An Adversarial Learning Approach to Irregular Time-Series Forecasting

Heejeong Nam*
hatbi2000@yonsei.ac.kr

Jihyun Kim
20160131@dongduk.ac.kr

Jimin Yeom
jyeom32@gatech.edu

Abstract

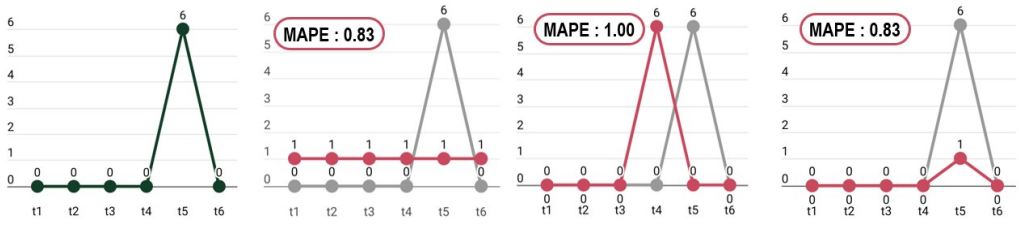
Forecasting irregular time series presents significant challenges due to two key issues: the vulnerability of models to mean regression, driven by the noisy and complex nature of the data, and the limitations of traditional error-based evaluation metrics, which fail to capture meaningful patterns and penalize unrealistic forecasts. These problems result in forecasts which are often misaligned with human intuition. To tackle these challenges, we propose an adversarial learning framework with a deep analysis of adversarial components. Specifically, we emphasize the importance of balancing the modeling of global distribution *(overall patterns)* and transition dynamics *(localized temporal changes)* to better capture the nuances of irregular time series. Overall, this research provides practical insights for improving models and evaluation metrics, and pioneers the application of adversarial learning in the domain of irregular time-series forecasting.

1 Introduction

Irregular time series, characterized by significant variations in inter-arrival times and quantities, pose unique challenges in analysis and forecasting. Unlike stationary and regular time series, which have seen substantial advancements in both methodologies and applications [1], the exploration of irregular time series remains limited due to their inherent low interpretability. Forecasting such data is particularly challenging for two primary reasons—one stemming from the inadequacies of evaluation metrics and the other from the limitations of models, with both factors compounding each other. The first challenge lies in the widespread reliance on error-based metrics in forecasting, such as MAPE (mean absolute percentage error), which are ill-suited for capturing the unique characteristics of lumpy or intermittent patterns [2, 3]. These metrics fail to penalize the unrealistic forecasts often produced by statistical models, while simultaneously overlooking the potential of models that accurately identify underlying patterns but exhibit minor temporal shifts. The second challenge stems from the mean regression problem faced by forecasting models. This issue primarily arises due to the inherently noisy nature of irregular time series, which often lack clear trends or seasonality, making them especially prone to this problem. Paradoxically, under commonly used metrics like MAPE, the mean regression problem is not penalized but often rewarded, leading to models that fail to align with human intuition. Fig. 1 provides a clear example where statistical models or models with mean regression problem deviate notably from intuitive expectations, despite achieving high scores with MAPE-based evaluations. Similar cases in real world datasets are detailed in Appendix A. While fields such as large language models and image generation have made significant strides in producing outputs that align with human intuition, traditional forecasting models and evaluation metrics lag behind, especially in irregular time-series. To address this issue, we propose an adversarial learning framework aimed at bridging the gap between forecasting outputs and intuitive expectations. Adversarial learning, widely recognized for its applications in time-series

*First author and corresponding author

AdvML-Frontiers'24: The 3nd Workshop on New Frontiers in Adversarial Machine Learning@NeurIPS'24, Vancouver, CA.



(a) True (b) Prediction 1 (c) Prediction 2 (d) Prediction 3

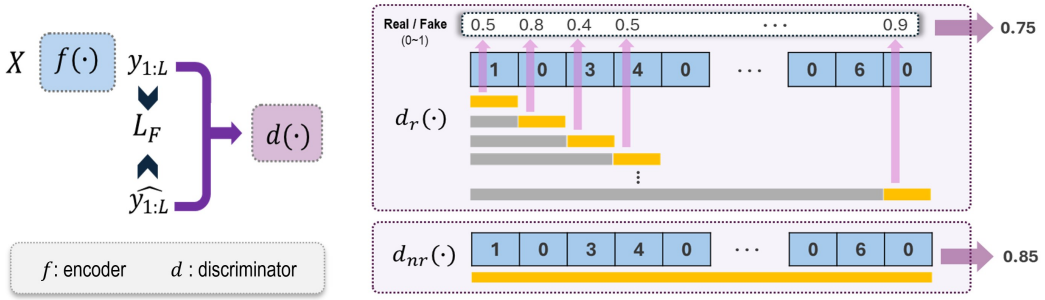
Figure 1: **Problems.** MAPE, where lower values indicate better performance, is highly unsuitable for evaluating the quality of irregular time-series forecasting. The true sequence (a) exhibits an irregular pattern. (b) matches the total quantity but reflects the monotonous tendencies common in many statistical models. (c), despite experiencing a slight temporal shift, receives the worst score. (d) provides a plausible result, capturing the correct entry point, yet it achieves the same MAPE as (b).

generation and forecasting, is predominantly inspired by generative adversarial networks (GANs) [4]. While GANs have demonstrated strong capabilities in generating realistic and detailed outputs, their application to time-series forecasting remains underexplored particularly in the context of irregular data. In the domain of time-series forecasting, discriminators typically fall into one of two architectural paradigms: recursive layers, which align with conventional methods for sequential data processing [5, 6, 7], and non-recursive layers integrated with activation functions [8, 9]. Despite their widespread use, a comprehensive analysis of the role and effectiveness of these architectures in time-series forecasting remains limited. To address this, we reexamine adversarial components and their impact on forecasting performance for irregular time-series data. To further tackle the limitations of traditional evaluation metrics which often neglect forecast plausibility, we propose novel qualitative approaches for analyzing irregular time-series forecasting. Our findings indicate that the architectural design of the adversarial components should align with the characteristics of the time series, balancing the capture of **global style** and **transition dynamics**. Finally, our research provides actionable insights into improving the models and the evaluation metrics for irregular time-series data. To our best knowledge, this is the first use of adversarial approaches in irregular time-series forecasting. We applied our approach to three real-world datasets, and our implementation is available at Hazel-Heejeong-Nam/adversarial-intermittent-lumpy-forecasting

2 Related Works

Adversarial learning in time series forecasting. Adversarial learning is often adopted in time series generation and forecasting [7, 8, 6, 9, 10, 5, 11]. However, despite its widespread use, a thorough analysis of the role of adversarial learning in the time-series domain remains lacking. The unidirectional nature of sequential data introduces unique challenges, particularly in forecasting rather than in generation. Hence, it is crucial to revisit the objective of using adversarial components in forecasting, given the differing nature of these tasks. The earlier works including C-RNN-GAN [5] simply replaced the generator and discriminator with LSTM networks. On the other hand, RCGAN [12] conditioned on additional input instead of previous outputs, while it still utilized recurrent units for both the generator and discriminator. TimeGAN [6] employed both supervised and adversarial objectives to address the mean regression problem and the lack of temporal dynamic consideration, respectively. Alongside other works [13, 14], TimeGAN integrated recurrent networks into both the generator and discriminator. GT-GAN [7], while not focused on forecasting, employs GRU-ODE [15] to construct the discriminator for generative purposes. AST [8] is the first adversarial model designed for forecasting, and it incorporates adversarial learning to prevent error accumulation from the autoregressive nature of predictions. Thus, the discriminator module consists of non-recursive layers and competes with the forecaster, which is expected to eliminate error accumulation. TrendGCN [9] employs a Graph Convolutional Network as a forecaster and utilized two discriminators, one for capturing spatial relations and the other for temporal relations.

Irregular time series forecasting. Irregular and sparse time-series forecasting is challenging as it is often characterized by multiple inter-arrival times, with some additionally distinguished by significant variations in quantity between these intervals. Although statistical methods (e.g., Croston [16]), machine learning approaches (e.g., SVR [17]), and deep learning techniques (e.g., LSTM [18]) have been applied to these problems [3], it remains unclear which method is the most suitable.



(a) (b)

Figure 2: **Overall frameworks.** (a) Simple encoder-discriminator framework of adversarial learning. (b) Concept of $d_r(\cdot)$ (recursive discriminator) and $d_{nr}(\cdot)$ (non-recursive discriminator).

This ambiguity arises not only due to insufficient investigation [1] but also because of the lack of appropriate metrics [2, 3]. Statistical methods like Croston [16], Holt-Winters [19], and ARIMA [20] have shown strong performance in M4 forecasting competitions [21]. Among these, Croston remains particularly effective for forecasting intermittent and lumpy data. In machine learning, SVR [22] has been identified as a strong performer; for example, Hua et al. [17] combined SVR with logistic regression. In deep learning, LSTM [18] models have shown promise, although their performance can vary. While neural networks may underperform compared to statistical methods on metrics like RMSE (root mean squared error) and MAPE, they may outperform in terms of service-level quality. For example, MAPE and RMSE have been criticized for their inability to account for shifts and temporal interactions [3], which are strengths of neural networks.

3 Rethinking Adversarial Components in Irregular Time-Series Forecasting

3.1 Problem Formulation

Let's assume we have $\mathbf{M}$ distinct time series, each spanning a time period indexed by $t = 1, \ldots, T$. To forecast, models utilize historical data of length $\mathbf{P}$, represented by $\{x_{im}\}_{m=1}^{P}$ for each i^{th} time series. The forecasting horizon is set to $\mathbf{L}$, and our objective is to accurately predict the values $\{y_{in}\}_{n=1}^{L}$. Given a forecasting model $\mathbf{F} : \mathbb{R}^P \to \mathbb{R}^L$, we express our forecasting process as $\{\hat{y_{in}}\}_{n=1}^{L} = \mathbf{F}(\{x_{im}\}_{m=1}^{P})$. In this paper, we only considered a global forecasting model across the different time series $i = 1, \ldots, \mathbf{M}$ within the same dataset, i.e., the observational space is $\mathcal{X} \in \mathbb{R}^{M \times T}$ while $\mathbb{X} = \{x_{i,1:P}\}_{i=1}^{M}$ and $\mathbb{Y} = \{y_{i,1:L}\}_{i=1}^{M}$ represent the sets of historical data and corresponding target values, respectively. Below, we specify the meaning of irregular time-series in our work.

Definition 3.1 ***(Irregular Time-Series)*** *We define irregular time series by their characteristic variability in inter-arrival times and further subcategorize them based on variability in quantity, drawing on the work of Syntetos et al. [23]. Irregular time series are identified by using the* ***ADI*** *(Average Demand Interval), a metric that quantifies regularity over time by calculating the average interval between successive non-zero entries. Following established conventions [3, 23, 24], we adopt a threshold of 1.32 for ADI ($ADI \geq 1.32$). Irregular time series can be further classified into two subcategories—**intermittent** and **lumpy**—based on the CV^2 metric, which measures the variability in non-zero quantities. We use a threshold of 0.49 for CV^2, as suggested in previous studies [3, 23, 24].*

3.2 Adversarial Components

Non-recursive Discriminator Discriminators with non-recursive layers treat irregular time-series as vectors without considering temporal relationships, similar to the approach used in image transfer or generation [25, 26]. For a fixed forecaster $\boldsymbol{F}$, the non-recursive discriminator $D_{nr}(\mathbf{y}, \theta_d)$ outputs a scalar which indicates the probability of input vector $\mathbf{y}$ originating from $\mathcal{X}$. We expect D_{nr} to operate as discribed in (1), and the global minimum is achieved if and only when $p_F = p_{\mathcal{X}}$.

$$D_{nr}(y_{1:L}) = \frac{p_{\mathcal{X}}(y_{1:L})}{p_{\mathcal{X}}(y_{1:L}) + p_F(y_{1:L})} \tag{1}$$

Recurvise Discriminator A discriminator with recursive layers D_r functions similarly to a chain of multiple PCL modules [27]. The PCL framework constructs two samples of vectors as shown in

Table 1: Summary of the key statistics for each dataset. For calculating CV^2 of demand size, the number of lumpy time-series and intermittent time-series, we follow SBC [23]

	AUTO	**M5**	**RAF**
Number of Time-Series (**M**)	1227	30489	5000
Time Series Length (**T**)	24	182	84
Mean Demand Interval	1.29	4.212	10.02
CV^2	4.38	2.91	10.67
Lumpy	286	5964	2403
Intermittent	941	23040	2597

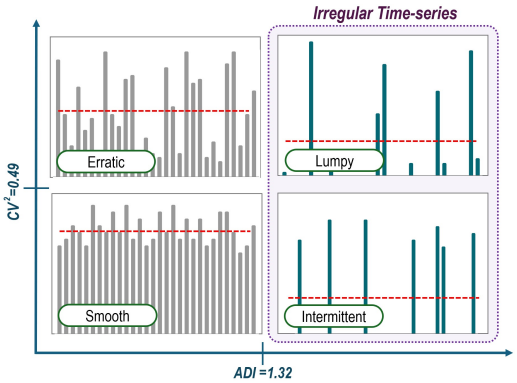


Figure 3: Illustration of irregular time-series

(2). Here, $v(n)$ intuitively provides a minimal description of the temporal dependencies in the data. For comparison, a fake data (in the context of adversarial learning) is sampled (*forecasted* in our case), and the objective is to learn to discriminate between $v(n)$ and $v^*(n)$ using logistic regression as expressed in (3), where h_n are scalar-valued functions that provide representation.

$$v(n) = \begin{pmatrix} y_{n-1} \\ y_n \end{pmatrix} \quad \text{and} \quad v^*(n) = \begin{pmatrix} y_{n-1} \\ y_n^* \end{pmatrix} \tag{2}$$

$$D_r(v) = \frac{1}{L} \sum_{n=1}^{L} B_n(h_n(v^1), h_n(v^2)) \tag{3}$$

It has been theoretically proven [27] that if the underlying sources of the time-series are 1) mutually independent, 2) temporally dependent, and 3) stationary, then $h_n(y_{n-1})$ can recover the underlying sources up to permutation and component-wise transformation. However, in our Def.3.1, it is challenging to assume stationary sources : for example, one might intuitively model one of the nonstationary sources as active and inactive phases, as proposed by Song et al .[28], given that our dataset is characterized by varying inter-arrival times. Nevertheless, incorporating D_r could significantly enhance the model's ability to capture and adapt to the underlying transition dynamics, providing a more robust framework for modeling the temporal shifts in the data.

4 An adversarial approach to irreuglar time-series forecasting

4.1 Dataset

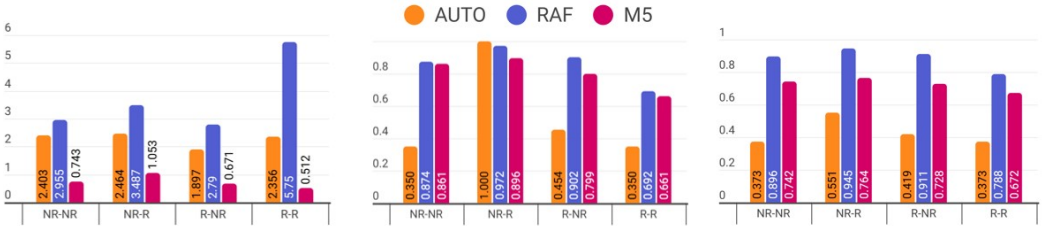
All experiments in this paper were evaluated on three real-world datasets. Following Def.3.1, we only include time-series classified as irregular according to the Syntetos-Boylan Classification [23]. **AUTO** dataset includes monthly demand data for 3,000 items over 24 months. Among these, we are able to get 1,227 irregular time-series. **RAF** dataset comprises aerospace parts demand data from the Royal Air Force and it covers 84 months of data for 5,000 parts. Lastly, **M5** dataset includes daily unit sales per product and store at Walmart over 5 years and we used last 182 timestamps for our experiment. When comparing these datasets, the RAF dataset has the highest ADI and CV^2 values, making it the most challenging for traditional forecasting methods. The AUTO dataset, while having fewer zero entries, exhibits greater variation in demand size. In contrast, the M5 dataset has less demand size variability but a higher number of zero entries compared to the AUTO dataset. Further details about these datasets are provided in Table 1.

4.2 Evaluation

Conventional evaluation metrics. We first employed four metrics that are widely used. MAPE (Mean Absolute Percentage Error) is commonly used metric although there is ongoing debate [3, 2, 9] regarding the suitability, as discussed in Fig. 1. Since MAPE has a fundamental limitation due to its asymmetry between predictions and true values, we also adopted sMAPE (Symmetric Mean Absolute Percentage Error) to mitigate this issue. Another widely used metric, RMSE (Root Mean Square Error), can effectively compare the similarity between vectors or matrices. However, it lacks the

Table 2: The forecasting results based on layer changes in the encoder and discriminator.

	AUTO				RAF				M5			
F	MLP		LSTM		MLP		LSTM		MLP		LSTM	
D	MLP	LSTM	MLP	LSTM	MLP	LSTM	MLP	LSTM	MLP	LSTM	MLP	LSTM
MSTD	2.403	2.464	**1.897**	2.356	2.955	3.487	**2.790**	5.75	0.743	1.053	0.671	**0.512**
V-Recall	0.350	**1.000**	0.454	0.350	0.874	**0.972**	0.902	0.692	0.861	**0.896**	0.799	0.661
V-F1	0.373	**0.551**	0.419	0.373	0.896	**0.945**	0.911	0.788	0.742	**0.764**	0.728	0.672
RMSE	7.822	9.728	7.922	7.608	16.855	20.399	16.609	27.485	2.693	2.575	2.317	3.145
MAPE	0.478	0.687	0.443	0.524	0.944	0.978	0.953	0.722	0.827	0.855	0.814	0.799
sMAPE	1.139	1.315	1.11	1.165	1.968	1.963	1.966	1.914	1.61	1.677	1.568	1.539
SPEC	2.005	2.05	1.517	1.926	1.534	0.348	0.803	0.72	24.144	30.644	18.25	26.949



(a) MSTD (b) V-Recall (c) V-F1

Figure 4: Visualization of three proposed metrics along different cofigurations.

ability to account for temporal shifts often observed in time-series data. Furthermore, we include SPEC (Stock-keeping-oriented Prediction Error Cost) [2], a metric designed for demand forecasting which incorporates penalties for overstock and understock situations balanced by parameters α_1 and α_2 (we set both to 0.5). All equations are provided in Appendix B.

Evaluation beyond error. To address the limitations of existing metrics, we adopted several alternative approaches. The first metric, MSTD (Mean Standard Deviation), measures the alignment between the forecasts $y_{i,\hat{1:L}}$ and the actual values $y_{i,1:L}$. This metric evaluates how well the forecasting results capture the marginal distribution of the training dataset. Achieving a desirable MSTD indicates robustness against minor shifts in values (e.g., Fig. 1 (c)) and focuses on reflecting the inherent patterns and statistical properties of the dataset, thus mitigating the mean regression problem. The second approach evaluates the ability to realistically represent void spaces, specifically the intervals in irregular time series where no entries exist. To assess how accurately these intervals are forecasted, we used the recall (V-Recall) and the F1 score (V-F1). A high V-Recall indicates that the model successfully predicts the absence of values in areas where no data is expected, while a strong V-F1 score demonstrates this capability while minimizing errors in interval predictions.

4.3 Adversarial Components in Irregular Time-Series Forecasting

We first conduct experiments to assess the roles of adversarial components, specifically the encoder and discriminator, each of which can be configured in one of two ways: recursive and non-recursive. The simple framework is illustrated in Fig. 2 (a). The encoder takes an input $x_{1:P}$ and produces an output $y_{\hat{1:L}}$. In the recursive configuration, the encoder is implemented using two layers of LSTM [18] and fully-connected layer, while in the non-recursive one consists of four linear layers. Similarly, the discriminator is implemented using either two layers of LSTM in the recursive case or a four linear layers in the non-recursive case. For the non-recursive discriminator, a single logit is directly produced by the successive linear layers to determine whether the input is real or fake. In contrast, in the recursive discriminator, the LSTM generates a logit at each state which comes from each hidden states, then the average of these logits is computed after after passing through the final fully-connected layer. This approach can be interpreted as evaluating the realisticity of the transition at each step, given the history. The basic training method follows the approach used in GANs [4]. For details, please refer to the Appendix D.

Table 3: Comparison results between the adversarial approach and other baselines.

	AUTO				RAF				M5			
	MSTD	V-Recall	V-F1	MAPE	MSTD	V-Recall	V-F1	MAPE	MSTD	V-Recall	V-F1	MAPE
Croston [16]	2.403	0.027	0.049	0.507	**2.653**	0.520	0.663	**0.844**	1.270	0.580	0.661	**0.637**
ARIMA [20]	2.294	0.228	0.285	0.557	2.661	0.748	0.827	0.926	1.198	0.845	0.766	0.786
ADIDA [29]	2.403	0.039	0.069	0.477	2.653	0.706	0.798	0.919	1.270	0.646	0.702	0.676
MLP	2.273	0.421	0.411	**0.404**	2.664	0.845	0.881	0.982	1.072	0.895	0.766	0.844
RNN [30]	2.223	0.000	0.000	0.425	2.711	0.904	0.912	0.955	1.220	0.880	0.764	0.838
LSTM [18]	2.208	0.483	0.432	0.465	2.692	0.865	0.891	0.942	1.107	0.881	**0.770**	0.859
R-NR	**1.897**	0.454	0.419	0.443	2.79	0.902	0.911	0.953	**0.671**	0.799	0.728	0.814
NR-R	2.464	**1.000**	**0.551**	0.687	3.487	**0.972**	**0.945**	0.978	1.053	**0.896**	0.764	0.855

Across all datasets, the results in 2 emphasize the effectiveness of combining recursive and non-recursive architectures for irregular time series forecasting, enabling better modeling of both temporal dependencies and the marginal distribution. LSTM encoders paired with MLP discriminators tend to produce lower MSTD values, indicating better alignment with the marginal distribution of the data set. For V-Recall and V-F1, the opposite configurations consistently outperform, highlighting that LSTM discriminators excel at recognizing void intervals. Although the use of non-recursive MLP forecasters may seem unconventional in regular time series, they are particularly effective for irregular time series due to the absence of trends and seasonality. For conventional metrics, RMSE and MAPE generally favor configurations with LSTM encoders, although the effect of discriminator architecture appears less pronounced. Additionally, sMAPE and SPEC do not show consistent trends and appear to be more dataset-dependent.

4.4 Comparison with Baseline Models

For the baseline models, we selected three statistical models (Croston [16], ARIMA [20], and ADIDA [29]) and three neural network models (MLP, RNN [30], and LSTM lstm). Each baseline models are either known for their strong performance on irregualr patterns or demonstrating simple architecture but showing strong performance. More detailed information about the baseline models can be found in the Appendix C. The results in Table 3 show that, in terms of MSTD, statistical models generally underperformed, while simpler neural networks such as MLP, RNN, and LSTM exhibited moderate success. The best performance was achieved with adversarial learning, which closely matches the true variation in the dataset. Although the RAF dataset displayed an opposite trend, configurations of LSTM with non-recursive discriminator outperformed a pure LSTM by a significant margin in the other two datasets. Furthermore, V-Recall and V-F1 scores improved significantly with the adversarial approach, especially compared to other baseline models. This highlights the limitations of statistical models in irregular time series forecasting, emphasizing their unsuitability for capturing the complexities of such data. Table 3 presents results for the four primary metrics of interest, while the full results, including RMES, SPEC and sMAPE, are provided in Appendix E.

5 Conclusion

In conclusion, this study represents the first application of adversarial learning for forecasting irregular time series. Our findings reveal that while existing methods often perform well according to conventional metrics, they frequently produce results that are overly simplistic or unrealistic. Through experiments, we demonstrate that adversarial learning can effectively address this issue, offering a more sophisticated approach for forecasting irregular time series. Although traditional error-based metrics tend to favor conservative forecasting models that prioritize accuracy, we argue that it is crucial to also assess forecasting quality through the lens of capturing the unique characteristics and marginal distribution of the data. The introduction of adversarial learning may lead to increased variation in forecasting values, which could result in a loss of accuracy in error-based metrics, yet we view this as an important step towards more flexible and realistic forecasting. This work, despite its limitations, lays the foundation for future research aimed at developing new metrics and methodologies that better capture the nuances of irregular time series data and more closely align with human intuition in forecasting tasks.

References

[1] Konstantinos Nikolopoulos. We need to talk about intermittent demand forecasting. *European Journal of Operational Research*, 291(2):549–559, 2021.

[2] Dominik Martin, Philipp Spitzer, and Niklas Kühl. A new metric for lumpy and intermittent demand forecasts: Stock-keeping-oriented prediction error costs, 04 2020.

[3] Daniel Kiefer, Florian Grimm, Markus Bauer, and Clemens van Dinther. Demand forecasting intermittent and lumpy time series: Comparing statistical, machine learning and deep learning methods. 01 2021.

[4] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.

[5] Olof Mogren. C-rnn-gan: Continuous recurrent neural networks with adversarial training, 2016.

[6] Jinsung Yoon, Daniel Jarrett, and Mihaela van der Schaar. Time-series generative adversarial networks. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[7] Jinsung Jeon, JEONGHAK KIM, Haryong Song, Seunghyeon Cho, and Noseong Park. Gt-gan: General purpose time series synthesis with generative adversarial networks. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 36999–37010. Curran Associates, Inc., 2022.

[8] Sifan Wu, Xi Xiao, Qianggang Ding, Peilin Zhao, Ying Wei, and Junzhou Huang. Adversarial sparse transformer for time series forecasting. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 17105–17115. Curran Associates, Inc., 2020.

[9] Juyong Jiang, Binqing Wu, Ling Chen, Kai Zhang, and Sunghun Kim. Enhancing the robustness via adversarial learning and joint spatial-temporal embeddings in traffic forecasting. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. ACM, October 2023.

[10] Yiwei Zhang, Jinyang Li, Haoran Wang, and Sou-Cheng T. Choi. Sentiment-guided adversarial learning for stock price prediction. *Frontiers in Applied Mathematics and Statistics*, 7, 2021.

[11] Zhenzhen Chu, Chengyu Wang, Cen Chen, Dawei Cheng, Yuqi Liang, and Weining Qian. Learning invariant representations for new product sales forecasting via multi-granularity adversarial learning. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, CIKM '23, page 3828–3832, New York, NY, USA, 2023. Association for Computing Machinery.

[12] Jing Liu, Wanning Sun, and Mengjie Li. Recurrent conditional generative adversarial network for image deblurring. *IEEE Access*, 7:6186–6193, 2019.

[13] Cristóbal Esteban, Stephanie L. Hyland, and Gunnar Rätsch. Real-valued (medical) time series generation with recurrent conditional gans, 2017.

[14] Giorgia Ramponi, Pavlos Protopapas, Marco Brambilla, and Ryan Janssen. T-cgan: Conditional generative adversarial network for data augmentation in noisy time series with irregular sampling, 2019.

[15] Edward De Brouwer, Jaak Simm, Adam Arany, and Yves Moreau. Gru-ode-bayes: Continuous modeling of sporadically-observed time series, 2019.

[16] J. D. Croston. Forecasting and stock control for intermittent demands. *Operational Research Quarterly (1970-1977)*, 23(3):289–303, 1972.

[17] Zhongsheng Hua and Bin Zhang. A hybrid support vector machines and logistic regression approach for forecasting intermittent demand of spare parts. *Applied Mathematics and Computation*, 181(2):1035–1048, 2006.

[18] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, nov 1997.

[19] C. Chatfield. The holt-winters forecasting procedure. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 27(3):264–279, 1978.

[20] George.E.P. Box and Gwilym M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, 1976.

[21] Spyros Makridakis, Evangelos Spiliotis, and Vassilios Assimakopoulos. The m4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1):54–74, 2020. M4 Competition.

[22] Harris Drucker, Chris J. C. Burges, Linda Kaufman, Alex Smola, and Vladimir Vapnik. Support vector regression machines. In *Proceedings of the 9th International Conference on Neural Information Processing Systems*, NIPS'96, page 155–161, Cambridge, MA, USA, 1996. MIT Press.

[23] Aris Syntetos and John Boylan. The accuracy of intermittent demand estimates. *International Journal of Forecasting*, 21:303–314, 04 2005.

[24] Ali Caner Turkmen, Tim Januschowski, Yuyang Wang, and Ali Taylan Cemgil. Intermittent demand forecasting with renewal processes, 2020.

[25] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks, 2019.

[26] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks, 2020.

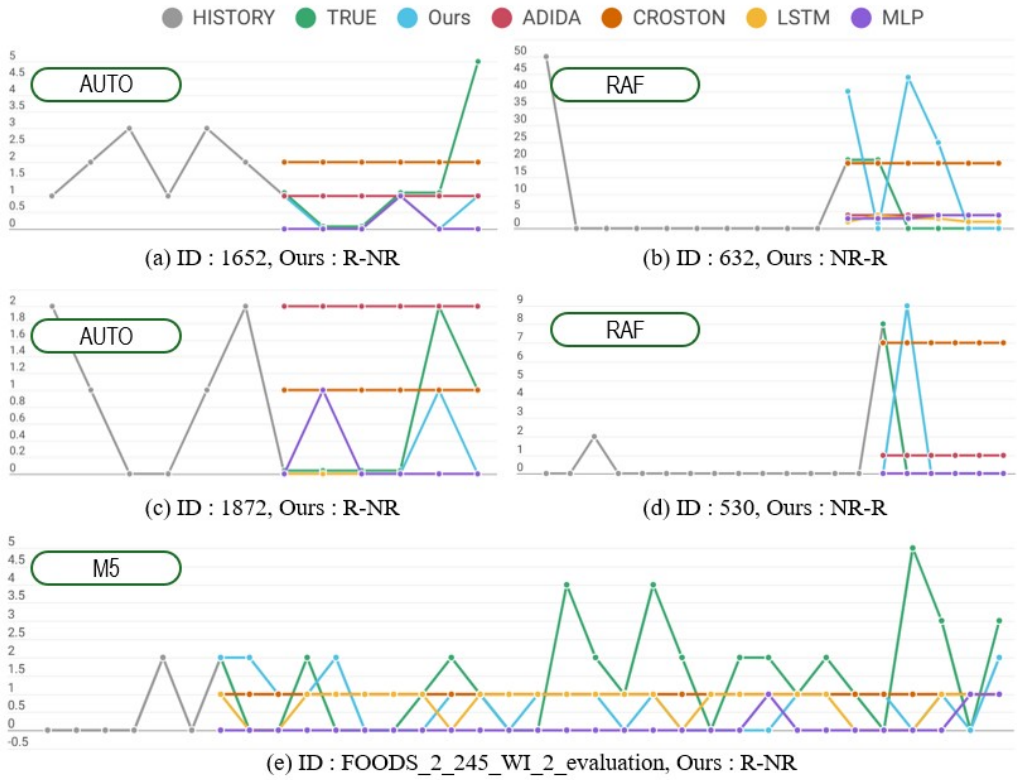
[27] AJ Hyvarinen and H Morioka. Nonlinear ica of temporally dependent stationary sources. 01 2017.

[28] Xiangchen Song, Weiran Yao, Yewen Fan, Xinshuai Dong, Guangyi Chen, Juan Carlos Niebles, Eric Xing, and Kun Zhang. Temporally disentangled representation learning under unknown nonstationarity, 2024.

[29] Konstantinos Nikolopoulos, Aris Syntetos, John Boylan, Fotios Petropoulos, and Vassilis Assimakopoulos. An aggregate-disaggregate intermittent demand approach (adida) to forecasting: An empirical proposition and analysis. *JORS*, 62:544–554, 03 2011.

[30] David E. Rumelhart and James L. McClelland. *Learning Internal Representations by Error Propagation*, pages 318–362. 1987.

A Real-world Examples of Mean Regression Problem and MAPE in Irregular time Series



(a) ID : 1652, Ours : R-NR (b) ID : 632, Ours : NR-R

(c) ID : 1872, Ours : R-NR (d) ID : 530, Ours : NR-R

(e) ID : FOODS_2_245_WI_2_evaluation, Ours : R-NR

Figure 5: Real-world examples that demonstrate the existence of the problem we are addressing.

Table 4: Calculated MAPE in each example.

	OURS	CROSTON	ADIDA	LSTM	MLP
(a)	0.45	0.2	0.9	0.75	0.75
(b)	1	0.8	0.05	0.85	0.85
(c)	0.75	0.75	0.25	1	1
(d)	1	0.875	0.125	1	1
(e)	0.06	0.05	0.01	0.06	0.06

B Conventional Evaluation Metric

The following equations represent the methods used to evaluate our experimental results. The notation is consistent with that in Section 3.1.

$$MAPE = \frac{1}{LM}\sum_{i=1}^{M}\sum_{n=1}^{L}\frac{|y_{in} - \hat{y_{in}}|}{y_{in}}) = \frac{1}{M}\sum_{i=1}^{M}(\frac{1}{\sum_{n=1}^{L}\mathbb{I}[y_{in} > 0]}\sum_{n=1}^{L}\mathbb{I}[y_{in} > 0]\frac{|y_{in} - \hat{y_{in}}|}{y_{in}}) \tag{4}$$

$$sMAPE = \frac{2}{LM}\sum_{i=1}^{M}\sum_{n=1}^{L}\frac{|y_{in} - \hat{y_{in}}|}{y_{in} + \hat{y_{in}}}$$

$$= \frac{2}{M}\sum_{i=1}^{M}(\frac{1}{\sum_{n=1}^{L}\mathbb{I}[y_{in} + \hat{y_{in}} > 0]}\sum_{n=1}^{L}\mathbb{I}[y_{in} + \hat{y_{in}} > 0]\frac{|y_{in} - \hat{y_{in}}|}{y_{in} + \hat{y_{in}}}) \tag{5}$$

$$SPEC = \frac{1}{ML}\sum_{i=1}^{M}\sum_{n=1}^{L}\sum_{m=1}^{n}((n-m+1)\cdot max(0; \alpha_1 \cdot min(y_{im}; \sum_{k=1}^{m} y_{ik} - \sum_{j=1}^{n} f_{ij});$$

$$\alpha_2 \cdot min(f_{im}; \sum_{k=1}^{m} f_{ik} - \sum_{j=1}^{n} y_{ij}))) \tag{6}$$

$$RMSE = \frac{1}{M}\sum_{i=1}^{M}\sqrt{\sum_{n=1}^{L}\frac{(y_{in} - \hat{y_{in}})^2}{n}} \tag{7}$$

C Baseline models

- **Croston** [16]: A method specifically designed for intermittent and lumpy time series forecasting. It decomposes demand into occurrence and size, making it effective for irregular patterns in time-series data.
- **ARIMA** [20]: A widely-used statistical model that combines autoregressive (AR) and moving average (MA) components with differencing to handle non-stationarity. It is effective for capturing linear temporal dependencies in regular time series but struggles with highly irregular patterns.
- **ADIDA (Aggregate-Disaggregate Intermittent Demand Approach) [29]**: A technique tailored for intermittent time series. It aggregates demand over fixed intervals to smooth irregular patterns and then applies standard forecasting methods to generate predictions.
- **MLP (Multilayer Perceptron)**: A feedforward neural network model that learns non-linear relationships between input and output. While flexible, it may require careful tuning for effective performance on time series with irregular patterns.
- **RNN (Recurrent Neural Network) [30]**: A neural network model with feedback connections that capture temporal dependencies in sequential data. RNNs are powerful for regular time series but can struggle with long-term dependencies or highly intermittent patterns.
- **LSTM (Long Short-Term Memory) [18]**: A specialized type of RNN designed to overcome the vanishing gradient problem, making it suitable for learning long-term dependencies in sequential data. LSTM can handle some degree of irregularity but requires significant computational resources.

D Experiment details

We first set look-back period and forecasting horizon for each dataset. We follow setting of kaggle competition for M5 dataset, thus both P and L are set to 28. AUTO and RAF are having L of 6, while P is set to 6 and 18 respectively. AUTO is having shorter look-back period due to the lack of historical data, as we only have 24 timestamp through out whole dataset.

In our experiments, we train encoder and discriminator jointly. Throughout all configurations, we trained for 100 epochs and hyperparameter tuning has been done in logarithmic scale. We selected our best model by using MAPE which can be considered as convention, and we observed how the models behave throughout in other metrics. All seeds in the experiment is set to 0 and we set a batch size to 256. During training, we first update parameters in encoder followed by discriminator.

E Additional Experiment Results

	AUTO				**RAF**				**M5**			
	MAPE	RMSE	sMAPE	SPEC	MAPE	RMSE	sMAPE	SPEC	MAPE	RMSE	sMAPE	SPEC
Croston [16]	0.507	8.790	1.109	2.831	0.844	24.672	1.949	6.562	0.637	2.288	1.425	18.796
ARIMA [20]	0.557	7.543	1.288	1.872	0.926	16.193	1.967	1.915	0.786	2.204	1.551	0.718
ADIDA [29]	0.477	7.838	1.122	2.074	0.919	16.483	1.966	2.172	0.676	2.268	1.445	17.375
MLP	0.404	7.879	1.068	1.802	0.982	16.284	1.986	0.906	0.844	2.421	1.664	29.477
RNN [30]	0.425	7.911	1.089	1.632	0.955	16.623	1.974	1.189	0.838	2.253	1.647	25.398
LSTM [18]	0.465	8.020	1.129	1.904	0.942	16.649	1.973	1.680	0.859	2.422	1.691	31.623
R-NR	0.443	7.922	1.110	1.517	0.953	16.609	1.966	0.803	0.814	2.317	1.568	18.250
NR-R	0.687	9.728	1.315	2.050	0.978	20.399	1.963	0.348	0.855	2.575	1.677	30.644