Timeless Foundations: Exploring DC-VAEs as Foundation Models for Time Series Analysis

Gastón García González*, Pedro Casas†, Emilio Martínez*, Alicia Fernández*
*IIE–FING, Universidad de la República, Montevideo – Uruguay
†AIT, Austrian Institute of Technology, Vienna – Austria

***Abstract*—We investigate a novel approach to time-series modeling, inspired by the successes of large pre-trained foundation models. We introduce FAE (Foundation Auto-Encoders), a foundation generative-AI model for anomaly detection in time-series data, based on Variational Auto-Encoders (VAEs). By Foundation Model (FM), we mean a model pre-trained on massive amounts of time-series data which can learn complex temporal patterns useful for accurate modeling and forecasting on previously unseen datasets. FAE leverages VAEs and Dilated Convolutional Neural Networks (DCNNs) to build a generic model for time-series modeling, which could eventually perform properly in out-of-the-box, zero-shot anomaly detection applications. We introduce the main concepts and ideas of this FM for time-series (TSFM), and present some preliminary results in a multi-dimensional mobile network monitoring dataset. We also present example results applying novel TSFMs to this dataset, both in a zero-shot manner and relying on fine-tuning, and show how complex it is in the practice to achieve accurate results.**

***Index Terms*—Time-Series Data, Generative AI, Anomaly Detection, VAE, Foundation Models for Time-Series**

I. Introduction

Inspired by the success of Foundation Models (FM) in other areas of AI, a new approach is emerging for time-series analysis: foundation models for time series (TSFMs). These models build upon the concept of large, pre-trained models that can be adapted to various tasks. Similar to how large language models (LLMs) process text, TSFMs are trained on massive amounts of time-series data. They leverage transformer architectures to identify patterns and relationships within the sequences. TSFMs represent a new generation of time-series analysis tools with improved performance and broader applicability.

We focus on devising a Generative AI model capable of matching or surpassing the performance of conventional time-series modeling methods without the need for training on the specific target dataset - a concept known as Zero-Shot Learning (ZSL). ZSL is a problem setup where, at test time, a learner needs to analyze samples from classes which were not observed during training. The ZSL concept is powerful and appealing for anomaly detection applications, offering multiple advantages: firstly, it simplifies the application of the model for time-series modeling, eliminating the requirement for specialized knowledge of fine-tuning techniques and the significant computational resources associated with them; secondly, it naturally aligns with scenarios characterized by limited data availability, where training or fine-tuning data is limited; lastly, by harnessing the comprehensive pattern extrapolation capabilities of large pre-trained models, it circumvents the substantial time, effort, and domain-specific expertise demanded for crafting dedicated time-series models.

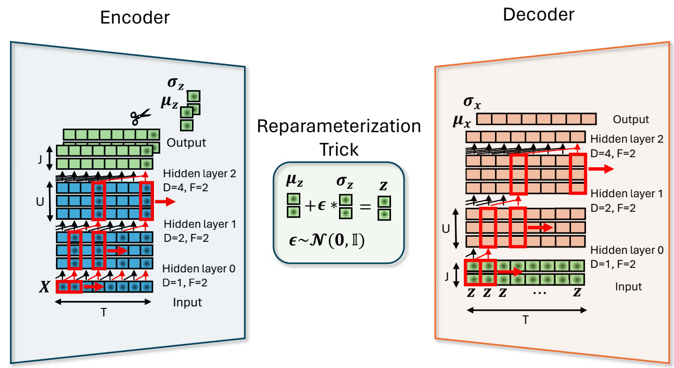


Fig. 1. FAE's encoder/decoder architecture using causal dilated convolutions, implemented through a stack of 1D convolutional layers.

We therefore investigate if a model pre-trained on multiple time-series data can learn temporal patterns useful for accurate forecasting on previously unseen time-series. We introduce FAE (Foundational Auto-Encoders), a foundational generative-AI model for anomaly detection in time-series data, based on Variational Auto-Encoders (VAEs) [1]. VAEs are generative AI models that learn the underlying distribution of the data and can generate new samples from this distribution. In the context of time-series data, VAEs can capture latent representations of temporal patterns and generate sequences that exhibit similar characteristics, making them powerful for generalization and ZSL. FAE uses DC-VAE's network architecture [2], originally designed for multivariate anomaly detection. In particular, it leverages VAEs and Dilated Convolutional Neural Networks (DCNNs) to build a generic model for *univariate time-series modeling*, which could eventually perform properly in out-of-the-box, zero-shot anomaly detection applications.

II. Related Work

Approaches to time-series anomaly detection based on deep learning technology have flourished in recent years [3]. Transformer-based models [4] have been gaining popularity in recent years for time-series analysis, given their remarkable performance in large-scale settings, such as long sequence time-series forecasting. There is a recent surge in papers targeting the conception of TSFMs, capable of generating accurate predictions for diverse datasets not seen during train-

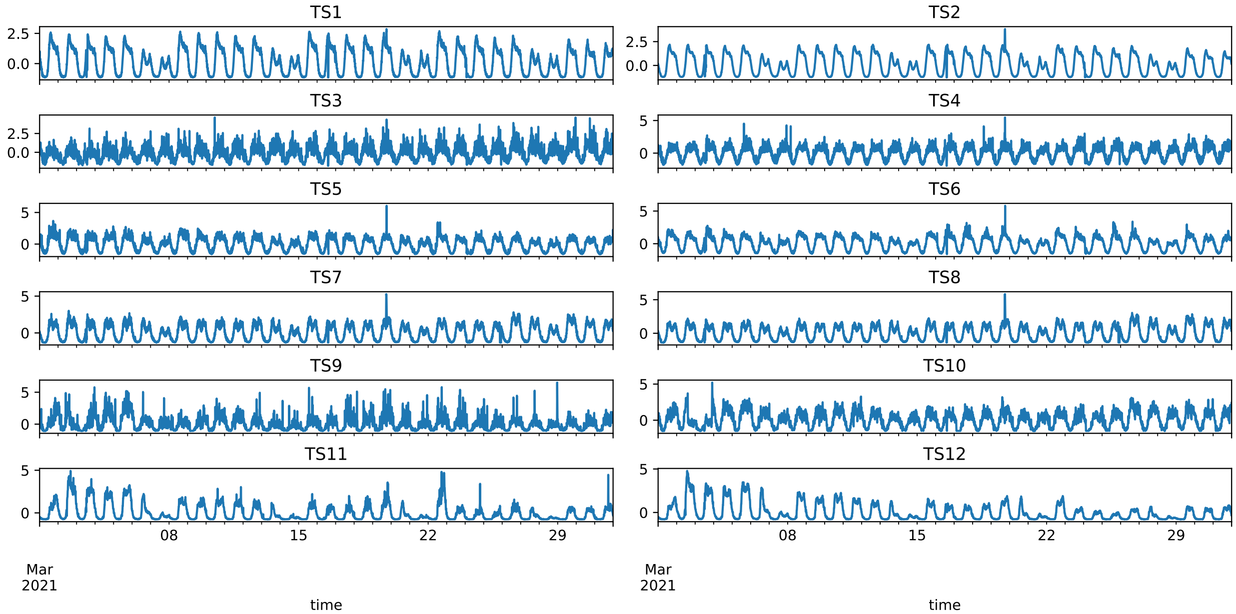


Fig. 2. TELCO time-series, for one month worth of data (March 2021), sampled at a five minutes rate.

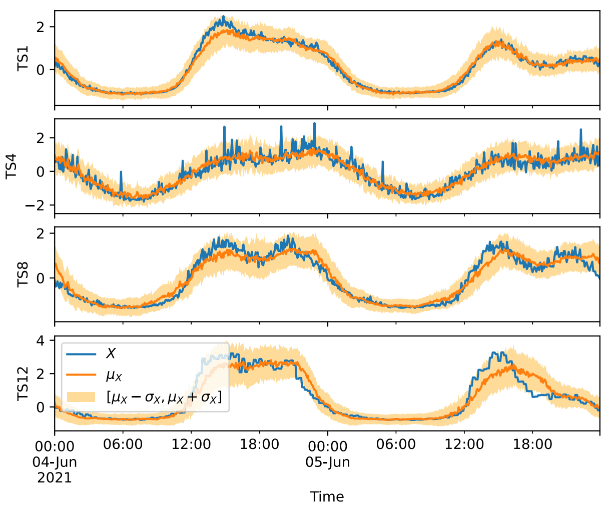
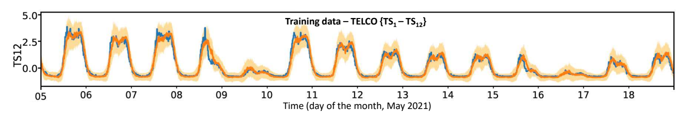
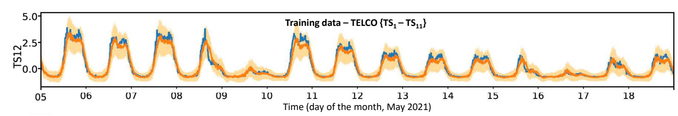


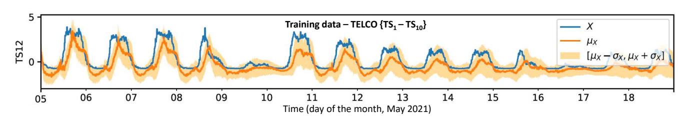
Fig. 3. Predictions with fully-trained FAE (12 time-series) in two days of testing samples from June 2021 (Friday and Saturday).



(a) FAE predictions for TS_{12}, with full-FAE (12 time-series).



(b) FAE predictions for TS_{12}, with FAE trained without TS_{12}.



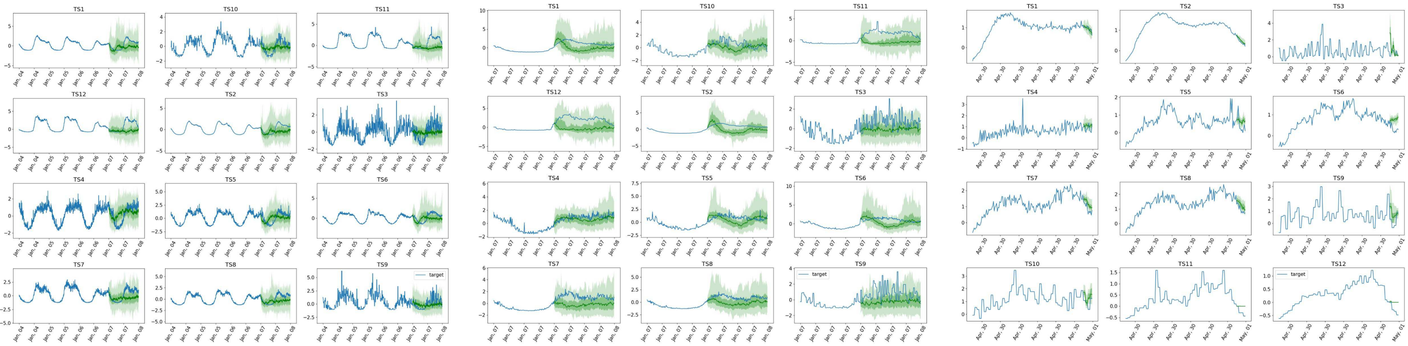
(c) FAE predictions for TS_{12}, with FAE trained without TS_{11} and TS_{12}.

Fig. 4. Zero-shot modeling experimentation, predicting TS_{12} for two weeks in the testing dataset (May 2021).

ing. The underlying concept of these models is to rely on highly expressive, large-scale Transformer architectures which are trained on billions of time-series data points, coming from very diverse domains and having high heterogeneity in terms of temporal behaviors and characteristics. TimeGPT-1 [5], PromptCast [6], LLMTime [7], TimesFM [8], Lag-Llama [9], and Time-LLM [10] are all recent examples of novel TSFMs for forecasting, targeting a ZSL application. FAE follows exactly this concept, but using a much smaller and simpler architecture. While this adds limitations in terms of expressiveness and therefore generalization capabilities, it also opens the door to the exploration of other venues, such as the combined utilization of smaller foundation models in the form of ensembles, in combination with domain detection strategies.

III. FAE Model, Performance, and TSFMs

Figure 1 depicts the encoder/decoder architecture used in FAE, which is an adaptation of DC-VAE's architecture, for the case of univariate time-series analysis. The FAE model functions as a univariate model trained on various series within a system simultaneously, treating them as distinct classes of series. Similar to the original DC-VAE version, FAE allows for monitoring of all time-series within a MTS process using a single model, albeit analyzing one time-series at a time. The architecture, based on dilated convolutional neural networks (DCNNs), is capable to exploit the temporal dependence of values for longer sequences. The main difference with DC-VAE is that the new architecture has to accommodate univariate input samples $\boldsymbol{X} \in \mathbb{R}^{1\times T}$, rather than multivariate ones. To maintain the concept of compression – i.e., the dimension of



(a) ZS: context is 72hs, prediction for 6hs. (b) ZS: context is 72hs, prediction for 6hs. (c) FS: context is 12hs, prediction for 1hr.

Fig. 5. TELCO time-series forecasting performance using Lag-Llama TSFM model. (a,b) Lag-Llama is directly applied in a zero-shot (ZS) manner, without fine-tunning. (c) Few-shot (FS) application: Lag-Llama is fine-tuned with TELCO data.

the latent space $\boldsymbol{Z}$ has to be lower than the input dimension of $\boldsymbol{X}$ – the latent space in FAE reduces dimensionality along the temporal dimension; in DC-VAE, the dimensionality reduction operates in the spatial dimension.

A. FAE Modeling Performance

We experiment with FAE in the analysis of the TELCO dataset [11], an open MTS dataset (12 time-series) arising from the monitoring of an operational mobile ISP, consisting of seven months-worth of time-series with different structural properties. Figure 2 depicts a one-month example from the complete TELCO MTS dataset. We take a 3/1/3-months temporal split for training/validation/testing purposes.

We evaluate the prediction performance of FAE in samples from the testing set, considering training on the full three months of data, for the 12 time-series, i.e., more than 300.000 samples. Figure 3 depicts the resulting predictions $\boldsymbol{\mu}_{\boldsymbol{X}_t}$ and $\boldsymbol{\sigma}_{\boldsymbol{X}_t}$ for two days of testing samples $\boldsymbol{X}_t$ from June 2021, for four representative time-series, including TS_1, TS_4, TS_8, and TS_{12}. To add more variability, we consider a working day (Friday 4th) and a weekend day (Saturday 5th). FAE can properly track different types of behavior in the time-series, including the strong seasonal daily component, but also the operation during workdays and weekends, clearly visible in TS_{12}. Interesting to note is how different periods of time-series variability result in more or less tight normal-operation regions estimated by FAE, as defined by $\boldsymbol{\sigma}_{\boldsymbol{X}_t}$.

We investigate now the performance of FAE in a zero-shot setting, testing the model for time-series not seen at training time. We focus the analysis on TS_{12}, due to its combined seasonality and particular temporal trend, as well as its strong correlated behavior to TS_{11} (cf. Figure 2). We train FAE on three different training datasets from TELCO, considering a different number of time-series TS_i. The first model uses all 12 time-series – we refer to it as *full-FAE*; the second model considers a zero-shot setting for TS_{12}, with a training dataset which includes time-series TS_1 to TS_{11}, leaving out all samples from TS_{12}; given the strong temporal correlation between TS_{12} to TS_{11}, we also train a third model leaving out all samples from TS_{11} and TS_{12}, i.e., training in TS_1 to TS_{10}.

Figure 4 presents the prediction performance of the three models, when applied to two weeks of TS_{12} samples, from May 5 to May 19, 2021. In Figure 4(a), the modeling performance for full-FAE is optimal, as it can properly track the different behaviors and patterns in the time-series, similarly to Figure 3. A similar performance is observed in Figure 4(b) for the second model, which learns the characteristics of TS_{12} at training time, from TS_{11}. Not surprisingly, the performance of the third model in Figure 4(c) is significantly worse than for the other two models, given the lack of a similar temporal pattern in the training data. To some extent, there is an identification with the patterns observed in time-series TS_1 – note how the daily sharp peaks are exacerbated. Nevertheless, FAE manages to capture and track the monthly downtrend, even without previous evidence of it.

B. TSFMs in TELCO - Applying Lag-Llama Forecasting

To conclude, we apply Lag-Llama [9], a TSFM model recently published as open software (https://github.com/time-series-foundation-models/lag-llama), to the 12 time-series in TELCO, and see how good it performs as a ZSL model. Lag-Llama is a general-purpose FM for univariate probabilistic time series forecasting, based on a decoder-only transformer architecture, pre-trained on a large corpus of diverse time series data from several domains. We test two different scenarios: (i) a pure zero-shot (ZS) application, where we apply the model directly to TELCO, following the recommendations of the authors for setting the length of the context for prediction – the context defines the number of samples it uses as input to forecast future values; and (ii) a few-shot-learning application (FS), where we fine-tune the model on TELCO training. Figures 5(a,b) depict the results in the ZS tests, and Figure 5(c) show the FS performance. The ZS performance is clearly poor, as the model is unable to forecast and correctly track any of the time series, despite using a context (input data) of three full days. For the fine-tunning scenario, results seem to be significantly better, for a short forecasting of one hour. While the approach employed by FAE is different, as it is not based on forecasting, we can appreciate how better the model is in tracking the correct shape of the time-series, as compared in this case to Lag-Llama.

Acknowledgments

This work has been partially supported by the Austrian FFG ICT-of-the-Future project *DynAISEC* – Adaptive AI/ML for Dynamic Cybersecurity Systems – project ID 887504, the Austrian Science Fund, FWF project *GRAPHS4SEC* – Graph Neural Networks for Robust AI/ML-driven Network Security Applications – grant number I-6653, and by the Uruguayan CSIC project Generalization and Domain Adaptation in Time-Series Anomaly Detection, project ID CSIC-I+D-22520220100371UD.

References

[1] D. P. Kingma and M. Welling, "Auto-encoding Variational Bayes," *CoRR*, vol. abs/1312.6114, 2013. [Online]. Available: https://arxiv.org/abs/1312.6114

[2] G. García González, S. Martinez Tagliafico, A. Fernández, G. Gómez, J. Acuña, and P. Casas, "One Model to Find Them All – Deep Learning for Multivariate Time-Series Anomaly Detection in Mobile Network Data," *IEEE Transactions on Network and Service Management*, pp. 1–1, 2023.

[3] G. Pang, C. Shen, L. Cao, and A. V. D. Hengel, "Deep learning for anomaly detection: A review," *ACM Comput. Surv.*, vol. 54, no. 2, Mar. 2021.

[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser, and I. Polosukhin, "Attention is All you Need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[5] A. Garza and M. Mergenthaler-Canseco, "TimeGPT-1," 2023.

[6] H. Xue and F. D. Salim, "PromptCast: A New Prompt-Based Learning Paradigm for Time Series Forecasting," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–14, 2023.

[7] N. Gruver, M. Finzi, S. Qiu, and A. G. Wilson, "Large Language Models Are Zero-Shot Time Series Forecasters," 2023.

[8] A. Das, W. Kong, R. Sen, and Y. Zhou, "A Decoder-only Foundation Model for Time-series Forecasting," 2024.

[9] K. Rasul, A. Ashok, A. R. Williams, H. Ghonia, R. Bhagwatkar, A. Khorasani, M. J. D. Bayazi, G. Adamopoulos, R. Riachi, N. Hassen, M. Biloš, S. Garg, A. Schneider, N. Chapados, A. Drouin, V. Zantedeschi, Y. Nevmyvaka, and I. Rish, "Lag-Llama: Towards Foundation Models for Probabilistic Time Series Forecasting," 2024.

[10] M. Jin, S. Wang, L. Ma, Z. Chu, J. Y. Zhang, X. Shi, P.-Y. Chen, Y. Liang, Y.-F. Li, S. Pan, and Q. Wen, "Time-LLM: Time series forecasting by reprogramming large language models," in *International Conference on Learning Representations (ICLR)*, 2024.

[11] G. García González, S. Martínez Tagliafico, A. Fernández, G. Gómez, J. Acuña, and P. Casas, "TELCO – a new Multivariate Time-Series Dataset for Anomaly Detection in Mobile Networks," 2023. [Online]. Available: https://dx.doi.org/10.21227/skpg-0539