0.597	0.710	0.790	0.850	0.710	0.648	0.618	0.576	0.583
Tennessee Eastman	0.450	0.440	0.440	0.450	0.530	0.640	0.450	0.442	0.442	0.416	0.464
Voraus	0.240	0.246	0.238	0.240	0.270	0.340	0.240	0.244	0.241	0.231	0.216
macro	*0.511*	*0.494*	*0.479*	*0.511*	*0.587*	*0.662*	*0.511*	*0.498*	*0.488*	*0.484*	*0.382*
$\rho\approx$20%											
CTSR	0.570	0.542	0.479	0.570	0.690	0.800	0.570	0.550	0.504	0.553	0.495
DAMADICS	0.278	0.311	0.289	0.278	0.444	0.500	0.278	0.310	0.301	0.284	0.119
Exathlon	0.830	0.824	0.821	0.830	0.870	0.890	0.830	0.825	0.822	0.819	0.573
HAI	0.214	0.229	0.218	0.214	0.429	0.607	0.214	0.222	0.218	0.236	0.269
MIT-BIH	0.530	0.526	0.516	0.530	0.580	0.590	0.530	0.527	0.520	0.513	0.290
Petrobras 3W	0.530	0.494	0.482	0.530	0.580	0.610	0.530	0.504	0.492	0.465	0.423
RATS40K	0.460	0.432	0.419	0.460	0.570	0.680	0.460	0.436	0.428	0.414	0.247
RCAEval	0.420	0.412	0.412	0.420	0.420	0.500	0.420	0.414	0.413	0.417	0.435
ROAD	0.480	0.416	0.404	0.480	0.560	0.640	0.480	0.427	0.419	0.442	0.394
TelecomTS	0.700	0.616	0.577	0.700	0.770	0.820	0.700	0.634	0.601	0.559	0.591
Tennessee Eastman	0.390	0.388	0.377	0.390	0.430	0.430	0.390	0.390	0.382	0.359	0.427
Voraus	0.190	0.192	0.172	0.190	0.240	0.320	0.190	0.193	0.180	0.172	0.205
macro	*0.466*	*0.448*	*0.430*	*0.466*	*0.549*	*0.616*	*0.466*	*0.453*	*0.440*	*0.436*	*0.372*

Table 60: Full metric grid for **TiRex + PRF**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.600	0.508	0.445	0.600	0.820	0.870	0.600	0.526	0.477	0.557	0.490
DAMADICS	0.278	0.289	0.328	0.278	0.500	0.833	0.278	0.285	0.322	0.329	0.109
Exathlon	0.860	0.818	0.814	0.860	0.920	0.990	0.860	0.826	0.821	0.783	0.780
HAI	0.286	0.314	0.286	0.286	0.750	0.857	0.286	0.325	0.302	0.302	0.166
MIT-BIH	0.620	0.602	0.591	0.620	0.670	0.670	0.620	0.607	0.597	0.598	0.333
Petrobras 3W	0.720	0.568	0.512	0.720	0.850	0.890	0.720	0.600	0.550	0.494	0.543
RATS40K	0.450	0.454	0.417	0.450	0.740	0.820	0.450	0.454	0.433	0.416	0.258
RCAEval	0.440	0.468	0.462	0.440	0.820	0.900	0.440	0.466	0.462	0.451	0.455
ROAD	0.440	0.448	0.408	0.440	0.800	0.960	0.440	0.453	0.442	0.447	0.431
TelecomTS	0.770	0.628	0.531	0.770	0.920	0.970	0.770	0.655	0.578	0.500	0.674
Tennessee Eastman	0.380	0.382	0.358	0.380	0.700	0.780	0.380	0.384	0.367	0.332	0.434
Voraus	0.250	0.196	0.187	0.250	0.640	0.850	0.250	0.207	0.199	0.183	0.200
macro	*0.508*	*0.473*	*0.445*	*0.508*	*0.761*	*0.866*	*0.508*	*0.482*	*0.463*	*0.449*	*0.406*
$\rho\approx$10%											
CTSR	0.610	0.508	0.445	0.610	0.820	0.870	0.610	0.528	0.478	0.557	0.490
DAMADICS	0.222	0.256	0.283	0.222	0.556	0.778	0.222	0.249	0.278	0.291	0.091
Exathlon	0.860	0.818	0.811	0.860	0.920	0.980	0.860	0.826	0.818	0.777	0.780
HAI	0.250	0.321	0.271	0.250	0.750	0.821	0.250	0.325	0.290	0.282	0.164
MIT-BIH	0.700	0.718	0.723	0.700	0.760	0.790	0.700	0.715	0.720	0.726	0.471
Petrobras 3W	0.690	0.534	0.480	0.690	0.830	0.890	0.690	0.564	0.516	0.459	0.523
RATS40K	0.400	0.412	0.374	0.400	0.730	0.800	0.400	0.407	0.385	0.375	0.286
RCAEval	0.380	0.412	0.410	0.380	0.820	0.880	0.380	0.406	0.408	0.405	0.389
ROAD	0.440	0.424	0.384	0.440	0.800	0.960	0.440	0.437	0.422	0.428	0.431
TelecomTS	0.750	0.602	0.508	0.750	0.920	0.970	0.750	0.629	0.554	0.477	0.683
Tennessee Eastman	0.340	0.360	0.344	0.340	0.640	0.720	0.340	0.361	0.349	0.312	0.430
Voraus	0.240	0.190	0.167	0.240	0.630	0.820	0.240	0.197	0.180	0.166	0.195
macro	*0.490*	*0.463*	*0.433*	*0.490*	*0.765*	*0.857*	*0.490*	*0.470*	*0.450*	*0.438*	*0.411*
$\rho\approx$20%											
CTSR	0.610	0.506	0.444	0.610	0.810	0.870	0.610	0.526	0.477	0.556	0.494
DAMADICS	0.167	0.244	0.222	0.167	0.500	0.667	0.167	0.236	0.230	0.253	0.075
Exathlon	0.860	0.812	0.794	0.860	0.920	0.980	0.860	0.822	0.807	0.765	0.802
HAI	0.286	0.293	0.250	0.286	0.679	0.786	0.286	0.302	0.271	0.251	0.193
MIT-BIH	0.510	0.492	0.501	0.510	0.590	0.650	0.510	0.495	0.500	0.509	0.326
Petrobras 3W	0.590	0.496	0.442	0.590	0.800	0.890	0.590	0.516	0.471	0.423	0.482
RATS40K	0.380	0.392	0.351	0.380	0.720	0.790	0.380	0.388	0.363	0.344	0.228
RCAEval	0.340	0.380	0.380	0.340	0.780	0.880	0.340	0.377	0.378	0.373	0.375
ROAD	0.440	0.400	0.348	0.440	0.800	0.960	0.440	0.415	0.392	0.412	0.409
TelecomTS	0.740	0.592	0.484	0.740	0.930	0.970	0.740	0.623	0.537	0.464	0.655
Tennessee Eastman	0.350	0.338	0.328	0.350	0.570	0.700	0.350	0.342	0.334	0.298	0.417
Voraus	0.210	0.180	0.155	0.210	0.620	0.790	0.210	0.185	0.167	0.152	0.185
macro	*0.457*	*0.427*	*0.392*	*0.457*	*0.727*	*0.828*	*0.457*	*0.436*	*0.411*	*0.400*	*0.387*

Table 61: Full metric grid for **TiRex + Purity**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.190	0.188	0.171	0.190	0.310	0.370	0.190	0.190	0.178	0.209	0.103
DAMADICS	0.222	0.267	0.250	0.222	0.333	0.333	0.222	0.265	0.254	0.269	0.151
Exathlon	0.830	0.788	0.764	0.830	0.900	0.940	0.830	0.795	0.776	0.744	0.548
HAI	0.214	0.214	0.236	0.214	0.321	0.500	0.214	0.211	0.226	0.228	0.051
MIT-BIH	0.590	0.602	0.606	0.590	0.710	0.810	0.590	0.600	0.603	0.624	0.334
Petrobras 3W	0.480	0.446	0.432	0.480	0.560	0.630	0.480	0.454	0.441	0.430	0.255
RATS40K	0.500	0.484	0.472	0.500	0.630	0.700	0.500	0.489	0.480	0.441	0.149
RCAEval	0.540	0.464	0.432	0.540	0.700	0.740	0.540	0.479	0.452	0.420	0.431
ROAD	0.480	0.480	0.488	0.480	0.480	0.640	0.480	0.480	0.486	0.487	0.130
TelecomTS	0.500	0.422	0.380	0.500	0.590	0.610	0.500	0.438	0.403	0.379	0.286
Tennessee Eastman	0.280	0.236	0.226	0.280	0.310	0.360	0.280	0.244	0.234	0.217	0.199
Voraus	0.200	0.208	0.198	0.200	0.260	0.350	0.200	0.207	0.200	0.197	0.110
macro	*0.419*	*0.400*	*0.388*	*0.419*	*0.509*	*0.582*	*0.419*	*0.404*	*0.394*	*0.387*	*0.229*
$\rho\approx$10%											
CTSR	0.210	0.186	0.171	0.210	0.310	0.370	0.210	0.191	0.180	0.210	0.089
DAMADICS	0.333	0.256	0.272	0.333	0.333	0.444	0.333	0.274	0.279	0.277	0.100
Exathlon	0.840	0.788	0.757	0.840	0.910	0.940	0.840	0.796	0.772	0.740	0.555
HAI	0.214	0.193	0.225	0.214	0.250	0.429	0.214	0.195	0.217	0.227	0.051
MIT-BIH	0.650	0.674	0.704	0.650	0.800	0.840	0.650	0.668	0.691	0.706	0.442
Petrobras 3W	0.460	0.436	0.424	0.460	0.530	0.590	0.460	0.442	0.431	0.415	0.253
RATS40K	0.510	0.470	0.447	0.510	0.640	0.690	0.510	0.477	0.461	0.423	0.135
RCAEval	0.540	0.452	0.426	0.540	0.700	0.760	0.540	0.469	0.444	0.404	0.393
ROAD	0.480	0.488	0.484	0.480	0.520	0.640	0.480	0.487	0.485	0.485	0.130
TelecomTS	0.500	0.422	0.386	0.500	0.580	0.620	0.500	0.436	0.406	0.382	0.291
Tennessee Eastman	0.280	0.236	0.225	0.280	0.310	0.360	0.280	0.244	0.234	0.217	0.199
Voraus	0.230	0.210	0.194	0.230	0.260	0.340	0.230	0.213	0.201	0.192	0.110
macro	*0.437*	*0.401*	*0.393*	*0.437*	*0.512*	*0.585*	*0.437*	*0.408*	*0.400*	*0.390*	*0.229*
$\rho\approx$20%											
CTSR	0.210	0.180	0.168	0.210	0.290	0.370	0.210	0.186	0.176	0.206	0.103
DAMADICS	0.333	0.256	0.278	0.333	0.333	0.444	0.333	0.272	0.283	0.276	0.100
Exathlon	0.820	0.780	0.743	0.820	0.910	0.950	0.820	0.788	0.760	0.732	0.549
HAI	0.214	0.200	0.214	0.214	0.250	0.393	0.214	0.204	0.213	0.217	0.059
MIT-BIH	0.500	0.508	0.522	0.500	0.620	0.700	0.500	0.507	0.517	0.532	0.278
Petrobras 3W	0.390	0.400	0.392	0.390	0.490	0.530	0.390	0.400	0.394	0.387	0.239
RATS40K	0.460	0.422	0.379	0.460	0.580	0.670	0.460	0.426	0.396	0.372	0.137
RCAEval	0.500	0.456	0.402	0.500	0.700	0.780	0.500	0.465	0.423	0.409	0.365
ROAD	0.440	0.424	0.440	0.440	0.520	0.640	0.440	0.430	0.440	0.443	0.114
TelecomTS	0.490	0.404	0.365	0.490	0.540	0.600	0.490	0.421	0.388	0.367	0.301
Tennessee Eastman	0.280	0.236	0.232	0.280	0.310	0.370	0.280	0.244	0.239	0.217	0.199
Voraus	0.070	0.082	0.073	0.070	0.150	0.190	0.070	0.082	0.077	0.080	0.106
macro	*0.392*	*0.362*	*0.351*	*0.392*	*0.474*	*0.553*	*0.392*	*0.369*	*0.359*	*0.353*	*0.213*

Table 62: Full metric grid for **TiRex + NR + GPC**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.610	0.610	0.610	0.610	0.640	0.650	0.610	0.610	0.610	0.668	0.536
DAMADICS	0.389	0.389	0.372	0.389	0.389	0.389	0.389	0.389	0.389	0.389	0.274
Exathlon	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.874
HAI	0.214	0.214	0.214	0.214	0.214	0.214	0.214	0.214	0.214	0.214	0.159
MIT-BIH	0.720	0.710	0.709	0.720	0.760	0.780	0.720	0.712	0.710	0.708	0.461
Petrobras 3W	0.570	0.588	0.598	0.570	0.700	0.700	0.570	0.587	0.594	0.598	0.487
RATS40K	0.190	0.272	0.284	0.190	0.640	0.780	0.190	0.265	0.276	0.290	0.100
RCAEval	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.515
ROAD	0.520	0.520	0.528	0.520	0.520	0.600	0.520	0.520	0.533	0.560	0.332
TelecomTS	0.800	0.802	0.801	0.800	0.810	0.810	0.800	0.802	0.801	0.802	0.809
Tennessee Eastman	0.490	0.490	0.490	0.490	0.490	0.490	0.490	0.490	0.490	0.490	0.466
Voraus	0.410	0.410	0.406	0.410	0.410	0.410	0.410	0.410	0.410	0.410	0.401
macro	*0.531*	*0.539*	*0.539*	*0.531*	*0.586*	*0.607*	*0.531*	*0.538*	*0.541*	*0.549*	*0.451*
$\rho\approx$10%											
CTSR	0.630	0.614	0.612	0.630	0.650	0.670	0.630	0.617	0.615	0.671	0.557
DAMADICS	0.278	0.278	0.261	0.278	0.278	0.278	0.278	0.278	0.278	0.295	0.272
Exathlon	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.874
HAI	0.214	0.214	0.214	0.214	0.214	0.214	0.214	0.214	0.214	0.214	0.159
MIT-BIH	0.760	0.772	0.768	0.760	0.800	0.800	0.760	0.772	0.769	0.770	0.525
Petrobras 3W	0.590	0.558	0.570	0.590	0.620	0.660	0.590	0.566	0.572	0.572	0.472
RATS40K	0.230	0.274	0.249	0.230	0.530	0.570	0.230	0.264	0.251	0.273	0.093
RCAEval	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.481
ROAD	0.480	0.480	0.480	0.480	0.480	0.520	0.480	0.480	0.488	0.507	0.271
TelecomTS	0.800	0.802	0.803	0.800	0.810	0.810	0.800	0.802	0.803	0.803	0.809
Tennessee Eastman	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.463
Voraus	0.410	0.410	0.406	0.410	0.410	0.410	0.410	0.410	0.410	0.410	0.397
macro	*0.524*	*0.525*	*0.522*	*0.524*	*0.558*	*0.569*	*0.524*	*0.525*	*0.525*	*0.535*	*0.448*
$\rho\approx$20%											
CTSR	0.600	0.608	0.608	0.600	0.640	0.660	0.600	0.607	0.608	0.665	0.519
DAMADICS	0.333	0.333	0.300	0.333	0.333	0.333	0.333	0.333	0.333	0.351	0.268
Exathlon	0.930	0.930	0.930	0.930	0.930	0.930	0.930	0.930	0.930	0.930	0.870
HAI	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.115
MIT-BIH	0.610	0.656	0.652	0.610	0.710	0.710	0.610	0.649	0.649	0.650	0.401
Petrobras 3W	0.600	0.582	0.570	0.600	0.600	0.630	0.600	0.588	0.577	0.571	0.478
RATS40K	0.360	0.220	0.236	0.360	0.480	0.590	0.360	0.243	0.246	0.258	0.068
RCAEval	0.460	0.460	0.460	0.460	0.460	0.460	0.460	0.460	0.460	0.460	0.468
ROAD	0.480	0.480	0.480	0.480	0.480	0.520	0.480	0.480	0.488	0.521	0.267
TelecomTS	0.860	0.854	0.854	0.860	0.860	0.860	0.860	0.855	0.854	0.854	0.849
Tennessee Eastman	0.450	0.450	0.450	0.450	0.450	0.450	0.450	0.450	0.450	0.450	0.435
Voraus	0.370	0.370	0.368	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.345
macro	*0.516*	*0.507*	*0.504*	*0.516*	*0.538*	*0.555*	*0.516*	*0.509*	*0.509*	*0.519*	*0.424*

Table 63: Full metric grid for **TiRex + NR + MajVote**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.643	0.539
DAMADICS	0.389	0.389	0.389	0.389	0.389	0.500	0.389	0.389	0.400	0.428	0.232
Exathlon	0.930	0.930	0.930	0.930	0.930	0.930	0.930	0.930	0.930	0.930	0.836
HAI	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.392
MIT-BIH	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.357
Petrobras 3W	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.570	0.478
RATS40K	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.217
RCAEval	0.440	0.440	0.440	0.440	0.440	0.440	0.440	0.440	0.440	0.440	0.422
ROAD	0.560	0.560	0.548	0.560	0.560	0.600	0.560	0.560	0.563	0.576	0.260
TelecomTS	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.701
Tennessee Eastman	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.419
Voraus	0.480	0.480	0.481	0.480	0.480	0.520	0.480	0.480	0.483	0.486	0.393
macro	*0.567*	*0.567*	*0.566*	*0.567*	*0.567*	*0.583*	*0.567*	*0.567*	*0.568*	*0.576*	*0.437*
$\rho\approx$10%											
CTSR	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.643	0.536
DAMADICS	0.278	0.278	0.272	0.278	0.278	0.389	0.278	0.278	0.285	0.319	0.189
Exathlon	0.930	0.930	0.930	0.930	0.930	0.930	0.930	0.930	0.930	0.930	0.836
HAI	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.392
MIT-BIH	0.790	0.790	0.790	0.790	0.790	0.790	0.790	0.790	0.790	0.790	0.412
Petrobras 3W	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.450
RATS40K	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.215
RCAEval	0.440	0.440	0.440	0.440	0.440	0.440	0.440	0.440	0.440	0.440	0.422
ROAD	0.560	0.560	0.548	0.560	0.560	0.600	0.560	0.560	0.563	0.576	0.260
TelecomTS	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.704
Tennessee Eastman	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.422
Voraus	0.480	0.480	0.481	0.480	0.480	0.520	0.480	0.480	0.483	0.486	0.392
macro	*0.565*	*0.565*	*0.564*	*0.565*	*0.565*	*0.581*	*0.565*	*0.565*	*0.566*	*0.574*	*0.436*
$\rho\approx$20%											
CTSR	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.643	0.536
DAMADICS	0.333	0.333	0.328	0.333	0.333	0.444	0.333	0.333	0.341	0.352	0.203
Exathlon	0.930	0.930	0.930	0.930	0.930	0.930	0.930	0.930	0.930	0.930	0.836
HAI	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.394
MIT-BIH	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.324
Petrobras 3W	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.434
RATS40K	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.228
RCAEval	0.440	0.440	0.440	0.440	0.440	0.440	0.440	0.440	0.440	0.440	0.422
ROAD	0.560	0.560	0.548	0.560	0.560	0.600	0.560	0.560	0.563	0.576	0.264
TelecomTS	0.730	0.730	0.730	0.730	0.730	0.730	0.730	0.730	0.730	0.730	0.684
Tennessee Eastman	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.431
Voraus	0.460	0.460	0.461	0.460	0.460	0.500	0.460	0.460	0.463	0.467	0.379
macro	*0.550*	*0.550*	*0.548*	*0.550*	*0.550*	*0.566*	*0.550*	*0.550*	*0.551*	*0.557*	*0.428*

Table 64: Full metric grid for **TiRex + NR + QPurity**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.620	0.536	0.451	0.620	0.750	0.820	0.620	0.556	0.491	0.532	0.554
DAMADICS	0.444	0.367	0.339	0.444	0.667	0.833	0.444	0.388	0.373	0.384	0.296
Exathlon	0.850	0.856	0.848	0.850	0.920	0.940	0.850	0.856	0.850	0.828	0.584
HAI	0.393	0.371	0.368	0.393	0.536	0.643	0.393	0.371	0.370	0.345	0.390
MIT-BIH	0.680	0.674	0.672	0.680	0.700	0.720	0.680	0.676	0.674	0.666	0.357
Petrobras 3W	0.540	0.510	0.486	0.540	0.620	0.690	0.540	0.516	0.498	0.470	0.431
RATS40K	0.510	0.468	0.466	0.510	0.680	0.760	0.510	0.475	0.474	0.451	0.183
RCAEval	0.420	0.432	0.400	0.420	0.540	0.560	0.420	0.430	0.408	0.401	0.414
ROAD	0.520	0.528	0.504	0.520	0.720	0.920	0.520	0.531	0.525	0.520	0.141
TelecomTS	0.870	0.724	0.661	0.870	0.940	0.950	0.870	0.757	0.702	0.628	0.774
Tennessee Eastman	0.440	0.448	0.441	0.440	0.540	0.600	0.440	0.446	0.441	0.431	0.396
Voraus	0.500	0.432	0.389	0.500	0.700	0.840	0.500	0.443	0.411	0.376	0.406
macro	*0.566*	*0.529*	*0.502*	*0.566*	*0.693*	*0.773*	*0.566*	*0.537*	*0.518*	*0.503*	*0.411*
$\rho\approx$10%											
CTSR	0.640	0.558	0.479	0.640	0.750	0.830	0.640	0.574	0.514	0.555	0.569
DAMADICS	0.278	0.300	0.289	0.278	0.611	0.833	0.278	0.297	0.303	0.331	0.259
Exathlon	0.860	0.860	0.854	0.860	0.930	0.940	0.860	0.861	0.857	0.836	0.602
HAI	0.393	0.379	0.371	0.393	0.464	0.571	0.393	0.384	0.377	0.347	0.392
MIT-BIH	0.790	0.784	0.786	0.790	0.790	0.800	0.790	0.785	0.786	0.783	0.412
Petrobras 3W	0.530	0.504	0.486	0.530	0.640	0.720	0.530	0.506	0.493	0.465	0.413
RATS40K	0.490	0.460	0.459	0.490	0.640	0.730	0.490	0.464	0.465	0.443	0.206
RCAEval	0.420	0.424	0.404	0.420	0.480	0.560	0.420	0.423	0.409	0.405	0.414
ROAD	0.520	0.520	0.504	0.520	0.720	0.920	0.520	0.526	0.525	0.510	0.141
TelecomTS	0.880	0.734	0.659	0.880	0.940	0.950	0.880	0.764	0.702	0.642	0.770
Tennessee Eastman	0.440	0.438	0.434	0.440	0.520	0.590	0.440	0.439	0.436	0.428	0.370
Voraus	0.500	0.422	0.388	0.500	0.670	0.850	0.500	0.437	0.410	0.372	0.400
macro	*0.562*	*0.532*	*0.509*	*0.562*	*0.680*	*0.775*	*0.562*	*0.538*	*0.523*	*0.510*	*0.412*
$\rho\approx$20%											
CTSR	0.630	0.562	0.492	0.630	0.730	0.820	0.630	0.576	0.523	0.558	0.564
DAMADICS	0.278	0.289	0.261	0.278	0.667	0.833	0.278	0.290	0.279	0.319	0.205
Exathlon	0.880	0.864	0.855	0.880	0.920	0.940	0.880	0.867	0.860	0.836	0.620
HAI	0.393	0.393	0.379	0.393	0.500	0.607	0.393	0.394	0.383	0.370	0.394
MIT-BIH	0.620	0.608	0.603	0.620	0.640	0.640	0.620	0.611	0.606	0.609	0.322
Petrobras 3W	0.490	0.482	0.470	0.490	0.600	0.720	0.490	0.480	0.473	0.446	0.353
RATS40K	0.470	0.450	0.442	0.470	0.630	0.730	0.470	0.452	0.448	0.424	0.219
RCAEval	0.420	0.424	0.406	0.420	0.480	0.540	0.420	0.424	0.412	0.405	0.414
ROAD	0.520	0.512	0.492	0.520	0.760	0.920	0.520	0.521	0.516	0.494	0.283
TelecomTS	0.880	0.722	0.650	0.880	0.930	0.950	0.880	0.755	0.694	0.634	0.778
Tennessee Eastman	0.430	0.426	0.425	0.430	0.500	0.540	0.430	0.426	0.426	0.422	0.379
Voraus	0.440	0.394	0.335	0.440	0.770	0.850	0.440	0.405	0.362	0.318	0.366
macro	*0.538*	*0.510*	*0.484*	*0.538*	*0.677*	*0.758*	*0.538*	*0.517*	*0.498*	*0.486*	*0.408*

Table 65: Full metric grid for **TiRex + NR + PRF**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.470	0.438	0.415	0.470	0.750	0.860	0.470	0.442	0.426	0.500	0.344
DAMADICS	0.444	0.367	0.278	0.444	0.667	0.667	0.444	0.382	0.324	0.369	0.214
Exathlon	0.850	0.840	0.838	0.850	0.950	0.960	0.850	0.842	0.840	0.812	0.682
HAI	0.321	0.329	0.311	0.321	0.643	0.857	0.321	0.335	0.322	0.299	0.259
MIT-BIH	0.690	0.708	0.702	0.690	0.750	0.760	0.690	0.706	0.703	0.702	0.379
Petrobras 3W	0.480	0.480	0.431	0.480	0.680	0.740	0.480	0.482	0.447	0.411	0.404
RATS40K	0.410	0.408	0.378	0.410	0.730	0.800	0.410	0.406	0.389	0.370	0.158
RCAEval	0.360	0.348	0.332	0.360	0.600	0.700	0.360	0.350	0.339	0.343	0.348
ROAD	0.400	0.400	0.380	0.400	0.800	0.920	0.400	0.400	0.402	0.424	0.195
TelecomTS	0.550	0.480	0.436	0.550	0.790	0.860	0.550	0.496	0.460	0.420	0.535
Tennessee Eastman	0.390	0.388	0.377	0.390	0.580	0.650	0.390	0.389	0.381	0.355	0.364
Voraus	0.460	0.322	0.274	0.460	0.810	0.870	0.460	0.349	0.307	0.272	0.400
macro	*0.485*	*0.459*	*0.429*	*0.485*	*0.729*	*0.804*	*0.485*	*0.465*	*0.445*	*0.440*	*0.357*
$\rho\approx$10%											
CTSR	0.470	0.438	0.414	0.470	0.750	0.860	0.470	0.442	0.426	0.500	0.344
DAMADICS	0.444	0.344	0.283	0.444	0.667	0.722	0.444	0.363	0.324	0.351	0.214
Exathlon	0.850	0.840	0.837	0.850	0.950	0.960	0.850	0.842	0.839	0.810	0.682
HAI	0.321	0.336	0.318	0.321	0.643	0.857	0.321	0.336	0.325	0.296	0.257
MIT-BIH	0.800	0.780	0.777	0.800	0.820	0.830	0.800	0.784	0.781	0.776	0.407
Petrobras 3W	0.480	0.472	0.424	0.480	0.690	0.760	0.480	0.475	0.440	0.401	0.384
RATS40K	0.430	0.390	0.368	0.430	0.730	0.820	0.430	0.396	0.381	0.356	0.163
RCAEval	0.360	0.328	0.308	0.360	0.580	0.680	0.360	0.337	0.321	0.321	0.361
ROAD	0.440	0.392	0.364	0.440	0.800	0.920	0.440	0.399	0.392	0.403	0.141
TelecomTS	0.540	0.478	0.433	0.540	0.800	0.880	0.540	0.492	0.456	0.414	0.532
Tennessee Eastman	0.380	0.390	0.374	0.380	0.590	0.670	0.380	0.390	0.379	0.350	0.371
Voraus	0.460	0.332	0.267	0.460	0.810	0.870	0.460	0.355	0.303	0.267	0.418
macro	*0.498*	*0.460*	*0.431*	*0.498*	*0.736*	*0.819*	*0.498*	*0.468*	*0.447*	*0.437*	*0.356*
$\rho\approx$20%											
CTSR	0.470	0.438	0.413	0.470	0.750	0.860	0.470	0.442	0.425	0.499	0.344
DAMADICS	0.444	0.344	0.294	0.444	0.611	0.722	0.444	0.360	0.330	0.323	0.220
Exathlon	0.840	0.838	0.831	0.840	0.950	0.960	0.840	0.838	0.833	0.804	0.676
HAI	0.321	0.343	0.311	0.321	0.643	0.821	0.321	0.341	0.321	0.288	0.260
MIT-BIH	0.560	0.594	0.596	0.560	0.660	0.700	0.560	0.587	0.591	0.592	0.315
Petrobras 3W	0.460	0.436	0.398	0.460	0.660	0.740	0.460	0.444	0.414	0.376	0.318
RATS40K	0.400	0.372	0.356	0.400	0.710	0.810	0.400	0.377	0.366	0.342	0.155
RCAEval	0.320	0.316	0.298	0.320	0.560	0.700	0.320	0.325	0.310	0.298	0.314
ROAD	0.400	0.376	0.356	0.400	0.800	0.920	0.400	0.383	0.381	0.393	0.137
TelecomTS	0.540	0.468	0.425	0.540	0.770	0.870	0.540	0.483	0.448	0.404	0.527
Tennessee Eastman	0.390	0.386	0.373	0.390	0.580	0.660	0.390	0.389	0.379	0.349	0.379
Voraus	0.400	0.312	0.255	0.400	0.770	0.870	0.400	0.328	0.284	0.251	0.378
macro	*0.462*	*0.435*	*0.409*	*0.462*	*0.705*	*0.803*	*0.462*	*0.442*	*0.424*	*0.410*	*0.335*

Table 66: Full metric grid for **TiRex + NR + Purity**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.290	0.254	0.223	0.290	0.440	0.500	0.290	0.261	0.237	0.276	0.170
DAMADICS	0.333	0.333	0.267	0.333	0.500	0.556	0.333	0.333	0.285	0.273	0.242
Exathlon	0.850	0.844	0.791	0.850	0.970	0.990	0.850	0.845	0.809	0.749	0.659
HAI	0.250	0.221	0.243	0.250	0.500	0.679	0.250	0.221	0.235	0.234	0.107
MIT-BIH	0.660	0.672	0.662	0.660	0.750	0.820	0.660	0.670	0.663	0.664	0.367
Petrobras 3W	0.500	0.460	0.427	0.500	0.610	0.650	0.500	0.466	0.441	0.411	0.357
RATS40K	0.500	0.460	0.435	0.500	0.630	0.710	0.500	0.471	0.451	0.407	0.177
RCAEval	0.380	0.344	0.326	0.380	0.620	0.720	0.380	0.348	0.335	0.321	0.283
ROAD	0.480	0.496	0.480	0.480	0.520	0.600	0.480	0.491	0.484	0.478	0.130
TelecomTS	0.780	0.638	0.559	0.780	0.920	0.930	0.780	0.668	0.603	0.534	0.612
Tennessee Eastman	0.300	0.282	0.284	0.300	0.420	0.490	0.300	0.287	0.287	0.278	0.181
Voraus	0.290	0.280	0.272	0.290	0.570	0.670	0.290	0.282	0.277	0.264	0.152
macro	*0.468*	*0.440*	*0.414*	*0.468*	*0.621*	*0.693*	*0.468*	*0.445*	*0.426*	*0.407*	*0.286*
$\rho\approx$10%											
CTSR	0.300	0.260	0.223	0.300	0.440	0.500	0.300	0.267	0.239	0.279	0.177
DAMADICS	0.333	0.300	0.283	0.333	0.500	0.667	0.333	0.307	0.293	0.255	0.208
Exathlon	0.830	0.842	0.778	0.830	0.950	0.980	0.830	0.840	0.797	0.743	0.659
HAI	0.214	0.214	0.218	0.214	0.357	0.607	0.214	0.216	0.219	0.229	0.143
MIT-BIH	0.780	0.754	0.748	0.780	0.850	0.910	0.780	0.758	0.753	0.743	0.412
Petrobras 3W	0.450	0.426	0.398	0.450	0.570	0.620	0.450	0.428	0.408	0.391	0.313
RATS40K	0.500	0.454	0.422	0.500	0.640	0.680	0.500	0.466	0.440	0.397	0.176
RCAEval	0.380	0.372	0.344	0.380	0.620	0.700	0.380	0.373	0.354	0.337	0.374
ROAD	0.480	0.496	0.472	0.480	0.560	0.600	0.480	0.495	0.483	0.482	0.130
TelecomTS	0.770	0.638	0.560	0.770	0.900	0.930	0.770	0.666	0.602	0.533	0.603
Tennessee Eastman	0.300	0.284	0.279	0.300	0.420	0.480	0.300	0.288	0.283	0.274	0.187
Voraus	0.280	0.272	0.245	0.280	0.530	0.600	0.280	0.275	0.256	0.245	0.169
macro	*0.468*	*0.443*	*0.414*	*0.468*	*0.611*	*0.689*	*0.468*	*0.448*	*0.427*	*0.409*	*0.296*
$\rho\approx$20%											
CTSR	0.300	0.250	0.221	0.300	0.440	0.500	0.300	0.259	0.236	0.276	0.173
DAMADICS	0.389	0.300	0.228	0.389	0.500	0.667	0.389	0.317	0.262	0.247	0.204
Exathlon	0.830	0.834	0.765	0.830	0.950	0.980	0.830	0.835	0.787	0.734	0.651
HAI	0.214	0.221	0.225	0.214	0.357	0.607	0.214	0.219	0.222	0.218	0.099
MIT-BIH	0.520	0.594	0.607	0.520	0.680	0.780	0.520	0.582	0.595	0.609	0.320
Petrobras 3W	0.410	0.398	0.379	0.410	0.510	0.580	0.410	0.401	0.387	0.369	0.264
RATS40K	0.490	0.424	0.402	0.490	0.630	0.670	0.490	0.440	0.420	0.379	0.155
RCAEval	0.340	0.368	0.344	0.340	0.620	0.660	0.340	0.365	0.350	0.331	0.338
ROAD	0.480	0.472	0.436	0.480	0.560	0.560	0.480	0.477	0.454	0.435	0.130
TelecomTS	0.750	0.612	0.542	0.750	0.870	0.930	0.750	0.645	0.585	0.515	0.587
Tennessee Eastman	0.290	0.280	0.270	0.290	0.400	0.470	0.290	0.282	0.275	0.269	0.200
Voraus	0.120	0.142	0.136	0.120	0.340	0.480	0.120	0.135	0.135	0.138	0.175
macro	*0.428*	*0.408*	*0.380*	*0.428*	*0.571*	*0.657*	*0.428*	*0.413*	*0.392*	*0.377*	*0.275*

Table 67: Full metric grid for **Chronos-2 + NR + GPC**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.622	0.488
DAMADICS	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.240	0.095
Exathlon	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.877
HAI	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.477
MIT-BIH	0.770	0.770	0.770	0.770	0.770	0.770	0.770	0.770	0.770	0.770	0.406
Petrobras 3W	0.650	0.650	0.650	0.650	0.650	0.650	0.650	0.650	0.650	0.650	0.565
RATS40K	0.460	0.464	0.467	0.460	0.470	0.510	0.460	0.464	0.472	0.485	0.241
RCAEval	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.529
ROAD	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.130
TelecomTS	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.714
Tennessee Eastman	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.550	0.534
Voraus	0.560	0.560	0.556	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.468
macro	*0.581*	*0.581*	*0.581*	*0.581*	*0.582*	*0.585*	*0.581*	*0.581*	*0.582*	*0.591*	*0.460*
$\rho\approx$10%											
CTSR	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.613	0.472
DAMADICS	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.184	0.079
Exathlon	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.877
HAI	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.477
MIT-BIH	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.427
Petrobras 3W	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.610	0.506
RATS40K	0.430	0.434	0.430	0.430	0.440	0.450	0.430	0.434	0.437	0.442	0.241
RCAEval	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.529
ROAD	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.130
TelecomTS	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.750	0.717
Tennessee Eastman	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.521
Voraus	0.560	0.560	0.556	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.472
macro	*0.572*	*0.573*	*0.572*	*0.572*	*0.573*	*0.574*	*0.572*	*0.573*	*0.573*	*0.582*	*0.454*
$\rho\approx$20%											
CTSR	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.617	0.481
DAMADICS	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.083
Exathlon	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.877
HAI	0.464	0.464	0.464	0.464	0.464	0.464	0.464	0.464	0.464	0.464	0.451
MIT-BIH	0.780	0.780	0.780	0.780	0.780	0.780	0.780	0.780	0.780	0.780	0.414
Petrobras 3W	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.491
RATS40K	0.450	0.448	0.442	0.450	0.450	0.460	0.450	0.450	0.451	0.454	0.251
RCAEval	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.563
ROAD	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.130
TelecomTS	0.730	0.730	0.730	0.730	0.730	0.730	0.730	0.730	0.730	0.730	0.702
Tennessee Eastman	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.496
Voraus	0.560	0.560	0.556	0.560	0.560	0.560	0.560	0.560	0.560	0.560	0.475
macro	*0.563*	*0.562*	*0.562*	*0.563*	*0.563*	*0.563*	*0.563*	*0.563*	*0.563*	*0.569*	*0.451*

Table 68: Full metric grid for **Chronos-2 + NR + MajVote**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.591	0.465
DAMADICS	0.278	0.278	0.283	0.278	0.278	0.333	0.278	0.278	0.281	0.310	0.139
Exathlon	0.870	0.870	0.870	0.870	0.870	0.870	0.870	0.870	0.870	0.870	0.715
HAI	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.401
MIT-BIH	0.760	0.760	0.760	0.760	0.760	0.760	0.760	0.760	0.760	0.760	0.396
Petrobras 3W	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.442
RATS40K	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.158
RCAEval	0.460	0.460	0.460	0.460	0.460	0.460	0.460	0.460	0.460	0.460	0.456
ROAD	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.130
TelecomTS	0.710	0.710	0.710	0.710	0.710	0.710	0.710	0.710	0.710	0.710	0.659
Tennessee Eastman	0.470	0.470	0.470	0.470	0.470	0.470	0.470	0.470	0.470	0.470	0.416
Voraus	0.550	0.550	0.549	0.550	0.550	0.560	0.550	0.550	0.551	0.552	0.437
macro	*0.545*	*0.545*	*0.545*	*0.545*	*0.545*	*0.550*	*0.545*	*0.545*	*0.545*	*0.552*	*0.401*
$\rho\approx$10%											
CTSR	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.583	0.454
DAMADICS	0.222	0.222	0.228	0.222	0.222	0.278	0.222	0.222	0.226	0.231	0.118
Exathlon	0.870	0.870	0.870	0.870	0.870	0.870	0.870	0.870	0.870	0.870	0.715
HAI	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.401
MIT-BIH	0.840	0.840	0.840	0.840	0.840	0.840	0.840	0.840	0.840	0.840	0.434
Petrobras 3W	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.440
RATS40K	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.159
RCAEval	0.440	0.440	0.440	0.440	0.440	0.440	0.440	0.440	0.440	0.440	0.439
ROAD	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.130
TelecomTS	0.710	0.710	0.710	0.710	0.710	0.710	0.710	0.710	0.710	0.710	0.662
Tennessee Eastman	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.459
Voraus	0.510	0.510	0.509	0.510	0.510	0.520	0.510	0.510	0.511	0.512	0.410
macro	*0.546*	*0.546*	*0.546*	*0.546*	*0.546*	*0.551*	*0.546*	*0.546*	*0.546*	*0.551*	*0.402*
$\rho\approx$20%											
CTSR	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.590	0.465
DAMADICS	0.167	0.167	0.172	0.167	0.167	0.222	0.167	0.167	0.170	0.175	0.094
Exathlon	0.870	0.870	0.870	0.870	0.870	0.870	0.870	0.870	0.870	0.870	0.715
HAI	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.429	0.401
MIT-BIH	0.800	0.800	0.800	0.800	0.800	0.800	0.800	0.800	0.800	0.800	0.423
Petrobras 3W	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.373
RATS40K	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.162
RCAEval	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.408
ROAD	0.520	0.520	0.524	0.520	0.520	0.560	0.520	0.520	0.529	0.531	0.212
TelecomTS	0.710	0.710	0.710	0.710	0.710	0.710	0.710	0.710	0.710	0.710	0.670
Tennessee Eastman	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.492
Voraus	0.490	0.490	0.489	0.490	0.490	0.500	0.490	0.490	0.491	0.492	0.393
macro	*0.533*	*0.533*	*0.534*	*0.533*	*0.533*	*0.542*	*0.533*	*0.533*	*0.534*	*0.539*	*0.401*

Table 69: Full metric grid for **Chronos-2 + NR + QPurity**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.550	0.444	0.368	0.550	0.670	0.790	0.550	0.468	0.407	0.446	0.466
DAMADICS	0.389	0.333	0.300	0.389	0.500	0.611	0.389	0.346	0.325	0.329	0.243
Exathlon	0.860	0.842	0.837	0.860	0.870	0.900	0.860	0.844	0.840	0.821	0.629
HAI	0.357	0.393	0.371	0.357	0.643	0.679	0.357	0.393	0.378	0.363	0.352
MIT-BIH	0.750	0.734	0.744	0.750	0.760	0.810	0.750	0.737	0.743	0.738	0.396
Petrobras 3W	0.620	0.520	0.488	0.620	0.750	0.800	0.620	0.540	0.511	0.468	0.500
RATS40K	0.490	0.482	0.456	0.490	0.660	0.750	0.490	0.484	0.468	0.436	0.182
RCAEval	0.460	0.448	0.450	0.460	0.540	0.760	0.460	0.453	0.452	0.438	0.456
ROAD	0.480	0.480	0.468	0.480	0.600	0.640	0.480	0.479	0.473	0.470	0.130
TelecomTS	0.710	0.658	0.603	0.710	0.910	0.920	0.710	0.669	0.627	0.596	0.603
Tennessee Eastman	0.470	0.440	0.427	0.470	0.560	0.620	0.470	0.444	0.434	0.418	0.480
Voraus	0.490	0.510	0.468	0.490	0.760	0.820	0.490	0.509	0.481	0.449	0.455
macro	*0.552*	*0.524*	*0.498*	*0.552*	*0.685*	*0.758*	*0.552*	*0.531*	*0.512*	*0.498*	*0.408*
$\rho\approx$10%											
CTSR	0.550	0.456	0.388	0.550	0.680	0.800	0.550	0.475	0.422	0.460	0.461
DAMADICS	0.222	0.233	0.200	0.222	0.444	0.611	0.222	0.235	0.217	0.230	0.205
Exathlon	0.850	0.846	0.841	0.850	0.880	0.920	0.850	0.847	0.843	0.828	0.629
HAI	0.393	0.400	0.379	0.393	0.607	0.643	0.393	0.400	0.385	0.359	0.378
MIT-BIH	0.840	0.840	0.837	0.840	0.840	0.860	0.840	0.840	0.838	0.835	0.434
Petrobras 3W	0.570	0.522	0.475	0.570	0.760	0.770	0.570	0.530	0.495	0.459	0.448
RATS40K	0.490	0.482	0.452	0.490	0.620	0.750	0.490	0.484	0.466	0.440	0.229
RCAEval	0.460	0.440	0.434	0.460	0.520	0.680	0.460	0.443	0.437	0.429	0.457
ROAD	0.480	0.472	0.472	0.480	0.560	0.640	0.480	0.473	0.474	0.460	0.130
TelecomTS	0.720	0.662	0.610	0.720	0.910	0.930	0.720	0.674	0.633	0.605	0.612
Tennessee Eastman	0.450	0.422	0.413	0.450	0.520	0.580	0.450	0.427	0.419	0.407	0.436
Voraus	0.520	0.496	0.452	0.520	0.730	0.810	0.520	0.502	0.470	0.443	0.387
macro	*0.545*	*0.523*	*0.496*	*0.545*	*0.673*	*0.749*	*0.545*	*0.528*	*0.508*	*0.496*	*0.400*
$\rho\approx$20%											
CTSR	0.550	0.464	0.394	0.550	0.670	0.790	0.550	0.483	0.428	0.467	0.454
DAMADICS	0.167	0.200	0.167	0.167	0.389	0.500	0.167	0.187	0.175	0.198	0.183
Exathlon	0.850	0.846	0.840	0.850	0.880	0.920	0.850	0.846	0.842	0.824	0.629
HAI	0.393	0.407	0.389	0.393	0.500	0.643	0.393	0.406	0.395	0.372	0.401
MIT-BIH	0.810	0.810	0.809	0.810	0.840	0.850	0.810	0.811	0.810	0.804	0.429
Petrobras 3W	0.570	0.486	0.452	0.570	0.720	0.750	0.570	0.501	0.472	0.433	0.408
RATS40K	0.490	0.476	0.440	0.490	0.630	0.770	0.490	0.479	0.455	0.426	0.233
RCAEval	0.400	0.392	0.380	0.400	0.520	0.660	0.400	0.394	0.385	0.376	0.403
ROAD	0.520	0.512	0.456	0.520	0.600	0.600	0.520	0.511	0.477	0.471	0.212
TelecomTS	0.710	0.658	0.603	0.710	0.910	0.930	0.710	0.669	0.626	0.599	0.597
Tennessee Eastman	0.440	0.412	0.407	0.440	0.490	0.580	0.440	0.416	0.411	0.399	0.389
Voraus	0.480	0.462	0.405	0.480	0.760	0.840	0.480	0.469	0.427	0.397	0.431
macro	*0.532*	*0.510*	*0.478*	*0.532*	*0.659*	*0.736*	*0.532*	*0.514*	*0.492*	*0.481*	*0.397*

Table 70: Full metric grid for **Chronos-2 + NR + PRF**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.420	0.372	0.318	0.420	0.700	0.780	0.420	0.382	0.341	0.405	0.396
DAMADICS	0.222	0.278	0.294	0.222	0.667	0.667	0.222	0.266	0.282	0.289	0.095
Exathlon	0.840	0.854	0.846	0.840	0.930	0.940	0.840	0.850	0.846	0.822	0.714
HAI	0.393	0.393	0.361	0.393	0.750	0.821	0.393	0.389	0.369	0.335	0.398
MIT-BIH	0.750	0.738	0.740	0.750	0.780	0.810	0.750	0.739	0.740	0.734	0.392
Petrobras 3W	0.450	0.442	0.423	0.450	0.730	0.780	0.450	0.447	0.432	0.397	0.308
RATS40K	0.490	0.382	0.377	0.490	0.740	0.820	0.490	0.400	0.394	0.369	0.169
RCAEval	0.380	0.428	0.408	0.380	0.740	0.820	0.380	0.421	0.409	0.408	0.398
ROAD	0.480	0.408	0.392	0.480	0.680	0.840	0.480	0.419	0.410	0.434	0.176
TelecomTS	0.670	0.572	0.529	0.670	0.890	0.960	0.670	0.591	0.554	0.491	0.609
Tennessee Eastman	0.470	0.410	0.393	0.470	0.640	0.700	0.470	0.419	0.404	0.385	0.408
Voraus	0.450	0.390	0.352	0.450	0.890	0.940	0.450	0.406	0.377	0.347	0.481
macro	*0.501*	*0.472*	*0.453*	*0.501*	*0.761*	*0.823*	*0.501*	*0.477*	*0.463*	*0.451*	*0.379*
$\rho\approx$10%											
CTSR	0.410	0.374	0.319	0.410	0.700	0.780	0.410	0.382	0.341	0.405	0.385
DAMADICS	0.278	0.267	0.256	0.278	0.611	0.722	0.278	0.264	0.259	0.274	0.095
Exathlon	0.840	0.852	0.846	0.840	0.930	0.940	0.840	0.849	0.846	0.821	0.714
HAI	0.429	0.386	0.350	0.429	0.750	0.821	0.429	0.388	0.363	0.325	0.428
MIT-BIH	0.850	0.830	0.825	0.850	0.850	0.890	0.850	0.833	0.829	0.824	0.433
Petrobras 3W	0.430	0.416	0.408	0.430	0.710	0.800	0.430	0.421	0.414	0.380	0.314
RATS40K	0.480	0.374	0.369	0.480	0.750	0.820	0.480	0.391	0.385	0.363	0.168
RCAEval	0.400	0.404	0.410	0.400	0.680	0.820	0.400	0.408	0.410	0.397	0.397
ROAD	0.480	0.392	0.376	0.480	0.720	0.840	0.480	0.405	0.396	0.411	0.176
TelecomTS	0.650	0.570	0.520	0.650	0.910	0.960	0.650	0.588	0.546	0.481	0.603
Tennessee Eastman	0.450	0.410	0.389	0.450	0.650	0.700	0.450	0.417	0.400	0.381	0.382
Voraus	0.450	0.384	0.351	0.450	0.890	0.940	0.450	0.401	0.374	0.342	0.476
macro	*0.512*	*0.472*	*0.452*	*0.512*	*0.763*	*0.836*	*0.512*	*0.479*	*0.464*	*0.450*	*0.381*
$\rho\approx$20%											
CTSR	0.410	0.374	0.319	0.410	0.700	0.780	0.410	0.382	0.341	0.404	0.382
DAMADICS	0.167	0.256	0.228	0.167	0.611	0.722	0.167	0.235	0.224	0.237	0.079
Exathlon	0.840	0.850	0.844	0.840	0.930	0.940	0.840	0.847	0.844	0.819	0.714
HAI	0.393	0.371	0.339	0.393	0.714	0.821	0.393	0.370	0.348	0.316	0.403
MIT-BIH	0.780	0.786	0.784	0.780	0.830	0.850	0.780	0.784	0.783	0.774	0.417
Petrobras 3W	0.420	0.402	0.387	0.420	0.690	0.780	0.420	0.409	0.397	0.363	0.308
RATS40K	0.460	0.372	0.369	0.460	0.770	0.810	0.460	0.388	0.384	0.355	0.165
RCAEval	0.400	0.396	0.378	0.400	0.700	0.800	0.400	0.398	0.384	0.374	0.423
ROAD	0.520	0.392	0.372	0.520	0.680	0.800	0.520	0.417	0.398	0.398	0.184
TelecomTS	0.670	0.572	0.520	0.670	0.910	0.950	0.670	0.592	0.549	0.477	0.639
Tennessee Eastman	0.420	0.416	0.386	0.420	0.650	0.690	0.420	0.418	0.396	0.376	0.391
Voraus	0.420	0.380	0.345	0.420	0.880	0.940	0.420	0.390	0.364	0.336	0.473
macro	*0.492*	*0.464*	*0.439*	*0.492*	*0.755*	*0.824*	*0.492*	*0.469*	*0.451*	*0.436*	*0.381*

Table 71: Full metric grid for **Chronos-2 + NR + Purity**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.270	0.206	0.176	0.270	0.370	0.430	0.270	0.219	0.194	0.231	0.142
DAMADICS	0.167	0.178	0.217	0.167	0.389	0.611	0.167	0.171	0.203	0.214	0.079
Exathlon	0.830	0.826	0.804	0.830	0.930	0.960	0.830	0.828	0.812	0.781	0.573
HAI	0.357	0.343	0.325	0.357	0.679	0.750	0.357	0.354	0.337	0.331	0.324
MIT-BIH	0.730	0.726	0.717	0.730	0.820	0.850	0.730	0.727	0.720	0.715	0.396
Petrobras 3W	0.540	0.468	0.426	0.540	0.650	0.700	0.540	0.484	0.449	0.405	0.296
RATS40K	0.490	0.478	0.441	0.490	0.640	0.730	0.490	0.483	0.457	0.420	0.160
RCAEval	0.340	0.332	0.342	0.340	0.560	0.720	0.340	0.338	0.343	0.336	0.304
ROAD	0.480	0.472	0.460	0.480	0.480	0.480	0.480	0.475	0.465	0.464	0.130
TelecomTS	0.650	0.570	0.541	0.650	0.800	0.860	0.650	0.587	0.561	0.517	0.510
Tennessee Eastman	0.280	0.290	0.291	0.280	0.410	0.440	0.280	0.290	0.291	0.276	0.154
Voraus	0.430	0.352	0.342	0.430	0.540	0.650	0.430	0.367	0.357	0.338	0.254
macro	*0.464*	*0.437*	*0.423*	*0.464*	*0.606*	*0.682*	*0.464*	*0.444*	*0.432*	*0.419*	*0.277*
$\rho\approx$10%											
CTSR	0.270	0.210	0.179	0.270	0.380	0.430	0.270	0.222	0.196	0.232	0.146
DAMADICS	0.111	0.167	0.211	0.111	0.389	0.556	0.111	0.154	0.189	0.223	0.079
Exathlon	0.830	0.834	0.811	0.830	0.930	0.950	0.830	0.834	0.818	0.779	0.576
HAI	0.321	0.343	0.361	0.321	0.714	0.750	0.321	0.345	0.357	0.333	0.304
MIT-BIH	0.780	0.808	0.805	0.780	0.880	0.900	0.780	0.803	0.803	0.807	0.431
Petrobras 3W	0.530	0.460	0.400	0.530	0.650	0.670	0.530	0.472	0.426	0.390	0.262
RATS40K	0.500	0.470	0.426	0.500	0.630	0.720	0.500	0.479	0.447	0.407	0.165
RCAEval	0.480	0.400	0.378	0.480	0.660	0.740	0.480	0.413	0.394	0.359	0.370
ROAD	0.480	0.472	0.460	0.480	0.480	0.520	0.480	0.475	0.465	0.462	0.130
TelecomTS	0.630	0.566	0.536	0.630	0.800	0.860	0.630	0.579	0.554	0.511	0.492
Tennessee Eastman	0.280	0.286	0.288	0.280	0.400	0.440	0.280	0.287	0.288	0.274	0.154
Voraus	0.430	0.362	0.332	0.430	0.570	0.640	0.430	0.371	0.348	0.330	0.230
macro	*0.470*	*0.448*	*0.432*	*0.470*	*0.624*	*0.681*	*0.470*	*0.453*	*0.440*	*0.426*	*0.278*
$\rho\approx$20%											
CTSR	0.270	0.206	0.176	0.270	0.370	0.420	0.270	0.219	0.194	0.229	0.142
DAMADICS	0.111	0.111	0.161	0.111	0.333	0.611	0.111	0.111	0.146	0.171	0.029
Exathlon	0.830	0.838	0.810	0.830	0.930	0.950	0.830	0.837	0.818	0.779	0.576
HAI	0.357	0.350	0.336	0.357	0.679	0.786	0.357	0.354	0.345	0.330	0.373
MIT-BIH	0.790	0.774	0.782	0.790	0.830	0.890	0.790	0.778	0.782	0.785	0.413
Petrobras 3W	0.480	0.410	0.369	0.480	0.600	0.670	0.480	0.424	0.390	0.359	0.276
RATS40K	0.460	0.424	0.388	0.460	0.620	0.700	0.460	0.434	0.406	0.364	0.170
RCAEval	0.440	0.360	0.354	0.440	0.620	0.740	0.440	0.376	0.367	0.344	0.340
ROAD	0.480	0.448	0.444	0.480	0.480	0.520	0.480	0.454	0.449	0.426	0.133
TelecomTS	0.630	0.554	0.523	0.630	0.780	0.840	0.630	0.569	0.543	0.500	0.492
Tennessee Eastman	0.280	0.282	0.282	0.280	0.390	0.420	0.280	0.284	0.284	0.270	0.156
Voraus	0.310	0.292	0.262	0.310	0.550	0.640	0.310	0.297	0.275	0.259	0.282
macro	*0.453*	*0.421*	*0.407*	*0.453*	*0.598*	*0.682*	*0.453*	*0.428*	*0.417*	*0.401*	*0.282*

Table 72: Full metric grid for **MR-Hydra**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.750	0.770	0.768	0.750	0.800	0.800	0.750	0.767	0.767	0.783	0.743
DAMADICS	0.667	0.667	0.667	0.667	0.667	0.778	0.667	0.667	0.689	0.712	0.478
Exathlon	0.960	0.966	0.965	0.960	0.970	0.970	0.960	0.964	0.964	0.964	0.927
HAI	0.321	0.379	0.382	0.321	0.571	0.571	0.321	0.367	0.374	0.414	0.200
MIT-BIH	0.760	0.838	0.823	0.760	0.910	0.910	0.760	0.824	0.819	0.826	0.640
Petrobras 3W	0.840	0.850	0.863	0.840	0.920	0.930	0.840	0.847	0.857	0.861	0.732
RATS40K	0.610	0.618	0.621	0.610	0.670	0.670	0.610	0.618	0.621	0.613	0.273
RCAEval	0.800	0.804	0.806	0.800	0.920	0.920	0.800	0.803	0.805	0.808	0.807
ROAD	0.800	0.768	0.732	0.800	0.920	0.920	0.800	0.819	0.832	0.858	0.725
TelecomTS	0.930	0.930	0.923	0.930	0.960	0.960	0.930	0.932	0.926	0.923	0.893
Tennessee Eastman	0.750	0.738	0.740	0.750	0.800	0.800	0.750	0.739	0.740	0.739	0.736
Voraus	0.920	0.924	0.913	0.920	0.970	0.970	0.920	0.923	0.922	0.921	0.843
macro	*0.759*	*0.771*	*0.767*	*0.759*	*0.840*	*0.850*	*0.759*	*0.772*	*0.776*	*0.785*	*0.666*
$\rho\approx$10%											
CTSR	0.400	0.146	0.073	0.400	0.510	0.510	0.400	0.463	0.463	0.463	0.370
DAMADICS	0.167	0.167	0.144	0.167	0.167	0.222	0.167	0.167	0.174	0.180	0.243
Exathlon	0.970	0.970	0.970	0.970	0.980	0.980	0.970	0.970	0.970	0.970	0.819
HAI	0.357	0.343	0.336	0.357	0.500	0.500	0.357	0.343	0.336	0.359	0.200
MIT-BIH	0.460	0.530	0.531	0.460	0.640	0.640	0.460	0.528	0.528	0.542	0.227
Petrobras 3W	0.860	0.850	0.850	0.860	0.900	0.900	0.860	0.852	0.852	0.851	0.751
RATS40K	0.630	0.600	0.590	0.630	0.650	0.660	0.630	0.605	0.597	0.603	0.256
RCAEval	0.780	0.784	0.808	0.780	0.900	0.900	0.780	0.781	0.799	0.798	0.766
ROAD	0.720	0.744	0.696	0.720	0.920	0.920	0.720	0.781	0.788	0.836	0.723
TelecomTS	0.930	0.932	0.932	0.930	0.970	0.970	0.930	0.933	0.933	0.931	0.894
Tennessee Eastman	0.690	0.690	0.692	0.690	0.710	0.720	0.690	0.690	0.692	0.692	0.712
Voraus	0.950	0.930	0.923	0.950	0.970	0.970	0.950	0.933	0.934	0.931	0.900
macro	*0.659*	*0.640*	*0.629*	*0.659*	*0.735*	*0.741*	*0.659*	*0.670*	*0.672*	*0.680*	*0.572*
$\rho\approx$20%											
CTSR	0.390	0.140	0.070	0.390	0.490	0.490	0.390	0.449	0.449	0.451	0.358
DAMADICS	0.111	0.111	0.106	0.111	0.111	0.167	0.111	0.111	0.119	0.124	0.183
Exathlon	0.970	0.972	0.972	0.970	0.980	0.990	0.970	0.972	0.972	0.971	0.822
HAI	0.429	0.336	0.329	0.429	0.464	0.464	0.429	0.348	0.338	0.326	0.306
MIT-BIH	0.840	0.818	0.809	0.840	0.870	0.870	0.840	0.822	0.814	0.817	0.524
Petrobras 3W	0.810	0.814	0.812	0.810	0.840	0.860	0.810	0.814	0.812	0.813	0.674
RATS40K	0.580	0.554	0.555	0.580	0.610	0.640	0.580	0.555	0.555	0.559	0.209
RCAEval	0.680	0.712	0.722	0.680	0.880	0.880	0.680	0.706	0.716	0.715	0.735
ROAD	0.640	0.640	0.628	0.640	0.800	0.800	0.640	0.665	0.689	0.738	0.610
TelecomTS	0.920	0.908	0.907	0.920	0.940	0.950	0.920	0.909	0.908	0.906	0.931
Tennessee Eastman	0.620	0.644	0.633	0.620	0.690	0.690	0.620	0.638	0.633	0.630	0.649
Voraus	0.920	0.886	0.884	0.920	0.930	0.950	0.920	0.891	0.893	0.896	0.749
macro	*0.659*	*0.628*	*0.619*	*0.659*	*0.717*	*0.729*	*0.659*	*0.657*	*0.658*	*0.662*	*0.562*

Table 73: Full metric grid for **RDST**, per dataset and macro-averaged, at each pollution level.

	P@1	P@5	P@10	HR@1	HR@5	HR@10	N@1	N@5	N@10	N@20	mF1
ρ=0%											
CTSR	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.673	0.605
DAMADICS	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.667	0.392
Exathlon	0.980	0.980	0.980	0.980	0.980	0.980	0.980	0.980	0.980	0.980	0.989
HAI	0.357	0.357	0.357	0.357	0.357	0.357	0.357	0.357	0.357	0.357	0.251
MIT-BIH	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.580	0.628
Petrobras 3W	0.740	0.740	0.740	0.740	0.740	0.740	0.740	0.740	0.740	0.740	0.627
RATS40K	0.440	0.440	0.440	0.440	0.440	0.440	0.440	0.440	0.440	0.440	0.114
RCAEval	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.861
ROAD	0.680	0.680	0.668	0.680	0.680	0.680	0.680	0.680	0.680	0.692	0.501
TelecomTS	0.890	0.890	0.890	0.890	0.890	0.890	0.890	0.890	0.890	0.890	0.869
Tennessee Eastman	0.720	0.720	0.720	0.720	0.720	0.720	0.720	0.720	0.720	0.720	0.709
Voraus	0.950	0.950	0.944	0.950	0.950	0.950	0.950	0.950	0.950	0.950	0.860
macro	*0.711*	*0.711*	*0.710*	*0.711*	*0.711*	*0.711*	*0.711*	*0.711*	*0.711*	*0.712*	*0.617*
$\rho\approx$10%											
CTSR	0.410	0.122	0.061	0.410	0.420	0.420	0.410	0.413	0.413	0.413	0.369
DAMADICS	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.182
Exathlon	0.980	0.980	0.980	0.980	0.980	0.980	0.980	0.980	0.980	0.980	0.989
HAI	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.205
MIT-BIH	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.530	0.275
Petrobras 3W	0.740	0.740	0.740	0.740	0.740	0.740	0.740	0.740	0.740	0.740	0.621
RATS40K	0.490	0.490	0.492	0.490	0.490	0.500	0.490	0.490	0.491	0.492	0.134
RCAEval	0.880	0.880	0.880	0.880	0.880	0.880	0.880	0.880	0.880	0.880	0.880
ROAD	0.600	0.600	0.592	0.600	0.600	0.640	0.600	0.600	0.603	0.626	0.402
TelecomTS	0.880	0.880	0.880	0.880	0.880	0.880	0.880	0.880	0.880	0.880	0.879
Tennessee Eastman	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680
Voraus	0.940	0.940	0.934	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.854
macro	*0.634*	*0.610*	*0.603*	*0.634*	*0.634*	*0.639*	*0.634*	*0.634*	*0.634*	*0.636*	*0.539*
$\rho\approx$20%											
CTSR	0.380	0.118	0.059	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.350
DAMADICS	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167
Exathlon	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.985
HAI	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.135
MIT-BIH	0.450	0.450	0.450	0.450	0.450	0.450	0.450	0.450	0.450	0.450	0.221
Petrobras 3W	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.557
RATS40K	0.430	0.430	0.430	0.430	0.430	0.430	0.430	0.430	0.430	0.430	0.114
RCAEval	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.831
ROAD	0.640	0.608	0.568	0.640	0.640	0.680	0.640	0.640	0.643	0.658	0.597
TelecomTS	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.820	0.849
Tennessee Eastman	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.645
Voraus	0.940	0.940	0.934	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.859
macro	*0.590*	*0.565*	*0.557*	*0.590*	*0.590*	*0.593*	*0.590*	*0.590*	*0.590*	*0.591*	*0.526*