Advancing Asynchronous Multivariate Time Series Forecasting: Insights from Oceanographic Data

Anonymous submission

Abstract

Asynchronous multivariate time series (AMTS) forecasting tasks are prevalent in various fields, but research on structures and methods to handle them is sparse. While most research on time series assumes fully-observed states and evenly-spaced observations, this is not the case for many real-world applications. Due to the lack of adequate definitions and representative datasets, it has become standard practice to dedicate significant effort to converting AMTS into regular series through complex data imputation processes, binning, and other techniques. Despite this, modern ML offers the opportunity for researchers to tackle time series irregularities directly, achieving better performance and more efficient data pipelines. We address this discrepancy between literature and practice in three ways: First, we properly define AMTS and the forecast problem associated with this object by gathering and improving contributions from several authors. Second, we introduce the *Port of Santos Oceanographic Data (PSOD)*, a dataset for model evaluation on AMTS forecasting and robustness to missing data. It comprises five years of measurements collected by a set of independent sensors installed in the main channel of the largest port in Latin America, exhibiting non-trivial irregularities that are not present in other AMTS datasets in the literature. Third, we evaluated models from diverse machine learning domains on the PSOD forecasting task, gaining valuable insights into the essential factors for designing models that can effectively handle AMTS. Our results provide evidence that architectures developed specifically to handle irregularities perform far better than those that have regularity assumptions in their formulation. With this work, we aim to foster advancements in learning on asynchronous data and promote the development of more flexible forecasting solutions that are fit to solve real-world problems.

1 Introduction

Asynchronous multivariate time series (AMTS) are prevalent in several fields such as climate sciences, healthcare, finance, and sales (Cooke et al. 2006; Silva et al. 2012; Zhang et al. 2017); however, despite their practical importance, ATMS are significantly underrepresented in machine learning research.

For instance, only two papers in AAAI2024 Main Track dealt with time series with varying sampling rates between variables (Yalavarthi et al. 2024; Xiao et al. 2024), and just one of them addressed the complexities of multivariate time series with both irregular sampling rates and partially-observed states.

Most researchers working on time series consider a regularized version of their data as the starting point of their contributions. However, in many instances, several hours of work that went into forcibly converting an AMTS into a regular time series are, at best, hidden in the appendix of their work. We believe this happens for historical reasons, since until recently the only tools we had to process time series were Signal Processing methods that rely on regularity assumptions (Wei 2019). Despite this, modern ML provides the opportunity to properly tackle time series irregularities directly, allowing for better performance and more efficient data pipelines.

Aspects of AMTS are discussed on several fronts and under different keywords. Alas, nomenclature, definitions and data structures used to work with these objects vary widely in the literature. Some authors discuss irregularly-sampled or irregular time series, mostly indicating that the sampling rate may vary with time (Chen et al. 2023; Mei and Eisner 2017), which is a central characteristic of the growing field of Neural Ordinary Differential Equations (Neural ODEs) (Kidger et al. 2020; Norcliffe et al. 2021; Auzina et al. 2023; Zhu et al. 2022). Some groups approach the subject from the perspective of continuous-time temporal graphs that may have explicit edges related to entities from the dataset (Kazemi et al. 2020; Huang et al. 2023), while others impose edges that represent pathways for diffusion of information (Zhang et al. 2021, 2023; Barros et al. 2024). Furthermore, a large research corpus relies on terms like missing data and data imputation (Little and Rubin 2002; Emmanuel et al. 2021; Wu et al. 2022), typically discussing regular multivariate time series that are missing some observations, but can still be regularized into a grid.

As most public multivariate time series datasets are either built using inherently regular data or are regularized before publication (Zhou et al. 2021; Reyes-Ortiz and Anguita 2013; Hebrail and Berard 2006; Vito 2008; Reiss and Stricker 2012), they do not match many real-world scenarios. Recent progress has been made with the introduction of AMTS datasets focused on classification and single-valued regression (Reyna et al. 2020; Huang et al. 2023); however, the forecasting task in more complex scenarios remains poorly understood and in fact lacking a formal definition.

To address these gaps that prevent ML techniques from properly addressing AMTS challenges, our contributions are threefold:

1. We properly define AMTS and the forecast task associated with them, and propose data structures to simplify its representation as a supervised learning task, unifying and improving several notions into a generalized structure.
2. Leveraged by our formalism, we introduce the Port of Santos Oceanographic Data (PSOD), a dataset for evaluating AMTS processing capabilities and robustness to missing data, and detail how it better represents the reality faced by many researchers working with real-world multivariate time series.
3. We evaluated models from diverse machine learning domains on the PSOD forecasting task, gaining valuable insights into the essential factors for designing models that can effectively handle AMTS.

Our findings provide evidence that architectures developed specifically to handle irregularities perform far better than those that have regularity assumptions in their formulation. More specifically, our results reveal the crucial role of time encoding in both accuracy and robustness to missing data.

2 Asynchronous Multivariate Time Series

Most authors define an MTS $\mathcal{M}$ as a sequence $\mathbf{Z} = [z_1, z_2, \cdots, z_T]$, where $z_t \in \mathbb{R}^K$. This sequence can be represented as a matrix $\tilde{\mathbf{Z}} \in \mathbb{R}^{L \times K}$, where L is the length (or number of observations/states) and K is the total number of variables in the MTS. This definition allows the use of a wide range of tools from Signal Processing (Wei 2019); however, it conceals some important assumptions.

The timestamps $\mathbf{T} = [t_1, t_2, \cdots, t_T]$ corresponding to the observations are often omitted, assuming a common and constant sampling rate across all variables, leading to a trivial $\mathbf{T} = [1, 2, \cdots, T]$. When timestamps are provided, they are typically assumed to be identical for all variables (implying that all variables share the same sampling rate and are aligned). Under this definition, any irregularities are reduced to missing values within the grid, which can be addressed through data imputation techniques.

The first step in explicitly modeling uneven observations and time intervals can be traced back to early work on Recurrent Neural Networks (RNNs) for continuous systems (Funahashi and Nakamura 1993), including Phased-LSTM (Neil, Pfeiffer, and Liu 2016), GRU-D (Che et al. 2018), and Neural ODEs (Chen et al. 2018). Some of these methods also account for missing data (i.e., partially-observed states) in the unevenly-sampled grid by introducing boolean masks as indicators for missing entries. The assumption of an underlying continuous system $\mathcal{D}_t$ from which the observations in $\mathcal{M}$ are sampled raises additional considerations. The formulation by Norcliffe et al. (2021) defines a function $F : \mathcal{T} \to \mathcal{Y}, \mathcal{Y} \subset \mathbb{R}^K$, where each time maps to a state. While this applies to the theoretical ODE that models $\mathcal{D}_t$, if $\mathcal{M}$ is obtained through a measurement apparatus, the recorded time is only an approximation of the actual time; thus, multiple events might be recorded at the same measured time. This situation is particularly relevant in applications like ML for particle physics, where several events can occur simultaneously due to the limited precision of measurement devices (Karagiorgi et al. 2022).

A second, and more pertinent, relaxation step is allowing for unaligned variables. Horn et al. (2020); Shukla and Marlin (2020); Huang, Sun, and Wang (2020) address the complexity of these scenarios by either defining each measurement $z \in \mathbb{R}^K$ of a variable m at time t as a triplet (t, z, m), or by treating each variable as an independent sequence. We adopt the latter approach, as it more accurately reflects real-world data acquisition, where AMTS data is often collected from a network of independent sensors. Figure 1 (left) illustrates how unaligned data sources complicate the process of regularizing AMTS into a grid.

A final assumption that often goes unnoticed is that AMTS comprehend data types other than tabular. Barros et al. (2024) proposes a hybrid model with RNNs and CNNs to process both images and tabular data. However, to unlock the many opportunities in multimodal AMTS for a broader range of models, the first step is to establish a definition that adequately address these situations.

Definition 2.1 (Asynchronous Multivariate Time Series)**.** An AMTS is a pair of finite sets of sequences $\mathcal{M} = (\mathcal{S}, \mathcal{T}_\mathcal{S})$. The set $\mathcal{S} = \{\mathbf{Z^i}\}_{i=1}^{|S|}$ contains $|S|$ sequences $\mathbf{Z^i} = [z^i_{t_1}, z^i_{t_2}, \cdots, z^i_{t_T}]$, where each event $z^i_t \in \mathcal{X}^i$ and is associated with a timestamp $t \in \mathbb{R}$. The set $\mathcal{T}_\mathcal{S} = \{\mathbf{T^i}\}_{i=1}^{|S|}$ contains all sequences of timestamps $\mathbf{T^i} = [t^i_1, \ldots, t^i_T]$ associated with $\mathcal{S}$.

The domain $\mathcal{X}$ varies depending on the type of data each $\mathbf{Z^i}$ represents. In this work, we focus exclusively on tabular data, where $\mathcal{X}^i \subset \mathbb{R}^{K^i}$, and K^i denotes the number of features. An example of tabular data is wind velocity measurements at a specific location, where each event captured by the sensor at time t is a vector $z^i_t \in \mathbb{R}^2$, representing wind speed and direction.

We loosely define *irregularities* in $\mathcal{M}$ as characteristics that prevent it from being represented by a matrix $L \times K$ of evenly sampled events, thereby preventing certain $\mathbf{Z^i}$ from being stacked into a single object. These irregularities can stem from differences in *sampling rate*, *missing data*, *time offsets*, and *data format*.

In contrast to the traditional approach in classical signal processing, where MTS are viewed as sequences of evenly sampled events from a continuous dynamical system $\mathcal{D}_t$, we generalize this by considering the set of sequences $\mathcal{S}$ sampled from $\mathcal{D}_t$ along with their timestamps $\mathcal{T}_\mathcal{S}$. Each sequence maintains a consistent *data format* within its domain $\mathcal{X}$, but $\mathcal{M}$ as a whole can exhibit all four types of irregularities. Although separating $\mathcal{S}$ from $\mathcal{T}_\mathcal{S}$ may seem counterintuitive at first, it greatly simplifies the definition of the Forecast Problem for AMTS, and unifies various relaxation steps introduced by previous works.

Forecasting Asynchronous Multivariate Time Series. For a given AMTS $\mathcal{M} = (\mathcal{S}, \mathcal{T}_\mathcal{S})$, consider an observer

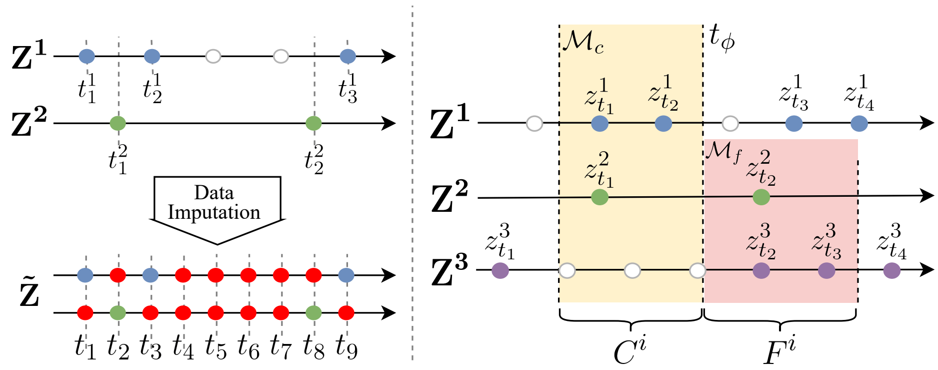

Figure 1: On the left, we depict how two independent, unaligned data sources, $\mathbf{Z^1}$ and $\mathbf{Z^2}$, are typically regularized into a grid. Although the original sequences contain only 5 events in total, the resulting matrix $\tilde{\mathbf{Z}} \in \mathbb{R}^{9\times 2}$ requires 18 entries. All the entries represented by red balls must be filled in using some form of data imputation algorithm. On the right, we show the extraction of context and forecast AMTS using window lengths C^i and F^i.

at time $t_\phi \in \mathbb{R}$ who wishes to forecast (or hindcast) the behavior of certain variables based on data available at t_ϕ. This problem can be approached by extracting a context and a forecast time series from $\mathcal{M}$. The context, $\mathcal{M}_c$, contains events that serve as inputs for the task, while the forecast series, $\mathcal{M}_f$, includes the events whose values we aim to predict.

Definition 2.2 (Forecast task on AMTS)**.** Given a context AMTS $\mathcal{M}_c = (\mathcal{S}_c, \mathcal{T}_{\mathcal{S}_c})$ and a forecast AMTS $\mathcal{M}_f = (\mathcal{S}_f, \mathcal{T}_{\mathcal{S}_f})$, the forecast task on AMTS consists in finding a function f such that $f(\mathcal{M}_c, \mathcal{T}_{\mathcal{S}_f}) = \mathcal{S}_f$.

A subtle aspect of this problem is that the observer at t_ϕ does not know when, or even if, the events in $\mathcal{S}_f$ will occur. However, the goal is not to predict the timing of these events, but rather to forecast the values they would have if they were to occur at certain predefined instants $\mathcal{T}_{\mathcal{S}_f}$. In other words, the objective is to predict the state of the underlying continuous system $\mathcal{D}_t$ at specific instants $\mathcal{T}_{\mathcal{S}_f}$ using the limited information available in $\mathcal{M}_c$.

A general procedure for extracting $\mathcal{M}_c$ and $\mathcal{M}_f$ from $\mathcal{M}$, suitable for most use cases, involves utilizing the observer time t_ϕ and defining C^i, F^i, and $O^i \in \mathbb{R}$. These represent the context, forecast, and offset time lengths, respectively. The first two, C^i and F^i, correspond to the maximum time lengths of the context and forecast windows. The latter, O^i, serves as an offset, allowing the inclusion of variables known for $t \geq t_\phi$:

$$\mathcal{M}_c = (\mathcal{S}_c, \mathcal{T}_{\mathcal{S}_c}), \tag{1}$$

$$\mathcal{S}_c = \{\, \mathbf{Z^i_t} \in \mathcal{S} \mid t_\phi - C^i \leq t + O^i < t_\phi \,\}, \tag{2}$$

$$\mathcal{T}_{\mathcal{S}_c} \vdash \mathcal{S}_c, \tag{3}$$

$$\mathcal{M}_f = (\mathcal{S}_f, \mathcal{T}_{\mathcal{S}_f}), \tag{4}$$

$$\mathcal{S}_f = \{\, \mathbf{Z^i_t} \in \mathcal{S} \mid t_\phi \leq t + O^i < t_\phi + F^i \,\}, \tag{5}$$

$$\mathcal{T}_{\mathcal{S}_f} \vdash \mathcal{S}_f, \tag{6}$$

where $\vdash$ indicates the association between events and their respective timestamps.

Figure 1 (right) illustrates this procedure. Reducing C^i is typically motivated by the computational costs associated with long context windows and the observation that, in many systems, the autocorrelation between two states decreases over time. Additionally, note that there is no requirement for any event to occur precisely at t_ϕ; it can be sampled from a uniform distribution over the entire dataset's time interval.

In a supervised learning setup, the goal is to approximate the unknown function f such that $f(\mathcal{M}_c, \mathcal{T}_{\mathcal{S}_f}) = \mathcal{S}_f$ for all t_ϕ by learning a parameterized function f_θ. The hypothesis is that f_θ can be learned by observing multiple pairs $(\mathcal{M}_c, \mathcal{M}_f)$ extracted from $\mathcal{S}$. If the events within $\mathcal{S}$ sufficiently characterize $\mathcal{D}_t$, then f_θ can be used to forecast (or hindcast) the state of this system based on the information available at any given time.

Once again, this approach stands in contrast to the traditional formulation of the problem, typically applied when $\mathcal{M}$ is grid-like. In such cases, $t_\phi \in \mathbb{Z}$ represents the step from which the forecasting task begins. Conventionally, one would utilize the information contained in the last C steps, $\tilde{\mathbf{Z}}_{t_\phi - C \leq t < t_\phi}$, to predict the next F steps, $\tilde{\mathbf{Z}}_{t_\phi \leq t < t_\phi + F}$. However, as highlighted earlier, this standard approach is fundamentally inadequate when the MTS exhibits *irregularities*. Our approach, on the other hand, embraces the complexity of asynchronous data, offering a more robust and comprehensive perspective to the forecasting challenge.

Data Structures for AMTS. In traditional supervised learning on time series, sliding windows are commonly used to construct both context and forecast intervals. For grid-like structures, it is sufficient to define $C^i, F^i \in \mathbb{N}$ and a stride based on sequence indices. However, when dealing with AMTS, this technique can produce windows of varying time lengths due to factors such as missing data and changes in sampling rates.

To more accurately capture the characteristics of AMTS, we redefine C^i and F^i as time lengths, $C^i, F^i \in \mathbb{R}$, which means the exact number of elements in any context or forecast sequence $\mathbf{Z^i}$ cannot be known beforehand. During train-

Dataset	Field	Task	Missing Data	Irregular Sampling	Timeshift	Known Covariates
GRID (Cooke et al. 2006)	Speech Recognition	C	✗	✓	✗	✗
PhysioNet 2012 (Silva et al. 2012)	Healthcare	C/R	✓	✓	✗	✗
PAMAP2 (Reiss and Stricker 2012)	Activity Monitoring	C/F/R	✓	✓	✓	✗
MIMIC-III (Johnson et al. 2016)	Healthcare	C/F/R	✓	✓	✓	✗
USHCN (Menne, Williams, and Vose 2016)	Climatology	F	✓	✗	✗	✗
Beijing Air PM2.5 (Zhang et al. 2017)	Air Quality	I/R	✓	✗	✗	✗
PhysioNet 2019 (Reyna et al. 2020)	Healthcare	C/R	✓	✓	✓	✗
TGB (Huang et al. 2023)	Multiple	C/R	✓	✓	✓	✗
PSOD (ours)	Oceanography	F	✓	✓	✓	✓

Table 1: Comparison of literature datasets. C=Classification, F=Forecast, I=Imputation, R=Single-Valued Regression

ing, t_ϕ can be sampled uniformly across the entire training dataset until model convergence. However, to ensure consistency in evaluation, the test dataset must standardize the context and forecast window pairs.

We achieve this standardization by adopting the approach proposed by Norcliffe et al. (2021), where context and forecast windows are extracted from $\mathbf{Z}^{\mathbf{i}}$ using two index sets, I_C^i and I_F^i. These sets are generated by selecting t_ϕ values and applying the expressions in 2 and 5. We refer to these as context and forecast masks—boolean masks that function as selectors applied to the original time series.

This method not only better reflects the complexities of AMTS but also serves as a valuable tool for evaluating how model performance deteriorates under varying levels of data loss. In Section 4, we apply masks with progressively higher data loss and compare the performance of various models, gaining further insights into the factors that contribute to robustness against missing data.

3 Port of Santos Oceanographic Data

To foster the development of new architectures capable of effectively modeling highly irregular AMTS, this section provides a brief overview of the datasets typically used to benchmark irregular time series and explains why these datasets do not fully capture the challenges formalized in Section 2. We then introduce the Port of Santos Oceanographic Data (PSOD) dataset, emphasizing its unique and novel characteristics.

AMTS Datasets in the Literature. The limited research on AMTS in the ML literature has largely relied on a small set of datasets, with the most popular ones listed in Table 1. This is a stark contrast to the abundance of datasets available for regular time series, which can be found by the dozens, spanning various themes and sizes. However, as previously mentioned, real-world scenarios rarely align with the assumption of completely regular multivariate time series. This discrepancy has resulted in a lack of innovative architectures dedicated to AMTS and a sense of resignation among researchers, who often accept the lengthy process of data preprocessing to forcibly regularize AMTS as an unavoidable part of their work.

Our contribution addresses this gap by introducing a dataset specifically designed for forecasting highly irregular AMTS, paving the way for advancements in this challenging area of machine learning.

We summarize relevant characteristics for a representative AMTS dataset as:

- **Missing data:** Gaps can happen sporadically due to sensor malfunctions or data transmission errors, leading to incomplete time series.
- **Irregular Sampling:** Frequency of data collection varies, either across different time series or within a single series.
- **Time offsets:** As shown in Figure 1 (left), misaligned time series due to time offsets rapidly become sparse in respect to the underlying grid.
- **Known covariates:** Also known as exogenous regressors. These are variables that are known for $t > t_\phi$ and influence the behaviour of target variables.

The PSOD Dataset

The PSOD dataset is a collection of metocean data related to the Port of Santos main channel. The Port of Santos, located in the state of São Paulo, Brazil, is the largest and most significant port in Latin America. Serving as a crucial gateway for Brazil's international trade, it handles a substantial volume of the country's imports and exports. The port's strategic position, near the major industrial and agricultural centers, makes it a vital hub for both the national and global economy. Over the years, the Port of Santos has undergone significant expansions and modernization efforts to accommodate the growing demand and to improve its operational efficiency, solidifying its role as a key player in global maritime logistics.

The dataset comprises four time series obtained from sensors installed in the port's main channel and two additional time series derived from numerical models. The channel and sensor locations are depicted in Figure 2, and the dataset's content is summarized as follows:

- **Water Current:** Captures the speed of the water current in the main channel at the Praticagem measuring station. This relatively stable and less noisy variable is crucial for effective port operations planning.
- **Waves:** Includes significant wave height, peak period, and speed, recorded at the Palmas measuring station.

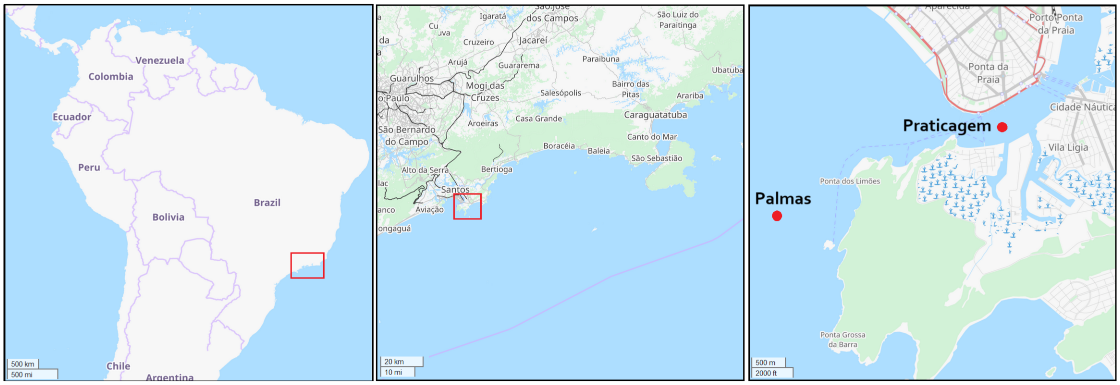

Figure 2: Channel location with the two sensor locations (Praticagem and Palmas stations) highlighted.

Model	Pre-Trained	Uses Known Covariates	Time Encoding	Multivariate
Chronos	✓	✗	✗	✗
CGNN	✗	✗	✗	✓
CGNN + Time Encoding	✗	✗	✓	✓
Gap-Ahead	✗	✓	✓	✓
GRU	✗	✗	✗	✗
GRU+GNN	✗	✗	✗	✓
GRU + Time Encoding	✗	✗	✓	✗
LiESN-D	✗	✗	✓	✓
MoNODE	✗	✗	✗	✗
N-HITS	✗	✗	✓	✓
Prophet	✓	✗	✓	✗

Table 2: List of selected models with their characteristics relevant to the AMTS forecast task.

Wave speed is noisier than water current, while peak period and wave height are particularly challenging due to their high noise levels. Accurate forecasting of these features is essential for port efficiency and safety.

- **Winds:** Collected from the Praticagem station, winds are vital in wave formation and water currents, influencing the overall port dynamics.
- **Sea Surface Height:** Exogenous variables available for $t > t_\phi$, significantly impacting target variable behavior, crucial for precise forecasting.
- **SOFS:** The port's production numerical model, providing forecasts for current speed and sea surface height, serves as a key covariate in the PSOD dataset.
- **Astronomical Tide:** A highly predictable covariate caused by the Sun and Moon's movements, important for accurate forecasting.

It is crucial to emphasize that modeling known time-evolving covariates, such as SOFS and Astronomical Tide, presents a significant challenge that, to the best of our knowledge, has not been addressed by any existing AMTS dataset. This is a critical oversight, as in many applied research scenarios involving time series forecasting, a baseline model is often available and can be leveraged to inform predictions.

The PSOD Forecast Task. The AMTS forecast task associated with this dataset, as defined in Definition 2.2, is to predict *water current* (1 feature) and *wave* characteristics (3 features) over the next two days at specific instants $\mathcal{T}_{\mathcal{S}_f}$, using information from all six time series. For the test portion of the dataset, we introduce five data loss scenarios with missing data ratios of 0%, 20%, 40%, 60%, and 80%. Several studies have used data removal from regular time series as a method to generate AMTS (Che et al. 2018; Kidger et al. 2020; Chen et al. 2023). While missing data is not the only type of irregularity that is relevant, we incorporate this mechanism into our dataset to simulate extreme data loss scenarios, thereby offering the flexibility required to evaluate models' robustness to such conditions.

Each data loss scenario consists of 356 pairs (X, Y), where $X = (\mathcal{M}_c, \mathcal{T}_{\mathcal{S}_f})$ and $Y = \mathcal{S}_f$. To assess the performance, we use the Index of Agreement (IoA) (Willmott 1981), a standardized metric that compares two sequences $\hat{\mathbf{Z}}^{\mathbf{i}}$ and $\mathbf{Z}^{\mathbf{i}}$, expressed as

$$\mathrm{IoA}(\hat{\mathbf{Z}}^{\mathbf{i}}, \mathbf{Z}^{\mathbf{i}}) = 1 - \frac{\sum_t (\mathbf{Z}^{\mathbf{i}} - \hat{\mathbf{Z}}^{\mathbf{i}})^2}{\sum_t (|\hat{\mathbf{Z}}^{\mathbf{i}} - \bar{\mathbf{Z}}^{\mathbf{i}}| + |\mathbf{Z}^{\mathbf{i}} - \bar{\mathbf{Z}}^{\mathbf{i}}|)^2}. \tag{7}$$

The IoA provides a robust measure of model prediction accuracy, ranging from 0 (no agreement) to 1 (perfect match), and is extensively used in hydrological modeling to evaluate models by accounting for both systematic biases and data variability. In addition to IoA, we also report results using the Mean Absolute Error (MAE), offering a com-

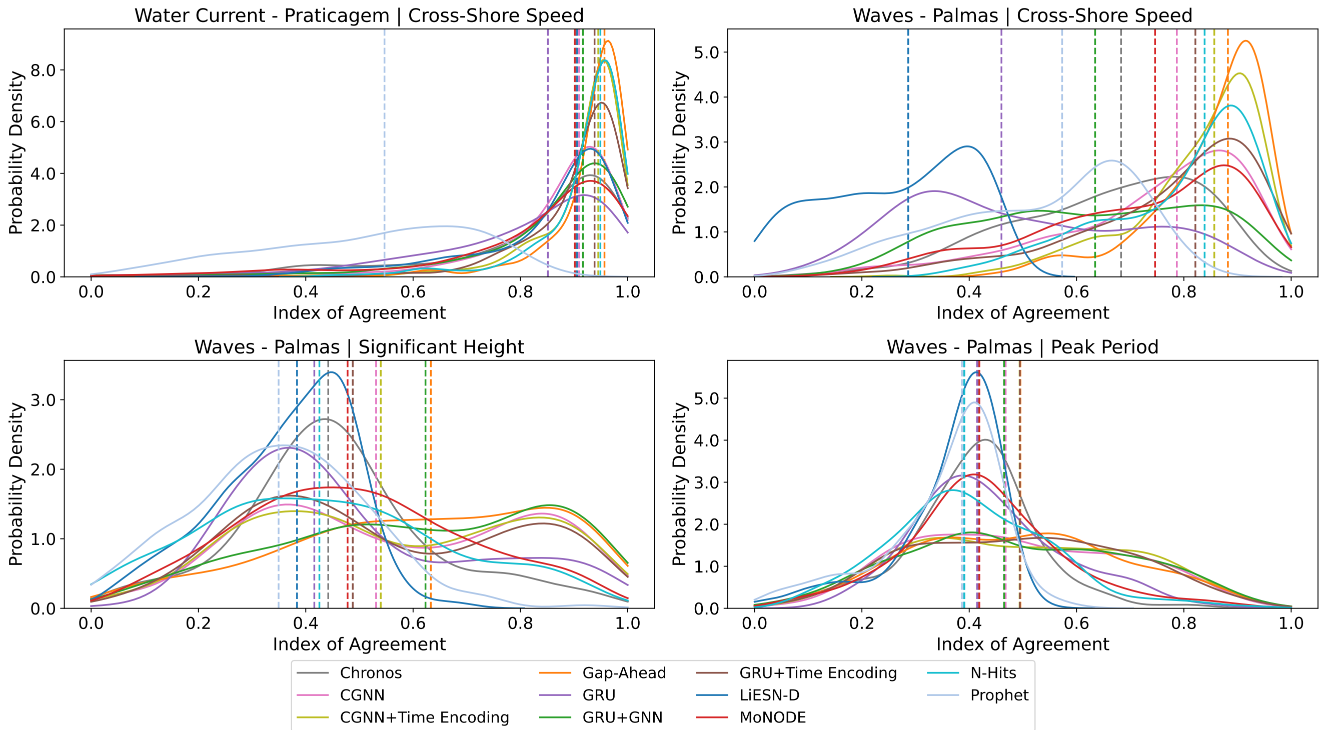

Figure 3: Kernel density estimation plot of the Index of Agreement (IoA) comparing model performance across all four features with 0% data removal. Dotted lines indicate the median values. Models that incorporate both time encoding and information sharing between sensors consistently demonstrate superior performance.

prehensive evaluation of model performance across varying scenarios.

4 Benchmarking: Insights from PSOD

We selected eleven models from the literature to tackle the PSOD forecasting task, with their characteristics detailed in Table 2. These models were chosen for their significance in the field and as representatives of various machine learning domains that have advanced time series forecasting. By analyzing how these different approaches perform under the unique challenges of PSOD, we gained important insights into the key factors to consider when designing a model capable of handling AMTS.

- **Chronos (Ansari et al. 2024):** A foundation model for time series that tokenizes data to utilize language models and extends mix-up data augmentation for time series, aimed at handling the data-intensive nature of these models. We use Chronos with the T5 (large) model, comprising 710M parameters.
- **CGNN (Xhonneux, Qu, and Tang 2020):** Continuous Graph Neural Networks use ODEs to model continuous dynamics in graphs, capturing long-range dependencies and mitigating over-smoothing issues in discrete GNNs.
- **Gap-Ahead (Barros et al. 2024):** A model featuring timestamp encoding, independent RNN encoding for each time series, and information diffusion through a Regularized Heterogeneous Graph Attention Network (RHGAT).
- **Auto-Regressive RNN (GRU) (Cho et al. 2014):** A Gated Recurrent Unit network (GRU) used as a baseline in the benchmark.
- **LiESN-D (Lukoševičius 2012):** A Leaky-Integrator Echo State Network using differential formulation, where neurons' time-continuous updates are discretized via the forward Euler method.
- **MoNODE (Auzina et al. 2023):** Modulated Neural ODEs introduce time-invariant modulator variables to separate dynamic states from static factors, enhancing generalization across dynamic parameterizations and improving long-term forecasting.
- **N-HITS (Challu et al. 2023):** A forecasting model that improves long-horizon predictions by combining multi-rate sampling with hierarchical interpolation, efficiently capturing different signal frequencies and scales.
- **Prophet (Taylor and Letham 2017):** Foundation time series model designed for business data, accommodating trend changes, seasonalities, and holidays, with intuitive parameters for easy adjustment by non-experts.
- **RNN+GNN:** A baseline combining RNNs and GNNs for feature propagation in multivariate time series, using a simplified version of the setup from Barros et al. (2024).

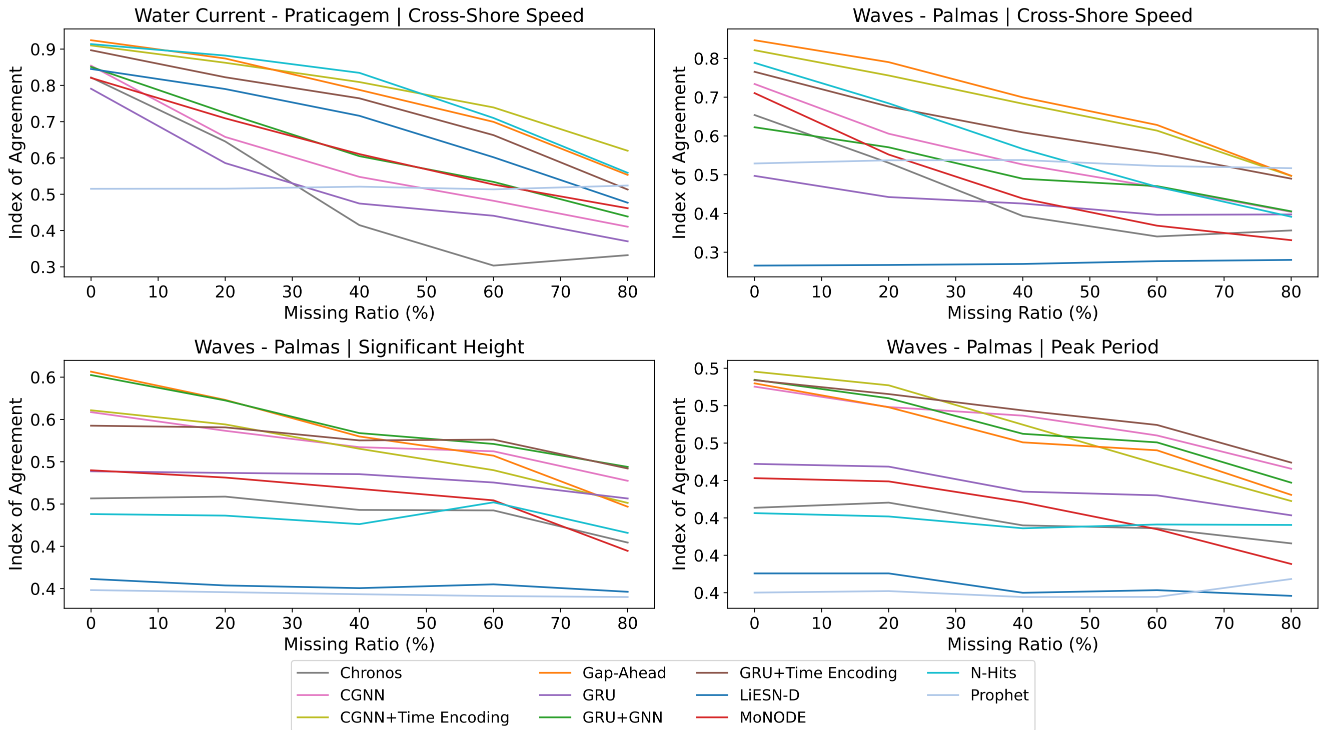

Figure 4: Index of agreement degradation as data removal increases. Approaches without time encoding solutions suffer from rapid performance degradation as time series gaps become prevalent.

Other architectures that focus on similar problems, such as AMTS classification, could be, at least in principle, adapted to solve this task. However, one of our goals is to instigate the development of dedicated AMTS forecasting solutions that do not rely on ad-hoc hybrid architectures and can easily be applied to replace some data preprocessing technique.

Figure 3 presents the IoA distributions and medians estimated from the test dataset for all models and features under the 0% data removal condition. A key finding is that multivariate approaches with time encoding techniques, such as Gap-Ahead and CGNN+Time Encoder, significantly outperform others. For wave speed, a moderately noisy variable that is highly correlated with water current speed, the results show a clear spread of medians, with approaches lacking time encoding performing worse. This trend is consistent across all features, except for LiESN-D that performed poorly on noisy features.

Interestingly, the outstanding performance typically seen with foundation models on regular time series did not materialize in this experiment, though different hyperparameters might yield better results. Figure 4 illustrates the degradation of IoA as the missing data ratio increases. The curve reveals that Chronos experiences a sharp decline in accuracy, alongside other approaches without time encoding. Notably, as the missing data ratio grows, the advantage of time-informed models becomes even more pronounced.

Furthermore, the significant improvement seen when simply adding time encoding to both GRU and CGNN underscores the importance of time encoding. This addition not only boosted the base IoA but also enhanced robustness to missing data, highlighting time encoding as a critical component regardless of the underlying architecture.

5 Conclusion

In this work, we addressed the significant gap in machine learning research concerning Asynchronous Multivariate Time Series (AMTS) by proposing a formal definition and structured approach to this challenge. We provided a unified framework for defining AMTS and introduced the Port of Santos Oceanographic Data (PSOD) dataset, designed to evaluate model robustness and processing capabilities in handling AMTS. Additionally, our benchmarking of various models revealed the critical role of time encoding in enhancing accuracy and robustness to missing data.

Our results clearly show that models specifically designed to handle irregularities in time series data outperform those relying on regularity assumptions. This underscores the necessity for dedicated AMTS forecasting solutions that move beyond extensive preprocessing and offer more reliable, efficient alternatives.

As AMTS becomes increasingly prevalent in real-world applications, we hope this work will inspire further advancements in machine learning, paving the way for more effective and robust models across a wide range of domains.

References

Ansari, A. F.; Stella, L.; Turkmen, C.; Zhang, X.; Mercado, P.; Shen, H.; Shchur, O.; Rangapuram, S. S.; Arango, S. P.; Kapoor, S.; Zschiegner, J.; Maddix, D. C.; Wang, H.; Mahoney, M. W.; Torkkola, K.; Wilson, A. G.; Bohlke-Schneider, M.; and Wang, Y. 2024. Chronos: Learning the Language of Time Series. ArXiv:2403.07815 [cs].

Auzina, I. A.; Yildiz, C.; Magliacane, S.; Bethge, M.; and Gavves, E. 2023. Modulated Neural ODEs. In *Thirty-seventh Conference on Neural Information Processing Systems*.

Barros, M.; Pinto, A.; Monroy, A.; Moreno, F.; Coelho, J.; Silva, A. P.; Netto, C. F. D.; Leite, J. R.; Mathias, M.; Tannuri, E.; Jordao, A.; Gomi, E.; Cozman, F.; Dottori, M.; and Costa, A. H. R. 2024. Early Detection of Extreme Storm Tide Events Using Multimodal Data Processing. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(20): 21923–21931. Number: 20.

Challu, C.; Olivares, K. G.; Oreshkin, B. N.; Ramirez, F. G.; Canseco, M. M.; and Dubrawski, A. 2023. NHITS: Neural Hierarchical Interpolation for Time Series Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(6): 6989–6997. Number: 6.

Che, Z.; Purushotham, S.; Cho, K.; Sontag, D.; and Liu, Y. 2018. Recurrent Neural Networks for Multivariate Time Series with Missing Values. *Scientific Reports*, 8(1): 6085. Number: 1 Publisher: Nature Publishing Group.

Chen, R. T. Q.; Rubanova, Y.; Bettencourt, J.; and Duvenaud, D. K. 2018. Neural Ordinary Differential Equations. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc.

Chen, Y.; Ren, K.; Wang, Y.; Fang, Y.; Sun, W.; and Li, D. 2023. ContiFormer: Continuous-Time Transformer for Irregular Time Series Modeling. *Advances in Neural Information Processing Systems*, 36: 47143–47175.

Cho, K.; van Merriënboer, B.; Bahdanau, D.; and Bengio, Y. 2014. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches. In Wu, D.; Carpuat, M.; Carreras, X.; and Vecchi, E. M., eds., *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, 103–111. Doha, Qatar: Association for Computational Linguistics.

Cooke, M.; Barker, J.; Cunningham, S.; and Shao, X. 2006. An audio-visual corpus for speech perception and automatic speech recognition. *The Journal of the Acoustical Society of America*, 120(5): 2421–2424.

Emmanuel, T.; Maupong, T.; Mpoeleng, D.; Semong, T.; Mphago, B.; and Tabona, O. 2021. A survey on missing data in machine learning. *Journal of Big Data*, 8(1): 140.

Funahashi, K.-i.; and Nakamura, Y. 1993. Approximation of dynamical systems by continuous time recurrent neural networks. *Neural Networks*, 6(6): 801–806.

Hebrail, G.; and Berard, A. 2006. Individual Household Electric Power Consumption, UCI Machine Learning Repository.

Horn, M.; Moor, M.; Bock, C.; Rieck, B.; and Borgwardt, K. 2020. Set Functions for Time Series. In *International Conference on Machine Learning*, 4353–4363. PMLR. ISSN: 2640-3498.

Huang, S.; Poursafaei, F.; Danovitch, J.; Fey, M.; Hu, W.; Rossi, E.; Leskovec, J.; Bronstein, M.; Rabusseau, G.; and Rabbany, R. 2023. Temporal Graph Benchmark for Machine Learning on Temporal Graphs. *Advances in Neural Information Processing Systems*, 36: 2056–2073.

Huang, Z.; Sun, Y.; and Wang, W. 2020. Learning Continuous System Dynamics from Irregularly-Sampled Partial Observations. In *Advances in Neural Information Processing Systems*, volume 33, 16177–16187. Curran Associates, Inc.

Johnson, A. E. W.; Pollard, T. J.; Shen, L.; Lehman, L.-w. H.; Feng, M.; Ghassemi, M.; Moody, B.; Szolovits, P.; Anthony Celi, L.; and Mark, R. G. 2016. MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3(1): 160035. Publisher: Nature Publishing Group.

Karagiorgi, G.; Kasieczka, G.; Kravitz, S.; Nachman, B.; and Shih, D. 2022. Machine learning in the search for new fundamental physics. *Nat Rev Phys*, 4(6): 399–412. Publisher: Nature Publishing Group.

Kazemi, S. M.; Goel, R.; Jain, K.; Kobyzev, I.; Sethi, A.; Forsyth, P.; and Poupart, P. 2020. Representation Learning for Dynamic Graphs: A Survey. *Journal of Machine Learning Research*, 21(70): 1–73.

Kidger, P.; Morrill, J.; Foster, J.; and Lyons, T. 2020. Neural controlled differential equations for irregular time series. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, 6696–6707. Curran Associates Inc. ISBN 978-1-71382-954-6.

Little, R. J. A.; and Rubin, D. B. 2002. Missing Data in Experiments. In *Statistical Analysis with Missing Data*, 24–40. John Wiley & Sons, Ltd. ISBN 978-1-119-01356-3.

Lukoševičius, M. 2012. *A Practical Guide to Applying Echo State Networks*, 659–686. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN 978-3-642-35289-8.

Mei, H.; and Eisner, J. 2017. The neural hawkes process: a neurally self-modulating multivariate point process. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, 6757–6767. Curran Associates Inc. ISBN 978-1-5108-6096-4.

Menne, M. J.; Williams, J.; and Vose, R. S. 2016. Long-Term Daily and Monthly Climate Records from Stations Across the Contiguous United States (U.S. Historical Climatology Network). Technical Report osti:1394920; cdiac:NDP-019; doi:10.3334/CDIAC/CLI.NDP019, Environmental System Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE) (United States); CDIAC.

Neil, D.; Pfeiffer, M.; and Liu, S.-C. 2016. Phased LSTM: Accelerating Recurrent Network Training for Long or Event-based Sequences. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc.

Norcliffe, A. L. I.; Bodnar, C.; Day, B.; Moss, J.; and Lio, P. 2021. Neural ODE Processes: A Short Summary. In *The Symbiosis of Deep Learning and Differential Equations*.

Reiss, A.; and Stricker, D. 2012. Introducing a New Benchmarked Dataset for Activity Monitoring. In *2012 16th Annual International Symposium on Wearable Computers (ISWC)*, 108–109. IEEE. ISBN 978-0-7695-4697-1 978-1-4673-1583-8.

Reyes-Ortiz, J.; and Anguita, D. 2013. Human Activity Recognition Using Smartphones.

Reyna, M. A.; Josef, C. S.; Jeter, R.; Shashikumar, S. P.; Westover, M. B.; Nemati, S.; Clifford, G. D.; and Sharma, A. 2020. Early Prediction of Sepsis From Clinical Data: The PhysioNet/Computing in Cardiology Challenge 2019. *Critical Care Medicine*, 48(2): 210–217.

Shukla, S. N.; and Marlin, B. 2020. Multi-Time Attention Networks for Irregularly Sampled Time Series. In *International Conference on Learning Representations*.

Silva, I.; Moody, G.; Scott, D. J.; Celi, L. A.; and Mark, R. G. 2012. Predicting In-Hospital Mortality of ICU Patients: The PhysioNet/Computing in Cardiology Challenge 2012. *Computing in cardiology*, 39: 245–248.

Taylor, S. J.; and Letham, B. 2017. Forecasting at scale. Technical Report e3190v2, PeerJ Inc. ISSN: 2167-9843.

Vito, S. 2008. Air Quality, UCI Machine Learning Repository.

Wei, W. W. S. 2019. Multivariate Time Series Analysis and Applications. *Wiley*, 528.

Willmott, C. J. 1981. On the Validation of Models. *Physical Geography*, 2(2): 184–194.

Wu, R.; Hamshaw, S. D.; Yang, L.; Kincaid, D. W.; Etheridge, R.; and Ghasemkhani, A. 2022. Data Imputation for Multivariate Time Series Sensor Data With Large Gaps of Missing Data. *IEEE Sensors Journal*, 22(11): 10671–10683.

Xhonneux, L.-P. A. C.; Qu, M.; and Tang, J. 2020. Continuous graph neural networks. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *ICML'20*, 10432–10441. JMLR.org.

Xiao, J.; Basso, L.; Nejdl, W.; Ganguly, N.; and Sikdar, S. 2024. IVP-VAE: Modeling EHR Time Series with Initial Value Problem Solvers. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(14): 16023–16031. Number: 14.

Yalavarthi, V. K.; Madhusudhanan, K.; Scholz, R.; Ahmed, N.; Burchert, J.; Jawed, S.; Born, S.; and Schmidt-Thieme, L. 2024. GraFITi: Graphs for Forecasting Irregularly Sampled Time Series. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(15): 16255–16263. Number: 15.

Zhang, S.; Guo, B.; Dong, A.; He, J.; Xu, Z.; and Chen, S. X. 2017. Cautionary tales on air-quality improvement in Beijing. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2205): 20170457. Publisher: Royal Society.

Zhang, W.; Zhang, L.; Han, J.; Liu, H.; Zhou, J.; Mei, Y.; and Xiong, H. 2023. Irregular Traffic Time Series Forecasting Based on Asynchronous Spatio-Temporal Graph Convolutional Network. *arXiv*, (arXiv:2308.16818).

Zhang, X.; Zeman, M.; Tsiligkaridis, T.; and Zitnik, M. 2021. Graph-Guided Network for Irregularly Sampled Multivariate Time Series. In *International Conference on Learning Representations*.

Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; and Zhang, W. 2021. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(12): 11106–11115. Number: 12.

Zhu, A.; Jin, P.; Zhu, B.; and Tang, Y. 2022. On Numerical Integration in Neural Ordinary Differential Equations. In *International Conference on Machine Learning*, 27527–27547. PMLR. ISSN: 2640-3498.

6 Reproducibility Checklist

This paper:

Includes a conceptual outline and/or pseudocode description of AI methods introduced (yes/partial/no/NA): Yes

Clearly delineates statements that are opinions, hypothesis, and speculation from objective facts and results (yes/no): Yes

Provides well marked pedagogical references for less-familiare readers to gain background necessary to replicate the paper (yes/no): Yes

Does this paper make theoretical contributions? (yes/no): Yes

If yes, please complete the list below.

All assumptions and restrictions are stated clearly and formally. (yes/partial/no): Yes

All novel claims are stated formally (e.g., in theorem statements). (yes/partial/no): Yes

Proofs of all novel claims are included. (yes/partial/no): Yes

Proof sketches or intuitions are given for complex and/or novel results. (yes/partial/no): Yes

Appropriate citations to theoretical tools used are given. (yes/partial/no): Yes

All theoretical claims are demonstrated empirically to hold. (yes/partial/no/NA): Yes

All experimental code used to eliminate or disprove claims is included. (yes/no/NA): Yes

Does this paper rely on one or more datasets? (yes/no): Yes

If yes, please complete the list below.

A motivation is given for why the experiments are conducted on the selected datasets (yes/partial/no/NA): Yes

All novel datasets introduced in this paper are included in a data appendix. (yes/partial/no/NA): Yes

All novel datasets introduced in this paper will be made publicly available upon publication of the paper with a license that allows free usage for research purposes. (yes/partial/no/NA): Yes

All datasets drawn from the existing literature (potentially including authors' own previously published work) are accompanied by appropriate citations. (yes/no/NA): Yes

All datasets drawn from the existing literature (potentially including authors' own previously published work) are publicly available. (yes/partial/no/NA): Yes

All datasets that are not publicly available are described in detail, with explanation why publicly available alternatives are not scientifically satisficing. (yes/partial/no/NA): NA

Does this paper include computational experiments? (yes/no)

If yes, please complete the list below.

Any code required for pre-processing data is included in the appendix. (yes/partial/no): Yes

All source code required for conducting and analyzing the experiments is included in a code appendix. (yes/partial/no): Yes

All source code required for conducting and analyzing the experiments will be made publicly available upon publication of the paper with a license that allows free usage for research purposes. (yes/partial/no): Yes

All source code implementing new methods have comments detailing the implementation, with references to the paper where each step comes from (yes/partial/no): Yes

If an algorithm depends on randomness, then the method used for setting seeds is described in a way sufficient to allow replication of results. (yes/partial/no/NA): Yes

This paper specifies the computing infrastructure used for running experiments (hardware and software), including GPU/CPU models; amount of memory; operating system; names and versions of relevant software libraries and frameworks. (yes/partial/no): Yes

This paper formally describes evaluation metrics used and explains the motivation for choosing these metrics. (yes/partial/no): Yes

This paper states the number of algorithm runs used to compute each reported result. (yes/no): Yes

Analysis of experiments goes beyond single-dimensional summaries of performance (e.g., average; median) to include measures of variation, confidence, or other distributional information. (yes/no): Yes

The significance of any improvement or decrease in performance is judged using appropriate statistical tests (e.g., Wilcoxon signed-rank). (yes/partial/no): Yes

This paper lists all final (hyper-)parameters used for each model/algorithm in the paper's experiments. (yes/partial/no/NA): Yes

This paper states the number and range of values tried per (hyper-) parameter during development of the paper, along with the criterion used for selecting the final parameter setting. (yes/partial/no/NA): Yes