

Decoding Latent Spaces: Assessing the Interpretability of Time Series Foundation Models for Visual Analytics

Inmaculada Santamaria-Valenzuela¹ , Victor Rodriguez-Fernandez¹ , Javier Huertas-Tato¹ , Jong Hyuk Park² , David Camacho¹  *

¹ Computer Systems Engineering Dpt., Universidad Politécnica de Madrid, Madrid (Spain)

² Department of Computer Science and Engineering, Seoul National University of Science and Technology, Seoul (Korea)

* Corresponding author: mi.santamaria@upm.es (I. Santamaria-Valenzuela), victor.rfernandez@upm.es (V. Rodriguez-Fernandez), javier.huertas.tato@upm.es (J. Huertas-Tato), jhpark1@seoultech.ac.kr (J. H. Park), david.camacho@upm.es (D. Camacho)

Received 18 April 2025 | Accepted 2 June 2025 | Early Access 11 May 2026



ABSTRACT

The present study explores the interpretability of latent spaces produced by time series foundation models, focusing on their potential for visual analysis tasks. Specifically, we evaluate the MOMENT family of models, a set of transformer-based, pre-trained architectures for multivariate time series tasks such as: imputation, prediction, classification, and anomaly detection. We evaluate the capacity of these models on five datasets to capture the underlying structures in time series data within their latent space projection and validate whether fine tuning improves the clarity of the resulting embedding spaces. Notable performance improvements in terms of loss reduction were observed after fine tuning. Visual analysis shows limited improvement in the interpretability of the embeddings, requiring further work. Results suggest that, although time series foundation models such as those in MOMENT are robust, their latent spaces may require additional methodological refinements to be adequately interpreted, such as alternative projection techniques, loss functions, or data preprocessing strategies. Despite the limitations of MOMENT, foundation models suppose a big reduction in execution time and so a great advance for interactive visual analytics.

KEYWORDS

Foundation Models, Latent Spaces, Time Series, Visual Analytics.

DOI: 10.9781/ijimai.2026.6558

I. INTRODUCTION

VISUAL Analytics are essential for interpreting large time series data in fields like finance, enabling human understanding and supporting algorithm-assisted decision making [1], [2]. In this field, understanding Deep Learning (DL) models' embedding space has been a major focus. Different interactive visualization tools support dataset analysis by exploring these embedding spaces [3]–[5].

In other domains like computer vision and natural language processing, Transfer Learning (TL) has proven effective at reusing the model knowledge across datasets, enabling faster analysis and reducing training time in new scenarios. Bringing this efficiency into the visual analytics of time series could significantly enhance interactivity. However, unlike image or text data, TS does not always exhibit shared semantic information between datasets and domains -the same change in the shape of a TS could be an anomaly in a machine analysis but not in stock analysis. This raises important questions about how well TL can be adapted to TS [6]. However, TL is used for different TS tasks, such as forecasting workload on cloud platforms [7] or forecasting

influenza hospitalization with small datasets [8]. Also, Fawaz et al. [9] motivate practitioners to no longer train models from scratch for TS classification but instead to fine-tune pre-trained models, based on the idea that models can be adapted. But this still raises various questions. First, “is the size of the dataset the only important thing, or is context explicitly necessary too?” Second, “even in the same domain, is it possible to swap semantics between datasets?”, “is it enough with a fine-tune, or does a TS model need extra semantic information for building the TS?”.

Following this path, recent studies have shown great advances in time series foundation models (TSFMs), with good performance in different tasks (classification, anomaly detection, etc.), leading to their widespread adoption in the field. New TSFM [10]–[12] are appearing that could enhance visual analytics by providing more interpretable, fast, and interactive analysis. Specifically, the MOMENT family [10] has been analyzed to show interpretable embedding projections for different synthetic datasets (see Fig. 1) and to perform well against the current state of the art [10].

Please cite this article as: I. Santamaria-Valenzuela, V. Rodriguez-Fernandez, J. Huertas-Tato, J. H. Park, D. Camacho. Decoding Latent Spaces: Assessing the Interpretability of time series foundation models for Visual Analytics, *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 9, no. 7, pp. 47-66, 2026, <http://doi.org/10.9781/ijimai.2026.6558>

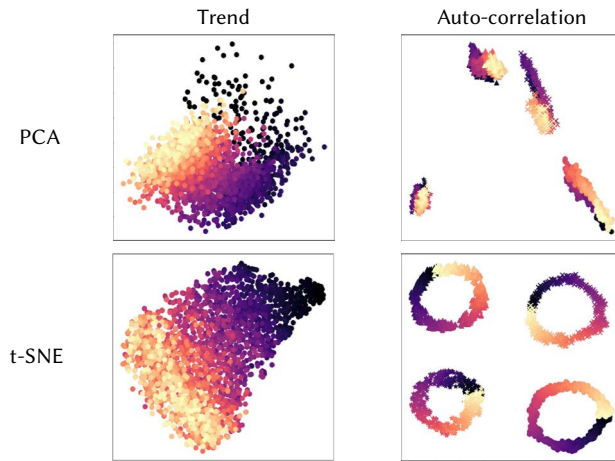


Fig. 1. Embedding space inferred by the large version of MOMENT (fine-tuned) for synthetically generated datasets projection using PCA (top) and t-SNE (bottom) projections. Shows how subtle trend and auto-correlation are captured. This suggests that TSFM embeddings may encode rich structural features that could be exploited for interactive exploration. Figure modified from [10] under the terms of Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>).

In this context, two main questions arise for the field of visual analytics:

- RQ1** Do quantitative improvements in classical tasks (measured as loss reduction during training) lead to enhanced interpretability of the embedding space?
- RQ2** How much model tuning (quantified as the percentage of labeled data used) is needed to obtain high-quality, descriptive clusters for a given time series dataset, as perceived through experimental evaluation of behavioral pattern detectability?

We explore these questions by conducting both statistical and visual analyzes using DeepVATS [5] as the evaluation framework. Initially developed as a Deep Learning-based tool, DeepVATS is a Visual Analytics tool for time series that aims to analyze the embedding space of long time series (TS) by leveraging DL models to extract vectorized representations, providing a more intuitive comprehension of their structure.

However, the high computational cost of this approach limits interactivity as dataset size increases. To enhance the interactivity - and with an eye in the Data Mining field - MPlot (similarity matrix plot) was integrated into DeepVATS [13]–[15]. Although computationally expensive, these methods have been successfully scaled for large-scale TS exploration. The integration of MPlot into DeepVATS has significantly enhanced trend detection capabilities for univariate time series while also providing a fast preview of the series' properties [13]. This combination facilitates an initial exploration of univariate TS while the Deep Learning model is still being trained, improving the overall workflow efficiency.

While this integration supports fast previews and initial exploration, the question remains whether taking a step back into the DL field, TSFM embeddings can offer deeper insights—not just a preview—of both univariate and multivariate time series properties. To analyze this apportionment, we integrate MOMENT into DeepVATS and analyze its embedding space using the same datasets and projection techniques as in Rodriguez-Fernandez et al. [5] to ensure a fair comparison.

The findings suggest that quantitative improvements are not necessarily linked to the embeddings' precision, and that even minimal fine-tuning can significantly change a model's internal knowledge. This indicates that extensive training may not be required for effective adaptation, making it a great idea to integrate foundation models into

visual analytics tools. In the next steps, we aim to refine this fine-tuning approach to produce clearer embeddings that better capture the diverse properties of time series data for specific tasks.

Thus, the central contribution of this work is the introduction of the MOMENT family into DeepVATS, resulting - to the best of our knowledge - in the first integration of foundation models into visual analytics tools. This integration has been done, enabling the adaptation and fine-tuning of any pretrained model in a zero-shot mode. This approach offers a flexible and efficient way of visually analyzing large time series data.

The next sections are structured as follows. Section II introduces foundation models for time series and, more specifically, the MOMENT family. Section III shows the analysis of four synthetic datasets and one real data dataset within deepVATS using MOMENT's models and analyzes its advantages and disadvantages. Section IV summarizes the analysis and proposes future lines for improving visual interactive analytic tools for time series.

II. BACKGROUND

The past two years have been especially important for the development of time series models. At the beginning, Transformers were thought to lose temporal correlation, raising doubts about their use for time series analysis [17]. Few time later, Wu et al. [18] added a new AutoCorrelation mechanism that fixed the temporal correlation problem, making transformers adaptable to time series. Fig. 2 shows how, only in 2023, different transformer models [19] were adapted to time series, and their history still continues in the present year (see timexer [20], which adds exogenous variates, or Medformer [21], developed for specific medical classification tasks). Their flexibility made them easy to adapt for pre-trained models and foundation models, showing great results in other fields like image [22], [23] or text [24], [25], leading to the option of using foundation models for time series.

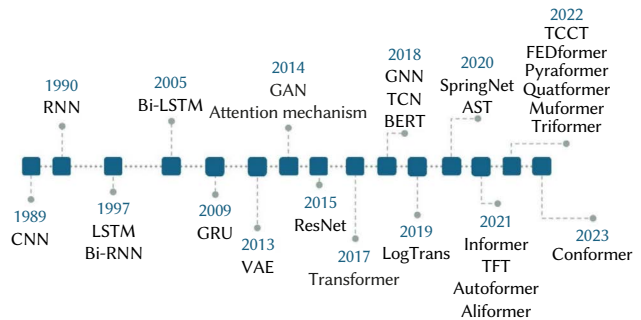


Fig. 2. Historical development of time series forecasting deep learning. Figure obtained from [16] under the terms of Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

At first, transfer learning was thought not to be appropriate for TS as Woo et al. [6] describe: “Unlike image and text data which naturally share semantic information across datasets and domains, time series data may not enjoy such properties of transferability as the semantics of time series data may be unique to their dataset or domain. As such, it is still unclear how time series models can benefit from pre-training and transfer learning”. However, in the same year, transfer learning for specific tasks appeared, motivating practitioners to no longer train models from scratch when classifying time series, but instead fine-tune pre-trained models [7]–[9]. Eventually, following the positive path, the TSFMs appeared, which have increasingly gained importance in the field [10], [12], [26], [27].

A. Most Influential TSFMs

To the best of our knowledge, the most influential TSFMs, based on the transformer architecture, are the following five:

Chronos An univariate time series forecasting model based on language architectures [11].

MOIRAI A Masked Encoder-based Universal time series Forecasting Transformer, Trained on a Large-scale Open time series Archive (LOTSAs) featuring over 27 billion observations across nine domains (energy, climate, cloud operations, transport, web, sales, finance, nature, and healthcare) [12].

Lag-LLaMa A pre-trained decoder-only transformer with lag features, leveraging LLaMA to univariate probabilistic time series forecasting [28].

TimeGPT-1 A proprietary TSFM designed for generative forecasting across diverse domains [27].

MOMENT A family of multivariate TSFMs, designed for multiple tasks [10].

B. Visual Analytics for Time Series

Embedding space visualization is an important tool used to interpret how DL models internally represent high-dimensional data. These projections - typically computed using techniques like t-SNE [29], UMAP [30] or the more recent PacMAP [31] - allow analysts to explore the structure of the data, detect outliers, and evaluate how well a model captures the intrinsic characteristics. This is used across different domains such as genomics [32] or healthcare-related time series [33]. These areas usually represent data in high-dimensional tabular matrices, and visualizing their embeddings enables researchers to observe patterns, disease progressions, or treatment effects.

Tools like TimeClustser [34] project embeddings extracted from sliding-windowed time series into scatter plots to identify intrinsic behaviors, while DeepVATS [5] combines masked autoencoders and data mining tools to better capture patterns. In addition, it integrates an interactive interface that allows users to link points in embedding space regions with raw temporal behavior, enhancing the interpretability within the embedding space analysis.

Focusing on temporal tabular data, in the healthcare domain, Stym-Popper et al. [33] analyze tabular information that could be summarized as time series, using a DL model embedding space projection to support hypertension diagnosis. Adding interaction to such visual analyses could provide additional insights into patient evolution. In the sports and transport domain, Adrienko et al. [35] use visual embedding projections to expose groups of time windows with similar feature values, allowing interactive selection of groups for inspection and eventual labeling. However, the tool does not show an interactive relation between the scatter plot and the raw data. Combining their labeling capacity with the interactivity of DeepVATS would greatly add interpretability to the results.

C. DeepVATS Application

DeepVATS [5] is a Deep Visual Analytics tool for Time Series that focuses on the modeling of intrinsic characteristics of the time series through Deep Learning models, showing the results in a visual interactive plot of their latent space and the original plot. The DL backbone model is a TimeInception-based masked time series autoencoder (MTSAE) that uses a masked value prediction callback (MVP) from 'tsai' package [36], [37]. The application is thought to be suitable for four main tasks: detection of anomalies, segmentation, pattern detection, and trend detection.

All foundation models have an embedding space that should be able to detect those characteristics. In particular, MOMENT's models have already been trained for multitasks including imputation,

anomaly detection, and classification; therefore, it should achieve fair results within those tasks. In fact, Goswami et al. [10] defend that the embedding space using the PCA and TSNE dimensionality reduction techniques shows that the models fit adequately to the trend of the timeseries, showing interpretable embedding spaces (see Fig. 1). Thus, the integration of MOMENT into DeepVATS could aid in the achievement of an interactive application with few training times for large time series; the integration includes both tools to test the interpretability of MOMENT within the deepVATS benchmark.

D. The MOMENT Family

The MOMENT TSFM family relates to transformer models (see Fig. 3) with three different size versions pre-trained: small, base, and large (see Table I). Also, each version can be loaded in four different modes (tasks): reconstruction, embedding, forecasting, and classification. Each task modifies the structure of the transformer (see Fig. 3) to use a specific output head based on the task to be solved, without modifying the base model.

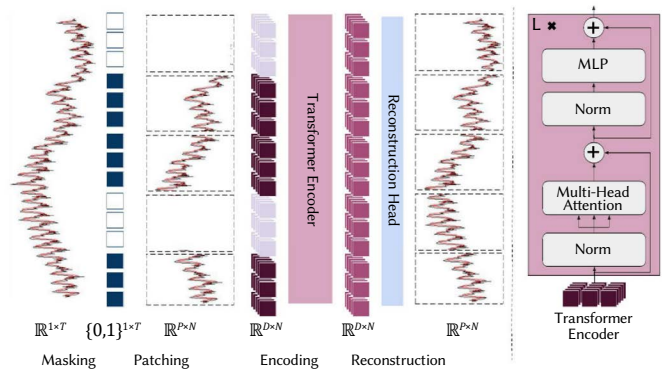


Fig. 3. Overview of MOMENT. A time series is broken into P disjoint fixed-length (N) patches. Each of them is mapped into a D -dimensional patch embedding. During pretraining, patches are uniformly masked at random by replacing their embeddings. The goal of pre-training is to learn patch embeddings which can be used to reconstruct the input time series using a light-weight reconstruction head. It is important to advise that, even though the image shows the univariate process, the model is defined for multivariate time series. Figure obtained from Fig. 3 of [10] under the terms of Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>).

TABLE I. MODEL SIZES AND ARCHITECTURES OF MOMENT PRE-TRAINED VARIANTS AND THE MODEL

Model	Layers	Hidden Dimension	Parameters (\approx)
MOMENT-1 Small	8	512	38 M
MOMENT-1 Base	12	768	113 M
MOMENT-1 Large	24	1024	346 M
DeepVATS-MTSAE	6	128	457K

Thus, MOMENT learns general representations of the time series and adapts to different tasks by adding a specific head for each task:

- **Reconstruction task.** Used for imputation and anomaly detection. This version is trained to reconstruct missing values within masked patches of the time series.
- **Embedding task.** Used for representation learning, clustering, and classification. In this case, MOMENT generates a fixed embedding vector for each time series, capturing its essential features.
- **Forecasting task.** Used to predict future values in time series. MOMENT learns representations and maps them to a forecast horizon.

- **Classification task.** Used to label time series based on learned patterns. MOMENT extracts representations and maps them to output classes.

Since the embedding task generates general latent representations of the time series, it is the most adequate for its integration into DeepVATS as it is not limited to a specific task and models the intrinsic characteristics of the time series. From now on the notation “MOMENT-small”, “MOMENT-base” and “MOMENT-large” is used to refer to the different sizes of the embedding task version of the model.

III. INTEGRATION OF MOMENT INTO DEEPVATS

In order to analyze the integration of MOMENT into DeepVATS in an equitable manner, we analyze and visualize latent spaces for most of the data sets analyzed in Rodriguez-Fernandez et al. [5] and Santamaria-Valenzuela et al.[14]: synthetic datasets generated by additive decomposition S1, S2, S3, the synthetic multivariate series M-Toy [38], and the Kohl’s dataset, based on real data [39]. Dataset S1 is used to analyze segmentation, S2 for anomaly detection, S3 for trends, M-Toy for anomaly detection, and Kohl’s for both segmentation and trend detection¹.

The goal is to check that MOMENT achieves at least the capabilities of MTSAE. This architecture has proven to have interpretable latent space projections for anomalies, patterns, and segments, lacking that interpretability in detecting trends. The addition of MOMENT to DeepVATS would provide more interactive and interpretable latent space plots, solving the question RQ1.

A. Zero-Shot Analysis of S1 Using MOMENT-Small

The S1 time series is designed to have four segments (df^c), providing a good example to test the segmentation task within the application. The execution of MOMENT for S1 in the same configuration as in the original research paper results in five very interlaced clusters, like a tangle. Four of them cover the whole series but the other one seems to be getting the whole series but not the df_2 segment (see Fig. 4). Using a zoom in all the clusters (see Fig. A13, Fig. A14, Fig. A15, Fig. A16, Fig. A17 in the appendix) we can detect that MOMENT is trying to get patterns, but they do not seem to be very clear. Also, the only segment that MOMENT can detect correctly is df_2 .

This fact leads to two questions. The first is: “Is the size of the window hiding important information from MOMENT as it tries to learn more specific information?”. And second, “Can I force MOMENT to learn specific information of the time series with few shots of the data?”. Based on the fact that foundation models work better when fine tuning is applied, we focus on the second one, analyzing the amount of dataset MOMENT needs to know to get better insights of it in the inference step.

B. Fine-Tune in DeepVATS

For fine tuning, a wrapper has been built so that any torch-based model can easily be included in deepVATS’s fine-tuning process. The fine-tuning is done like an imputation task training, in a similar way to training, thought for the imputation task: based on the percentage of the training dataset and masking part, use and validate the model for the reconstruction task using both a validation and train dataset.

The goal is to check if this reinforcement makes MOMENT fit the time series better and to check if the embedding space plot shows this enhancement. The data will be selected from parts of the time series to make sure we do not focus on a specific/special part of the time series.

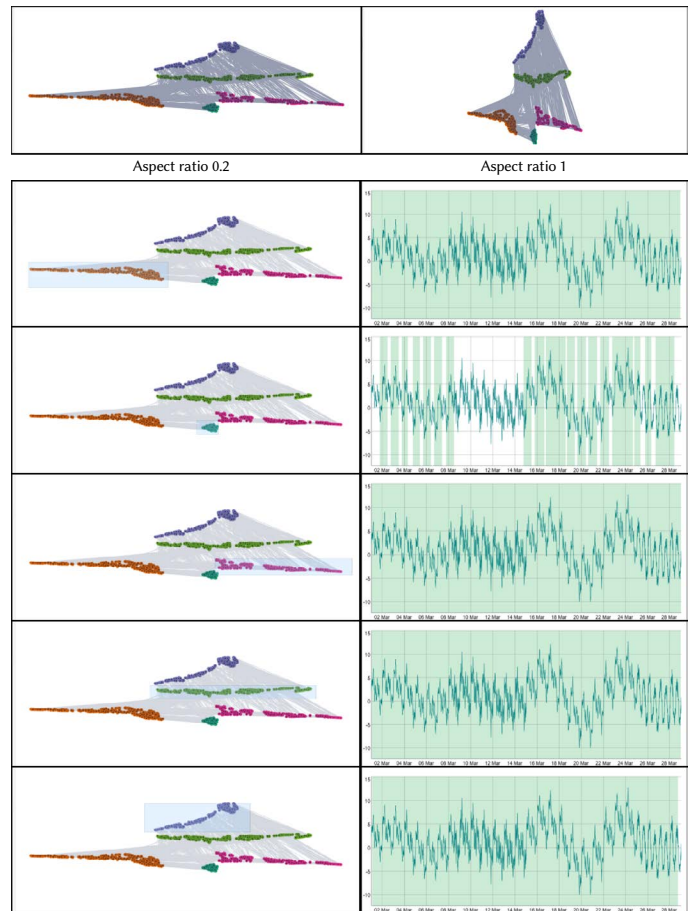


Fig. 4. Execution of MOMENT-small for S1 taking the mean in batches of 20 minutes. The execution is done for window length 54 and stride 2.

The fine-tuning algorithm is based on four main parameters:

- **window_lengths.** Hosts the window sizes that will be used for the batches in the fine-tuning.
- **training_percent.** This value is the percentage of the total percentage of the dataset that will be used to get the training batches.
- **valid_percent.** This value is the percentage of the total percentage of the dataset that will be used to obtain the validation batches.
- **mask_percent.** This value is the percentage that will be masked in each batch.

The selection of the window lengths and the position of the selected windows within the time series resulted in two different versions of the fine-tuning algorithms, modeled in the code based on a `mix_windows` flag:

- **mix_windows False.** This version loops through `window_lengths`, splitting the total dataset for the current `wlen` size and randomly selecting (within the `training_percent` and `valid_percent` percentages) windows across the time series.
- **mix_windows True.** This version splits the time series using the validation and training percentages. The validation and training datasets are built from the beginning of the time series. The first part will be the training dataset and the second part, the validation dataset. The batches are dynamically built getting a random window length from `window_length` in each batch. To ensure the integrity of the training and validation batches, the values of the lengths are cached so each validation is done within the same batches.

¹ The code necessary for replication as well as the full-sized screenshots are available at [github:misantamaria/deepvats-foundation](https://github.com/misantamaria/deepvats-foundation).

The first version did not show much improvement. It seemed to be skewed, as each window size was retraining the full model just after the previous one, possibly leading to catastrophic forgetting of the previous model at each iteration. Thus, we tested mixing the window sizes to get a more stochastic process, one that ensured a global learning of the time series for the model.

This rationale led to the second version, where we got improvements of more than 20% loss decrease. However, this second version has a difficulty in selecting random parts of the time series to make sure the model learns the global shape of it. The problem is that we cannot directly divide the time series correctly in different window sizes. Maybe we could fix it by taking the largest window for the division and later reducing the batch window size by getting the first or last values so that we can fuse both fine-tune modes.

C. Statistical Loss Improvement Analysis

To select the values for the main parameters of the fine-tune of the MOMENT models, a previous analysis using the Kohl's time series is done. For each version of the MOMENT embedding model (small, base, and large), the validation percentage is fixed to 0.3 and four parameters are analyzed.

The first parameter is the number of epochs for the training. The model is trained until 20 epochs, saving the model with a smaller loss in the training step as the final model. Once the training is complete, we fix the best epoch as the one where the optimal model was saved. This approach reduces the execution time in the visualization by selecting the most suitable number of epochs.

The second parameter is the dataset percentage, which is tested for 0.15, 0.2, 0.25, 0.3u. This percentage is the part of the dataset used as shot -training dataset- for the fine-tune of the different model.

The third parameter is the number of window lengths. The first window length is fixed to 17 as in the previous MTSAE experiment at Rodriguez-Fernandez et al. [5]. The remaining window lengths are selected using the Fourier transform result, which identifies the most relevant frequencies, as window sizes until we reach the total number of sizes desired (through the `find_dominant_window_sizes` function in `aeon.segmentation`) [40]. The experiment considers the following number of window sizes: {1, 2, 4, 6, 8, 10}.

The combination of these options (a total of 72 cases) leads to a sufficiently complete scenario for a preliminary selection of the fine-tuning parameters for the visual analysis based on the loss improvement. However, due to computational costs, the scenario is reduced, setting the number of epochs to 20, dataset percentages at {0.15, 0.2, 0.3}, masked percentages at {0.25, 0.5, 0.75}, and window lengths at {1, 5}, resulting in 18 different use cases.

This setup allows us to compare MOMENT not only against itself but also against MTSAE. However, this comparison is challenging and not entirely fair, as we are evaluating against a pre-trained MTSAE model (replicating the one used in [5], trained with window sizes from 12 to 17).

The results of the experiments show really similar losses between the different versions of MOMENT, with much more adaptability for MOMENT-small, followed by MOMENT-base, and then MOMENT-large (see Fig. A1, Fig. A2, Fig. A3).

The expectation is that MOMENT will get good interpretable visual embeddings for the same cases as MTSAE, being better as the number of parameters increases - or the loss is reduced. The rest of the statistical analysis allows us to select the best versions and see if any of them are preferable to the others.

To ensure the best comparison metric, we use the original loss function, taking into account the masked part within the computation.

For each execution, the model was validated before and after training, obtaining $loss_{first}$ from the previous evaluation and $loss_{final}$ from the post-evaluation (using the best version of the model). Thus, the loss improvement is defined as $\frac{(loss_{first} - loss_{final}) \cdot 100}{loss_{first}}$. The greater the improvement, the larger the reduction in loss. Therefore, this improvement will be used as the optimization objective in analyzing the best combination of parameters.

The experiment achieves improvements up to 22% in the small model, up to 10% in the base model, and up to 4% in the large model (see Fig. A4, Fig. A5, Fig. A6). This reduction is expected since each larger model contains significantly more parameters than the previous one, allowing it to fit the original time series more effectively. Consequently, achieving further improvement becomes increasingly difficult.

To check the parameters that are more related to the improvement in loss, correlations are checked. First, a linear correlation matrix is displayed to check the relationship between the analyzed values: `time`, `best_epoch`, `dataset_percent`, `masked_percent`, `n_windows` and `improvement`. Our expectation was to have a high positive correlation between the percentage of the dataset and the improvement, as well as the masked dataset and the improvement. However, the only constant correlation found is the time with the percentage of the dataset used for the training (see Table II, Table III, Table IV and Fig. A10, Fig. A11, Fig. A12). Thus, we continued to apply other techniques to check non-linear influences.

TABLE II. FEATURE IMPORTANCE SCORES FROM SELECTKBEST, RANDOM FOREST, SHAP IMPACT AND LINEAR CORRELATION FOR THE MOMENTSMALL MODEL

Feature	R. Forest %	KBest %	SHAP %	Corr. %
masked_percent	71.47	69.53	21.09	-79.21
best_epoch	21.34	28.89	10.51	64.16
dataset_percent	7.19	1.58	5.58	-19.17
n_windows	0.00	0.00	0.02	0.07

TABLE III. FEATURE IMPORTANCE SCORES FROM SELECTKBEST, RANDOM FOREST, SHAP IMPACT AND LINEAR CORRELATION FOR THE MOMENTBASE MODEL

Feature	R. Forest %	KBest %	SHAP %	Corr. %
masked_percent	51.22	63.81	18.64	-76.35
best_epoch	48.21	31.92	16.38	64.12
dataset_percent	0.53	4.27	0.67	29.24
n_windows	0.03	0.00	0.1	v

TABLE IV. FEATURE IMPORTANCE SCORES FROM SELECTKBEST, RANDOM FOREST AND LINEAR CORRELATION FOR THE MOMENT-LARGE MODEL

Feature	R. Forest %	KBest %	SHAP %	Corr. %
masked_percent	56.77	80.19	18.11	97.33
best_epoch	41.52	19.76	12.51	90.32
dataset_percent	1.39	0.05	1.21	-10.56
n_windows	0.33	0.001	0.2	-1.82

First, we used different DL-based algorithms to detect the most relevant features and to assess how much they were related to the loss reduction or increase. Second, we looked for the best parameters within the two most relevant parameters. Finally, for the best values of these parameters, the best combination of the other parameters is selected.

For the first step, we use `sklearn.feature_selection`'s `SelectKBest` and `f_regression` functions and `sklearn.ensemble`'s `RandomForestRegressor`. Also, the SHAP impact percent is used as another correlation metric that gives an idea of the "direction" of the influence (the larger the best or the smaller the best). The result is constant among the versions of MOMENT: `masked_percent > best_epoch >> dataset_percent >> n_windows`. The `masked_percent`

has a direct correlation with the loss improvement in the MOMENT-large model while maintaining an inverse correlation in the small and base versions. The best epoch index maintains a direct relation in the three cases (see Table II, Table III, Table IV).

Based on these results, the best combination of parameters is computed using the two more relevant parameters best improvement rows, obtaining the best combination as a start (see Table V).

TABLE V. BEST PARAMETER VALUES FOR SMALL, BASE, AND LARGE MODELS BASED ON FEATURE IMPORTANCE ANALYSIS

Parameter	Small	Base	Large
masked_percent	25	25	75
best_epoch	17	13	17
n_windows	1	5	1
dataset_percent	15	15	20

The next step is to compare MOMENT versus MTSAE. The main challenges arise when constructing the validation dataset:

MTSAE does not support batches with varying window sizes, and the masked percentage was predefined during training. To address the first issue, we use a window size of 17 in validation, since it is the only common size (fixed by hand) in all execution cases. For the second issue, we compute the MSE using the full original input and predictions instead of focusing only on the masked part. We then perform the MOMENT evaluation for the best model version in the same case, using the same functions and percentages as in MTSAE for the generation of masks, ensuring equivalent experiment.

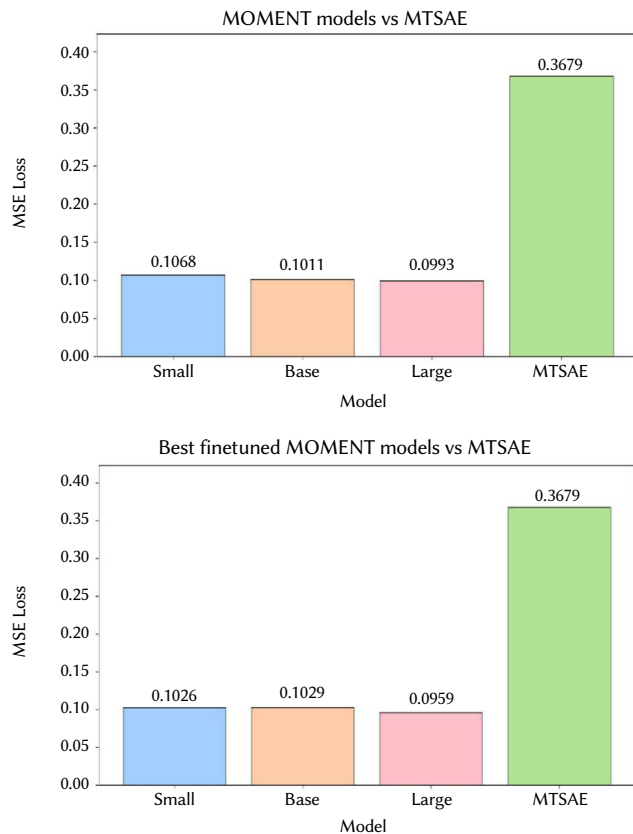


Fig. 5. Comparison of MOMENT models and the MTSAE model using the MSE loss comparing the full prediction to the original batch. At the top, the original version of MOMENT models. At the bottom, a re-training of the best cases.

The comparison of the best versions of MOMENT and MTSAE shows how increasing the number of trainable parameters does not necessarily imply an improvement in terms of MSE Loss. Also, MTSAE has a loss 70% higher than MOMENT models, regardless of the window length used for the inference (see Fig. 5, Fig. 6). Thus, the expectation is that MOMENT yields really interpretable visual embeddings for the same cases as MTSAE but does not result in too different plots between the three versions of MOMENT (once fine-tuned). The rest of the statistical analysis allows us to select the best versions and see if any of them is preferable to the others.

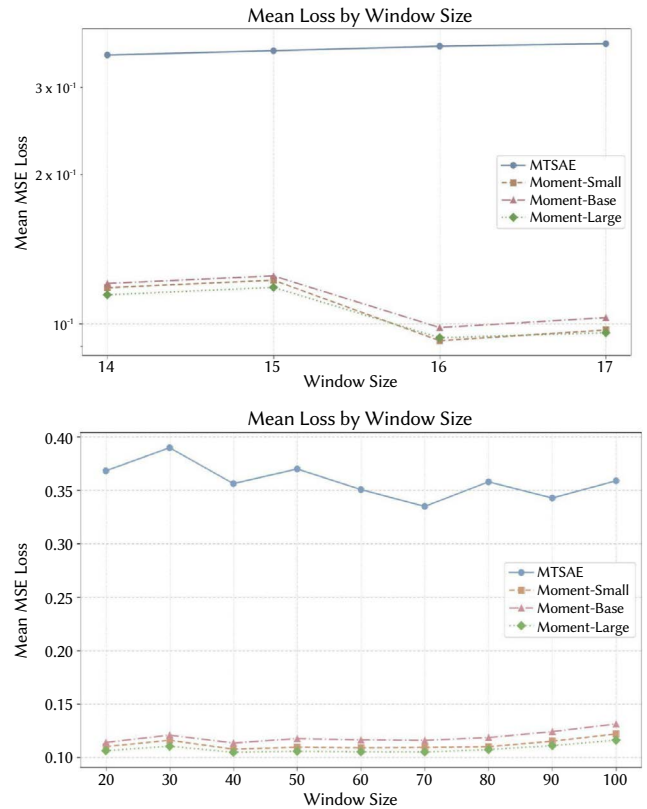


Fig. 6. Comparison of MOMENT models in their best version and the MTSAE model using the MSE loss comparing the full prediction to the original batch. At the top, using sizes in the range of those used in MTSAE training. At the bottom, using sizes in the range between 20 and 100 in steps of 10.

As the number of epochs may suppose a difference in obtaining the best model in any execution case, its frequencies have been computed to get the largest of the two more repeated values (or the first one if the difference was too big). This analysis resulted in the selection of 17 epochs for MOMENT-small, 13 for MOMENT-base and 20 for MOMENT-large (see Fig. A7, Fig. A8, Fig. A9).

To confirm that the number of window sizes seems irrelevant, a normalized Parallel Coordinates Plot is computed for the best improvement values. The plots in Fig. 7, Fig. 8, Fig. 9 show how, indeed, the value remains irrelevant. This conclusion is given by the sparsity of the best number of windows length value, which does not remain fixed to an option within the best executions. Thus, the best option on Table V is selected for each case.

Finally, based on the analysis conducted and looking for a good cost-to-improvement balance, the parameter described in Table VI where selected for the visual experimentation. In all cases, the percentage of the validation dataset was 30%. We focused on analyzing the best improvement scenario, but we did not analyze the relation between improvement and execution time - which would also be relevant for interactive applications.

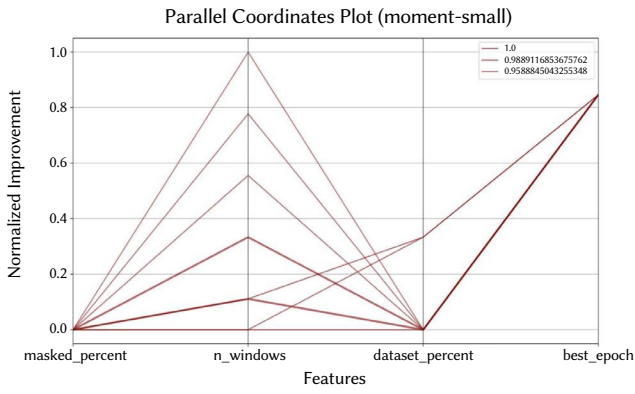


Fig. 7. Parallel coordinates plot matrices for MOMENT-small analysis.

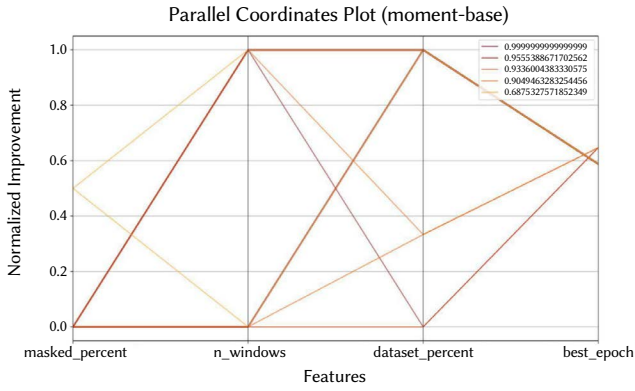


Fig. 8. Parallel coordinates plot matrices for MOMENT-base analysis.

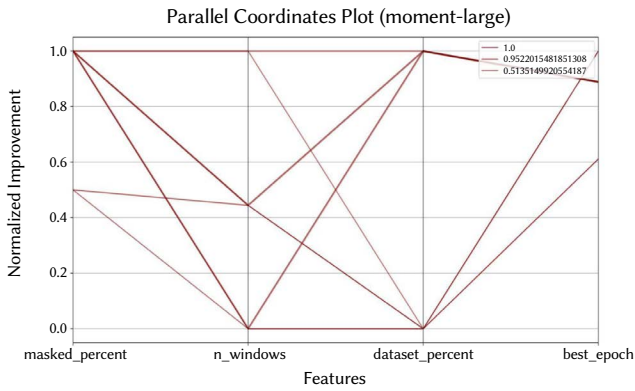


Fig. 9. Parallel coordinates plot matrices for MOMENT-large analysis.

TABLE VI. FINAL PARAMETER SELECTION FOR MOMENT-SMALL, MOMENT-BASE, AND MOMENT-LARGE MODELS FOR THE VISUAL EXPERIMENTATION

Parameter	Small	Base	Large
masked_percent	25	15	75
best_epoch	17	13	10
n_windows	1	5	1
dataset_percent	15	25	20

D. Analysis of S1 Using Fine-Tuned MOMENT-Small

Fig. 10 shows how MOMENT-small fine-tuning does not represent a visual difference with the zero-shot plot, resulting in the same four clusters shaped as a tree and (somehow) with the only segmentation of the segment $df2$. This is strange as TSFMs are supposed to get a really good adaptation to the new time series when fine-tuned. The

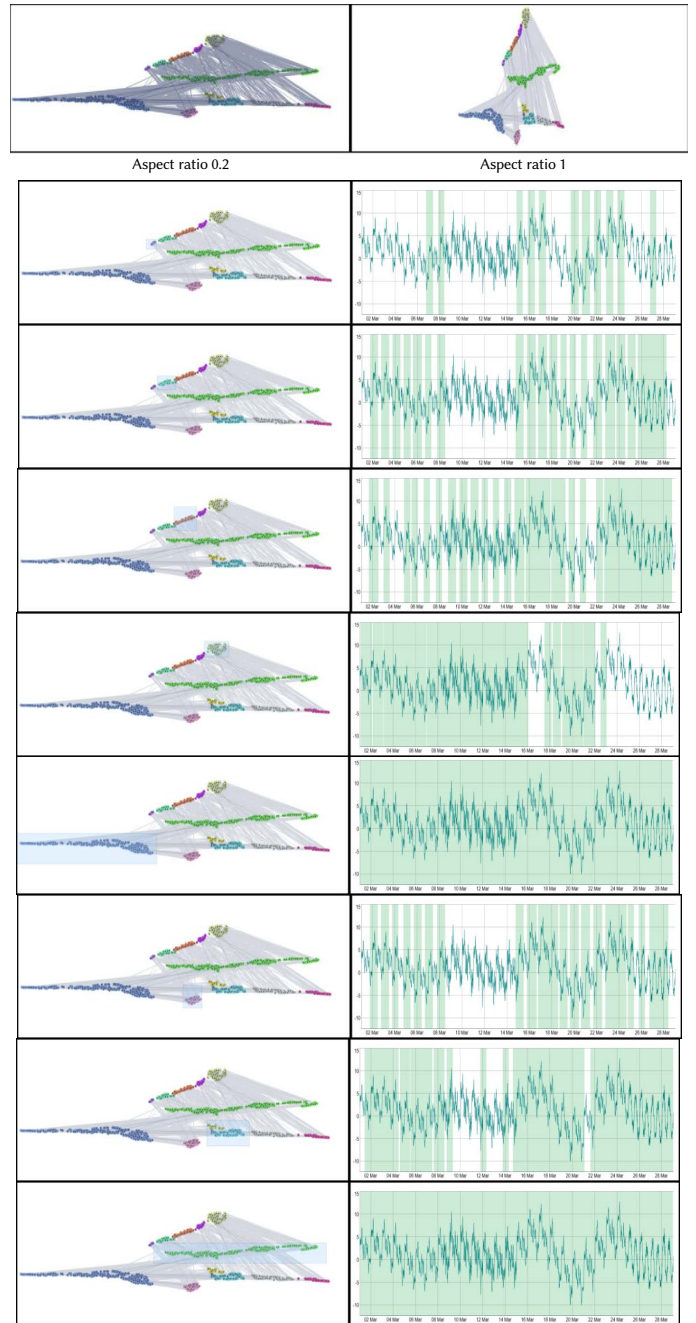


Fig. 10. Analysis of the embeddings projection plot of the fine-tuned version of MOMENT-small for S1.

loss improvement evidences this fact. However, the visual plot of the embedding space does not seem to capture this enhancement. Thus, related to RQ1, this data set does not appear to get a more visually interpretable embedding space when the loss decreases.

To better answer this question, the analysis is performed for the different tasks in DeepVATS: for segmentation, S1 (with the rest of the model versions); for anomaly detection, S2 and M-Toy; and for trend detection, S3.

E. Analysis of S1 Using MOMENT-Base

The plot of the MOMENT-base-zero-shot and the MOMENT-base-fine-tuned models are practically the same (see Fig. 11), showing no visual enhancement within the projection plot.

The embedding space is divided into three clusters. The first one seems to detect small patterns, not showing a clear classification of

TABLE VII. ADVANTAGES AND DISADVANTAGES OF USING MOMENT-SMALL, MOMENT-BASE, AND MOMENT-LARGE IN ZERO-SHOT AND FINE-TUNED CONFIGURATIONS FOR S1 DATASET (SEGMENTATION)

Model	Training	Observations
MOMENT-Small	zero-shot	<ul style="list-style-type: none"> ✓ Captures general patterns in the data. ✓ Requires no additional training, reducing computational cost. ✗ Produces poorly defined embeddings (“blurry” clusters). ✗ Struggles to differentiate time series segments effectively. However, it starts to detect $df1$, $df2$ and $df3$.
	Fine-tuned	<ul style="list-style-type: none"> ✓ Moderate loss reduction (~22%), improving embedding quality. ✓ Some improvement in pattern detection. ✗ Visual embedding projection shows minimal improvement over zero-shot. ✗ Cluster structure remains unclear despite fine-tuning.
MOMENT-Base	zero-shot	<ul style="list-style-type: none"> ✓ Clearly detects $df2$ and is closer to detecting $df1$ and $df3$ than Small. ✓ Captures some patterns of the time series. ✗ Still lacks precise segmentation of time series into clusters. We have even lost the precision with $df1$ and $df2$.
	Fine-tuned	<ul style="list-style-type: none"> ✓ Detects patterns. ✗ Fine-tuning provides only a marginal improvement. ✗ Cluster structure still lacks well-defined segmentation.
MOMENT-Large	zero-shot	<ul style="list-style-type: none"> ✓ Detects patterns. ✗ May be influenced by the time series’ noise. ✗ Clusters are not very separated. ✗ Segments are not clear.
	Fine-tuned	<ul style="list-style-type: none"> ✗ No special advantages. ✗ Fine-tuned embeddings are still highly similar to zero-shot, not getting any different result within the analysis.

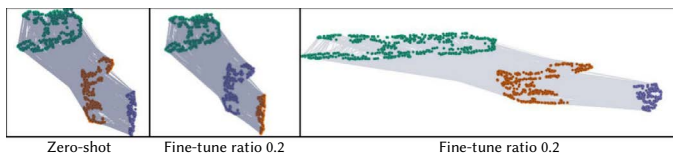


Fig. 11. Embeddings projection plot of the zero-shot and fine-tuned versions of the MOMENT-base model for S1.

the time series’ segments (see Fig. A18). The second cluster makes a really clear segmentation between $df2$ and the rest of the time series and starts to detect $df4$. However, it is not still as clear as it should be, and the split is not done within the 4 segments, so the model and embedding projection do not capture the full information we need (see Fig. A19). However, the third cluster starts to detect four segments really well, but mixed with other patterns (see Fig. A20).

At this point, it seems that the base version of the model is capturing much better the segmentation information of S1.

F. Analysis of S1 Using MOMENT-Large

The plots of the embeddings for S1 using the large model in both fine-tuned and zero-shot configurations exhibit a high degree of similarity (see Fig. 12, Fig. 13). Therefore, the analysis will focus primarily on the fine-tuned model, as it is expected to provide a more complete and interpretable representation.

The embedding space initially consists of three main clusters, resembling those observed in the base model embeddings (see Fig. 12, Fig. 11). However, these clusters appear less distinct and are positioned closer to each other.

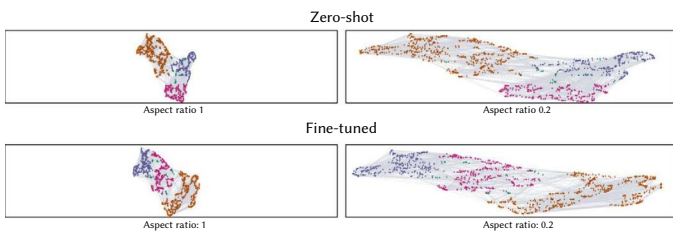


Fig. 12. Global view of the embeddings of the zero-shot and the fine-tuned large models for S1.

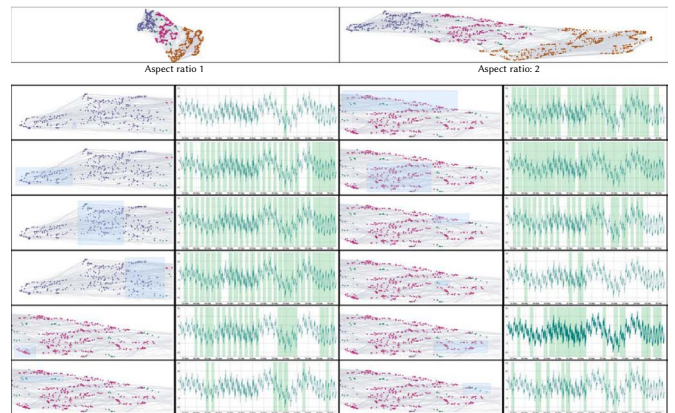


Fig. 13. Analysis of S1 using the fine-tuned large model.

Table VII summarizes the analysis for S1. The use of larger versions of MOMENT seems to be a good idea if the computation resources are good enough as the time series are better matched. However, the fine-tuning has not shown large differences with the zero-shot version of the models. In a future analysis for segmentation, some preprocessing should be done within the training dataset, maybe reducing the noise through Simple Moving Average (SMA). Also, we should try other dimensionality reduction methods to check if they end up with more time-ordered clusters. Related to the non-enhancement of the fine-tune for the visual analysis, we should check if there is a better distance for the segmentation when optimizing the loss. Also we could try to fine-tune the model with much larger window lengths and then using it to model the original length to check if it gives a long-term relationship. Training with larger SMA orders should also help in analyzing the global structure of the time series.

G. Analysis of S2

The S2 dataset contains two point anomalies. In Fig. 14 both of them are detected in the moment-small’s embedding space, but the interrelations between clusters make it difficult to detect them in a single view, losing interpretability.

If we focus on the left part of the time series and the right one, we observe anomalies in other parts of the embedding space at points that are not joined by time but short shapes. It seems that temporal

TABLE VIII. COMPARISON OF ZERO-SHOT VS. FINE-TUNE FOR S2 (ANOMALY DETECTION) USING MOMENT-SMALL, MOMENT-BASE, AND MOMENT-LARGE

Model	Training Type	Observations
MOMENT-Small	Zero-shot	<ul style="list-style-type: none"> ✓ Clearly defined clusters. ✓ Captures general patterns in S2. ✗ Clusters are highly intertwined. ✗ Anomalies detected but hard to isolate.
	Fine-tuned	<ul style="list-style-type: none"> ✓ No significant improvement. ✗ Anomalies remain hard to isolate. ✗ Cluster structure unchanged.
MOMENT-Base	Zero-shot	<ul style="list-style-type: none"> ✓ One anomaly detected at a cluster edge. ✗ The anomalies appear in different clusters at the same time.
	Fine-tuned	<ul style="list-style-type: none"> ✓ Minor improvement in anomaly detection. Clusters rotated but similar.
MOMENT-Large	Zero-shot	<ul style="list-style-type: none"> ✓ Most defined clusters. ✓ Anomalies detected at cluster edges. ✗ Anomalies still appear in multiple clusters.
	Fine-tuned	<ul style="list-style-type: none"> ✗ No significant improvement. ✗ Cluster structure remains nearly identical. ✗ The anomalies move to the middle of a cluster, making them more difficult to detect in a visual inspection.

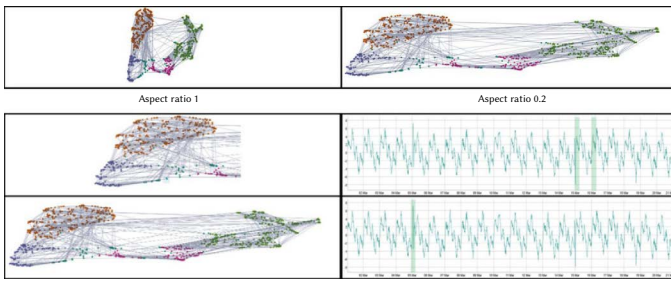


Fig. 14. Global view of the embeddings of the zero-shot small model for S2.

correlation is being lost and each window is treated like a word in a sentence, making it difficult to detect large parts, so anomalies are easy to check (see Fig. A21). In this part, we can find both anomalies. Also, if we focus on the right part of the time series, we see more patterns and also the presence of both anomalies of the time series (see Fig. A22).

After fine tuning moment-small, we can see that the embedding space does not change so much, with no improvement of its interpretability (see Fig. 15).

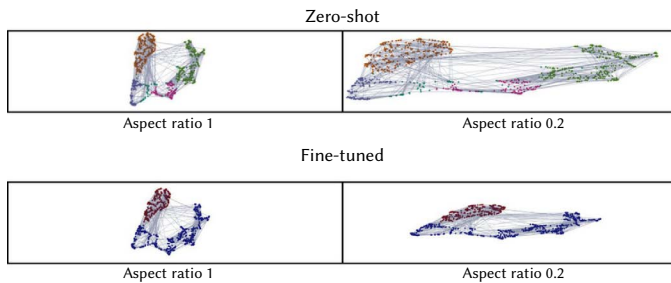


Fig. 15. Analysis of the projection plot for the moment-small fine-tuned version of the model for S2.

The MOMENT-base model results on more defined clusters (see Fig. 16, Fig. 17). The difference between the zero-shot and the fine-tuned models is still not that much, as the only change is the rotation of positions between the brown and green clusters in the right part. Again, MOMENT seems to correctly detect small patterns but it makes it difficult to check the anomalies as they are not in the appart zones but, at least, the anomaly by the right is at a corner of the pink cluster in an edge point (very similar “input-output” directions, like if they were the same line). This makes it easier to detect the anomalies, but not as easy as with the embeddings of MTSAE.

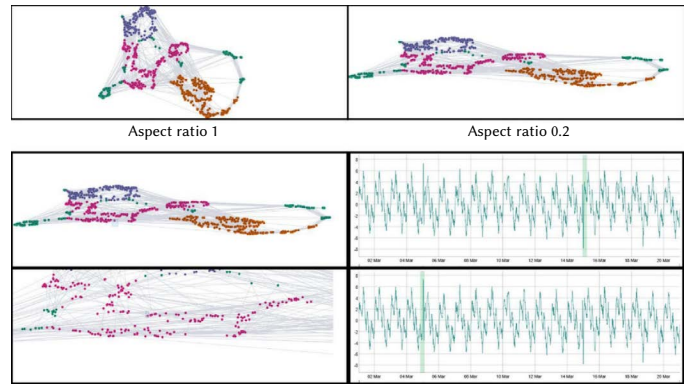


Fig. 16. Global view of the embeddings projections of the zero-shot version of MOMENT-base applied to S2.

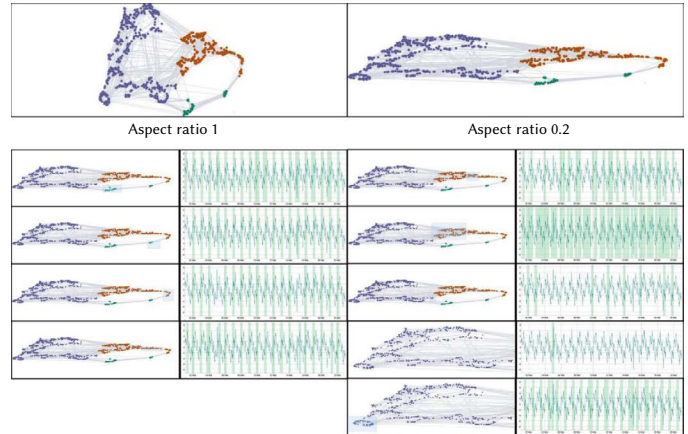


Fig. 17. Embeddings projection plot of the fine-tuned version of MOMENT-base applied to S2.

The embedding space of the MOMENT-large model is similar to the previous ones but seems to be more accurate in the right part with the new purple cluster (see Fig. 18). Focusing on the pink cluster we can see something interesting for the first time, both anomalies can be detected in its upper corner, which is a great advance in order to directly check the anomalies (see Fig. A23).

Fig. A24, Fig. A25, Fig. A26 and Fig. A27 show the insider view of the clusters. It is interesting that again the anomalies appear in different clusters at the same time (Fig. A23, Fig. A24).

TABLE IX. COMPARISON OF ZERO-SHOT VS. FINE-TUNED FOR S3 (TRENDS) USING MOMENT-SMALL, MOMENT-BASE, AND MOMENT-LARGE

Model	Training Type	Observations
MOMENT-Small	Zero-shot	<ul style="list-style-type: none"> ✓ Defined but interconnected clusters. ✗ No clear trend detected, with a spring effect.
	Fine-tuned	<ul style="list-style-type: none"> ✗ No significant changes. ✗ Spring effect remains in cluster visualization.
MOMENT-Base	Zero-shot	<ul style="list-style-type: none"> ✓ Some structures suggest trend detection. ✗ Still no fully defined trends.
	Fine-tuned	<ul style="list-style-type: none"> ✓ Slightly clearer structure. ✗ Trend still mixed within clusters.
MOMENT-Large	Zero-shot	<ul style="list-style-type: none"> ✓ Best model for detecting trend-like structures. ✗ Spring effect still present.
	Fine-tuned	<ul style="list-style-type: none"> ✗ No significant differences with the zero-shot.

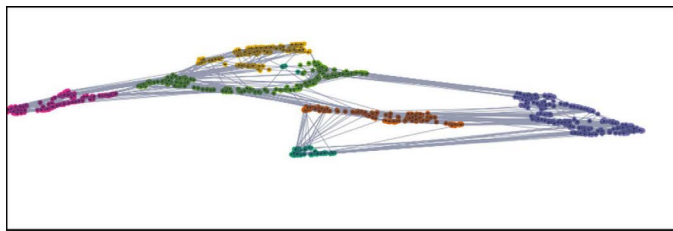


Fig. 18. Global view of the embeddings projections of the zero-shot version of MOMENT-large applied to S2.

The fine-tuned version has a very similar shape. In this case, the anomalies are in the middle of the pink cluster (Fig. 19), difficult to detect.

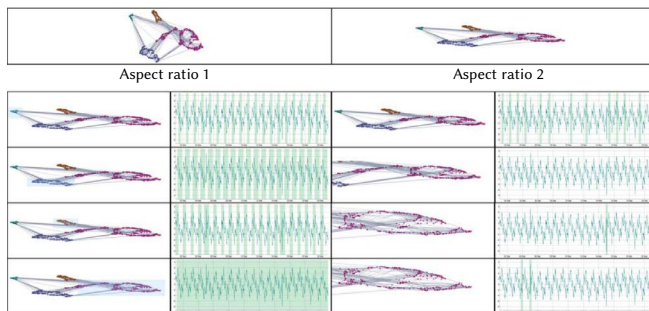


Fig. 19. Embeddings projections of the fine-tuned version of MOMENT-large applied to S2.

Thus, in summary, anomalies are not that easy to detect. Table VIII shows an overview of the analysis for each version of the dataset.

H. Analysis of S3 Using MOMENT

This dataset is used to check the interpretation of trends through the visual projection of the embedding space. Table IX shows an overview of the analysis for each version of the dataset. The zero-shot and fine-tuned versions of MOMENT-small have very similar projections, with small rotations. Taking a loop from left to right to the different clusters, no trend is detected, showing a spring effect (see Fig. 20).

The base projections are very similar to the small ones in both the zero-shot and the fine-tuned version. Although it seems to be getting the trend when selecting the two clusters as a full object, getting inside the blue part and going through the lines, we still have the same spring effect, showing that the trend is not related to the position of the points in the embedding space (see Fig. A28, Fig. A29).

The large version also shows really few differences between the zero-shot and fine-tuned embedding projections. This time, the trend seems to be detected, but it still has a spring effect in the blue cluster (see Fig. A30).

I. Analysis of Kohl's Time Series Using MOMENT

The Kohl's time series is used for analyzing both segmentation and trend. The model is supposed to fit both the upper trend and the segmentation between peaks and plains.

The MOMENT-small-zero-shot versions show two distinct clusters, however. The green one (in the bottom) is near to detect all the peaks - valleys - (with some temporal shift to the next timesteps)) and the first plain zone. Those peaks seems to be detected if using the other two small green clusters. However, when we select the rest of the embedding space, we see that the segmentation is not that clear as the blue part contains the full time series. Also, the "linear" pattern of the trend within the embedding space is not detected. The fine-tuned version has more defined clusters. The "plains" are nearly learned, in the green part. However, some parts are shared between all the clusters, like the peaks and the first plain. It is easier to detect, but there is no clear trend or segmentation pattern in the plot (see Fig. 21). The MOMENT-base model seems to be clearly detecting the peaks but not the plains, correctly detecting them as a pattern more than a segmentation. In fact, the blue clusters gets the full time series again instead of just the plain parts. Again, no trend visual pattern is found (see Fig. A33). For the MOMENT-large version, see that the zero-shot and the fine-tuned model are really similar, with a rotation in the right part (and a subtle more definition). This time, the fine-tune and the zero-shot models are more different (see Figs. A34, A35). The zero-shot model shows clear segmentation between peaks and plains (except for the first plain), showing a great performance in detecting the plains while detecting both patterns and intermediates. The fine-tuned version seemed to be distinct, but the results are really similar (see Table X).

J. Analysis of Toy Time Series

The M-Toy dataset is proposed for analyzing sequence anomalies in multivariate time series. The small versions can detect the two anomalies of the dataset. In both the fine-tuned and the zero-shot version, the anomaly by the left is really easy to check as an edge point, but the right anomaly is detected in the middle of a cluster, making it more likely a transition point and thus difficulting the detection (see Fig. 22). The base-zero-shot version is more difficult to analyze. However, when applying the fine-tune, the anomalies get easier to detect and get really near between them, showing a relation between both patterns (see Fig. A31). The large zero-shot version is more difficult to read, especially for the anomaly by the left, as it gets mixed with intermediate patterns (which may also be interesting). The large-finetune version results with similar plot with a better interpretability of the anomaly on the right. However, the anomaly by the left is even more difficult to detect within the embedding space.

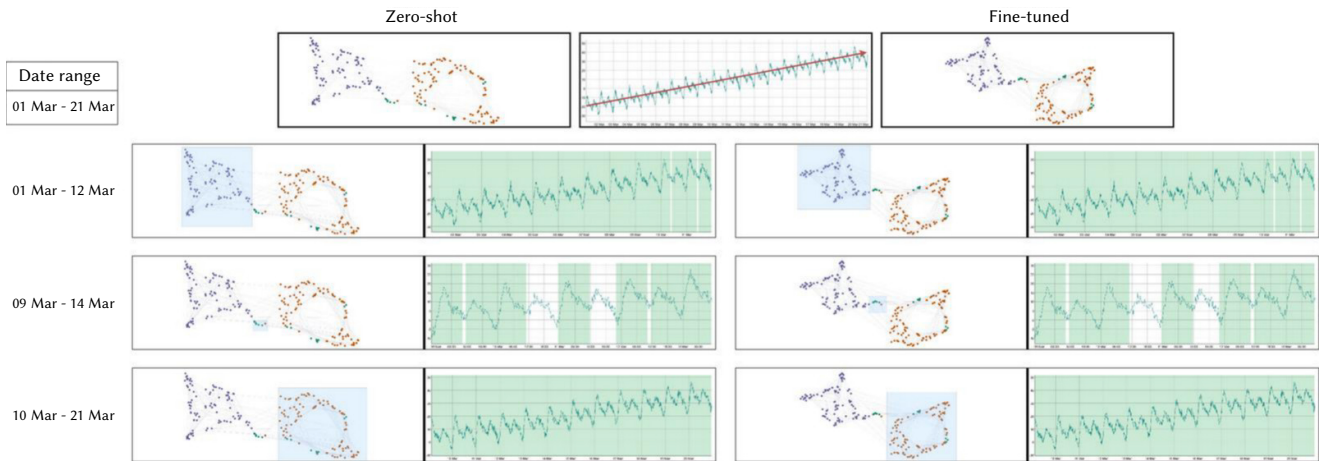


Fig. 20. Embeddings projections of MOMENT-small applied to S3.

TABLE X. COMPARISON OF ZERO-SHOT VS. FINE-TUNED FOR KOHL'S USING MOMENT-SMALL, MOMENT-BASE, AND MOMENT-LARGE

Model	Training Type	Observations
MOMENT-Small	Zero-shot	<ul style="list-style-type: none"> ✓ Clear separation of peaks. ✗ Trend not fully detected. ✗ First plain is mixed within clusters. ✗ Not clear separation of the plains.
	Fine-tuned	<ul style="list-style-type: none"> ✓ More defined clusters. ✓ Plains are closer to being detected. ✗ First plain shared between all the clusters. ✗ Plain section still containing peaks. ✗ Still no clear trend detection.
MOMENT-Base	Zero-shot	<ul style="list-style-type: none"> ✓ Peaks are well defined. ✗ Plains are not clearly detected.
	Fine-tuned	<ul style="list-style-type: none"> ✓ Slight improvement in segmentation. ✗ No clear improvement in trend detection.
MOMENT-Large	Zero-shot	<ul style="list-style-type: none"> ✓ Best model for segmentation. ✓ Peaks and plains clusters are visually distinct. ✗ Trend is still not fully learned.
	Fine-tuned	<ul style="list-style-type: none"> ✗ No major improvement. ✗ Trend detection remains similar to zero-shot.

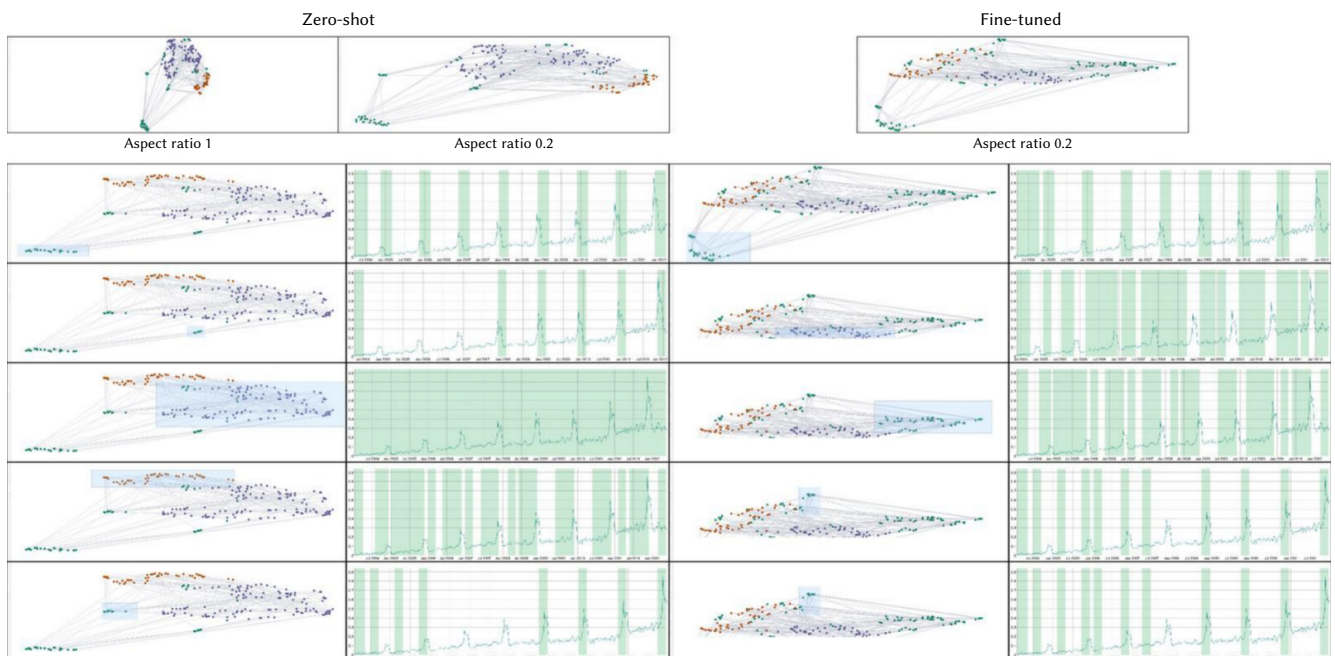


Fig. 21. Embeddings projections of MOMENT-small applied to Kohls.

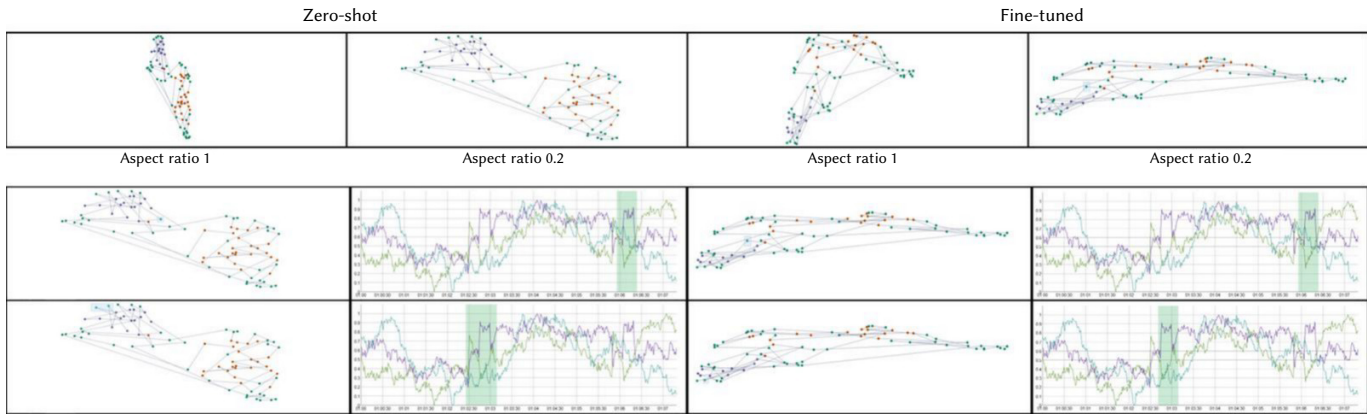


Fig. 22. Embeddings projections of MOMENT-small applied to M-Toy.

TABLE XI. SUMMARY OF MODEL PERFORMANCE ACROSS DIFFERENT TASKS (ANOMALY DETECTION, PATTERN DETECTION, SEGMENTATION, TREND DETECTION)

Model	Anomaly Detection	Pattern Detection	Segmentation	Trend Detection
MOMENT-Small	✗ Anomalies detected but difficult to isolate.	✓ Some patterns detected.	✗ No segmentation.	✗ No visible trend detection.
MOMENT-Base	✓ Minor improvements in detecting anomalies at edges.	✓ Patterns are clearer than using Small.	✗ Some, but incomplete.	✗ Trend detection is suggested but unclear.
MOMENT-Large	✓ Anomalies more difficult to detect than with base.	✓ Most structured pattern detection.	✗ No clear segments.	✗ Still lacks clear trend separation.

IV. CONCLUSIONS AND FUTURE WORK

Table XI shows how MOMENT's embeddings are not that easy to check visually in order to detect the timeseries characteristics. The best performance is obtained for short pattern detection. Also, the clusters are really interconnected, showing that the time correlation may not be correctly detected.

Focusing on the research questions RQ1 and RQ2, the fine-tuned loss performance is directly related to the percentage of the dataset used. However, after selecting the best cases within the statistical analysis, the model is not likely showing a great difference within the UMAP followed by PCA projections. Thus, it seems that the improvement in terms of loss is definitely not related to the fine-tune level (in fine-tune percentages), but we open the question whether MOMENT is able to detect the same patterns than MTSAE if trained with the full training dataset. However, this is against the idea of having a faster interaction due to the less need of training (still, we could check whether a full training of MOMENT in any of its versions is faster than MTSAE or not).

Looking at the performance of the fine-tuned versions of MOMENT selected in the Kohl's dataset, the improvements do not fully compensate for the extra time spent tuning. MOMENT-small was notably faster (approximately 16% faster than MOMENT-base or MOMENT-large). It also extracted embeddings faster: 0.15 seconds versus 0.21 and 0.39, respectively. This difference is very small, so it does not influence the final decision. To put it in perspective, fine-tuning required 13000% more time than embedding extraction for the Small model. That difference is less with larger models, but still huge. So, if fine-tuning is needed, MOMENT-Small gives you more for the time required. Otherwise, MOMENT-base may be the middle ground. It would be wiser to consider the proposed ideas as future lines rather than focusing on model selection.

In conclusion, the analysis suggests that a loss improvement for a moment model version is not directly related to the embeddings precision, so the small version should show good results in future analysis, resulting in lower time and memory cost within the analysis. Thus, the integration of foundation models with few parameters into visual analytics tools is a great idea to save waiting time. Thus, the

integration of MOMENT into deepVATS - with the option of changing it to any other foundation model - is a great contribution that offers a flexible and efficient way for visually analyzing large time series data.

To refine this fine-tune approach, different paths for improvement are identified: exploring alternative loss functions - such as soft-DTW [41] - could produce more refined embeddings, while testing different projection methods (e.g. PCA without UMAP and t-SNE) would help determine the impact of the chosen technique on the results. In addition, incorporating an option to preprocess the data prior to the fine-tune could result in clearer visualizations, and analyzing the effect of fine-tuning on each model later, including selective freezing of certain layers, could improve the embedding space projection interpretability. Either of these options would represent a viable strategy to further improve the application.

APPENDIX

A. Statistical Analysis Auxiliar Plots

This section contains the auxiliar plots for the statistical analysis section.

First, the loss comparison (before and after the fine-tuning) is shown in Fig. A1, Fig. A2, Fig. A3 using MSE as loss function. This plots allow to get an idea of the evolution of the loss within the execution cases.

Second, the loss improvement per analyzed case is shown in Fig. A4, Fig. A5, Fig. A6. These figures offer a better view in terms of percentage on how much we can improve the loss within the trainings, showing how the small changes on the previous plot really make a significant difference.

Third, the frequency of the best epoch within the fine-tune loop for each version of the model is shown in Fig. A7, Fig. A8, Fig. A9). This plot aids in the selection of a logical value for the number of epochs to ensure the best improvements with the lowest computation and time cost. This plot results in orange the two more repeated values, giving preference to the largest ones as they are the ones that give more assurance on getting the best model for each case configuration.

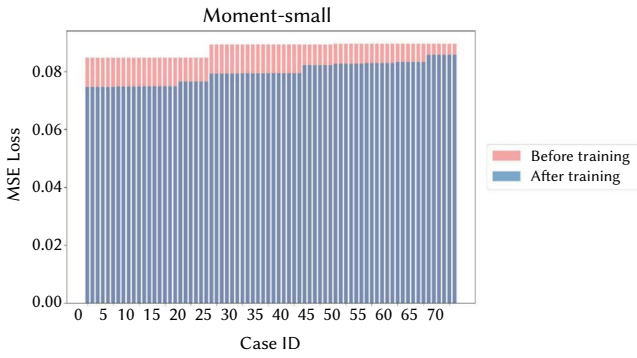


Fig. A1. MOMENT-Small. MSE Losses for all the cases in the experimentation. The red line is used to show the MSE Loss for the original model used for Kohl's series. Each bar represents the MSE Loss before (red) and after (blue) the fine-tuning.

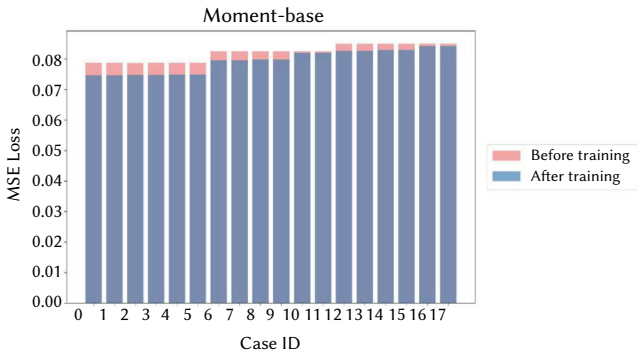


Fig. A2. MOMENT-Base. MSE Losses for all the cases in the experimentation. The red line is used to show the MSE Loss for the original model used for Kohl's series. Each bar represents the MSE Loss before (red) and after (blue) the fine-tuning.

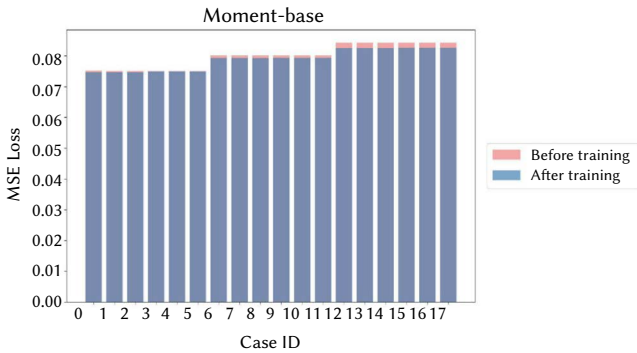


Fig. A3. MOMENT-Large. MSE Losses for all the cases in the experimentation. The red line is used to show the MSE Loss for the original model used for Kohl's series. Each bar represents the MSE Loss before (red) and after (blue) the fine-tuning.

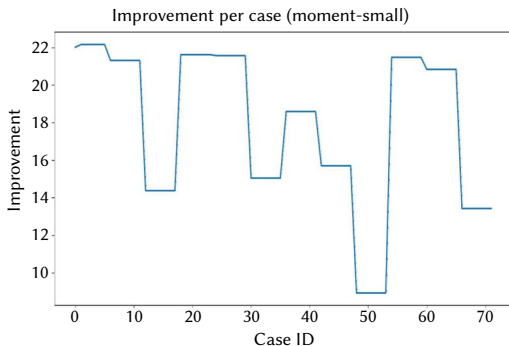


Fig. A4. Loss improvement for MOMENT-small across the experimentation.

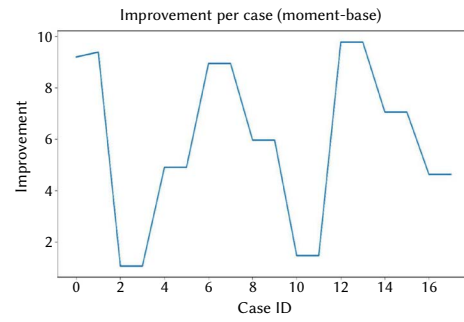


Fig. A5. Loss improvement for MOMENT-base across the experimentation.

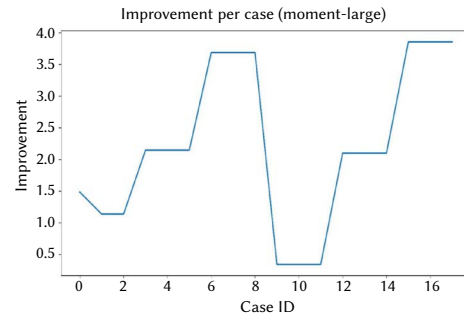


Fig. A6. Loss improvement for MOMENT-large across the experimentation.

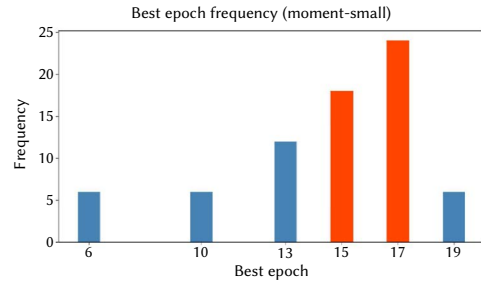


Fig. A7. Histogram showing the frequency of the best epoch (MOMENT-small). The two most frequently occurring best epochs are highlighted in orange.

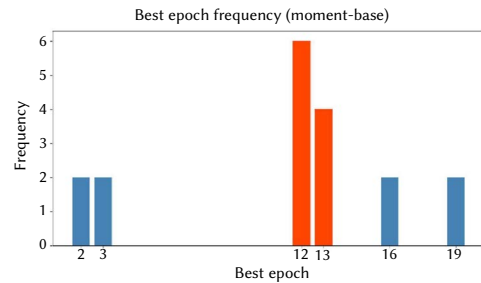


Fig. A8. Histogram showing the frequency of the best epoch (MOMENT-base). The two most frequently occurring best epochs are highlighted in orange.

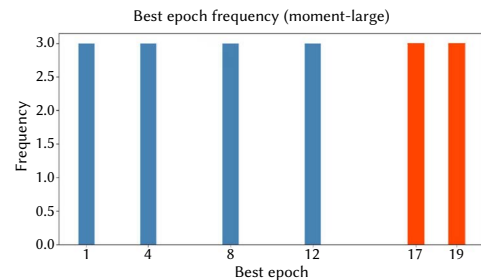


Fig. A9. Histogram showing the frequency of the best epoch (MOMENT-large). The two most frequently occurring best epochs are highlighted in orange.

Fourth, starting to interconnect all the parameters, the linear correlation matrices between them, and the loss improvement are shown in Fig. A10, Fig. A11, Fig. A12 for the three versions of MOMENT using tasks.

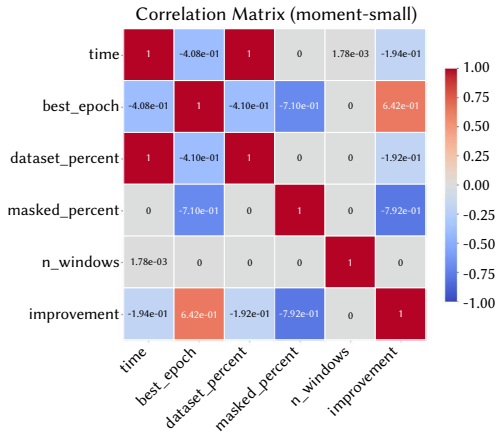


Fig. A10. Experimentation parametres correlation matrix for MOMENT-small.

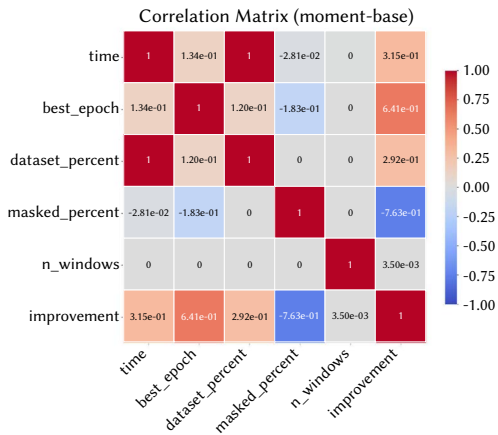


Fig. A11. Experimentation parametres correlation matrix for MOMENT-base.

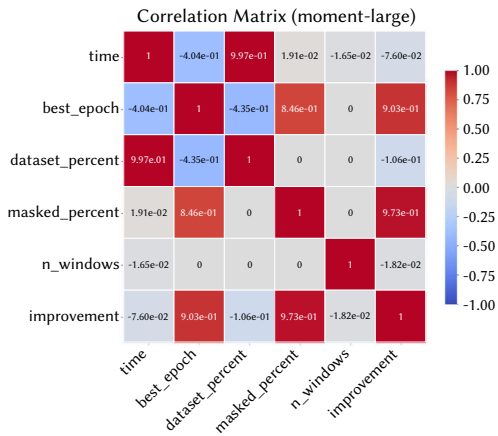


Fig. A12. Experimentation parametres correlation matrix for MOMENT-large.

B. Interactive Exploration Auxiliar Plot

This section includes auxiliary plots that support the interactive analysis of the datasets.

1. S1 Auxiliar Plots

First, the clusters for S1 analysis, zooming in on each of them to look for the different segments in the time series.

Fig. A13, Fig. A14, Fig. A15, Fig. A16, and Fig. A17 show the visualizations for the different clusters of the embedding space when executing MOMENT-small for S1.

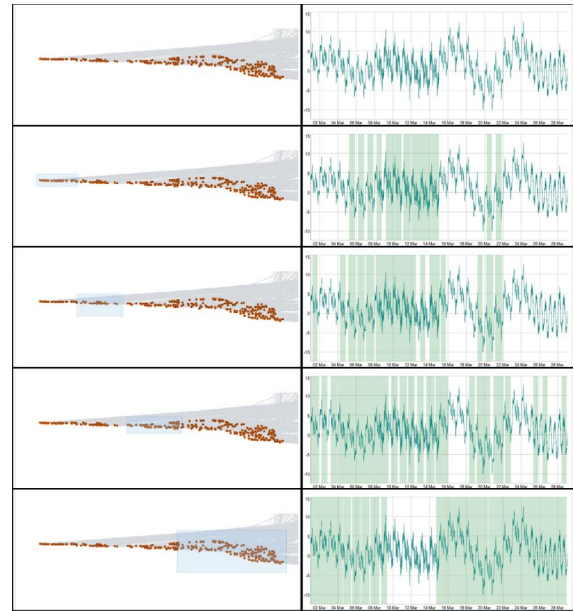


Fig. A13. Cluster I. Execution of MOMENT-small for S1.

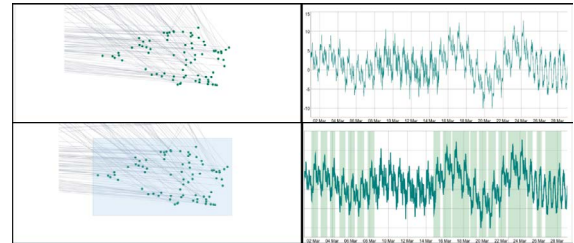


Fig. A14. Cluster II. Execution of MOMENT-small for S1.

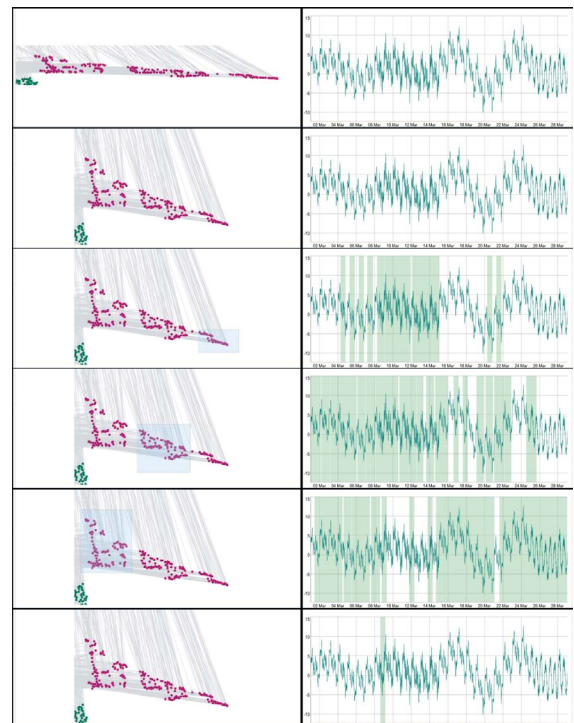


Fig. A15. Cluster III. Execution of MOMENT-small for S1.

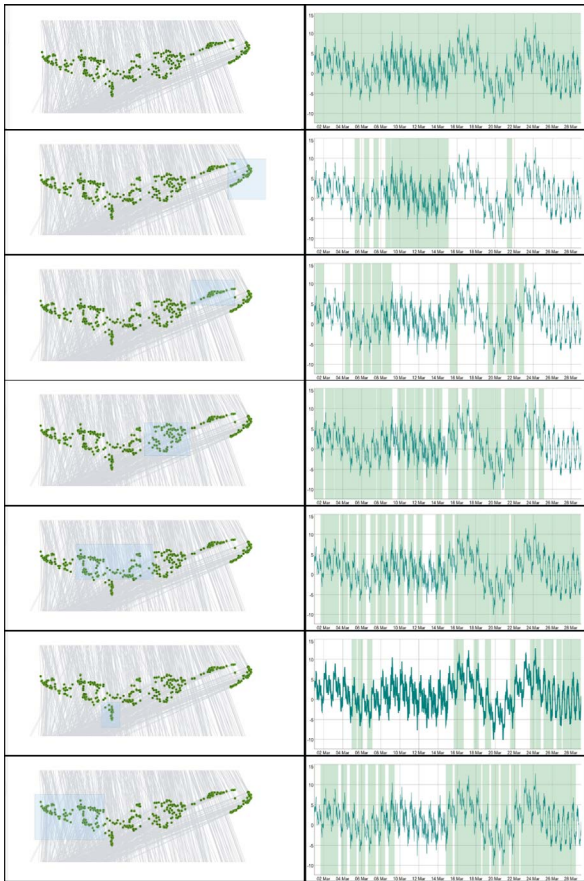


Fig. A16. Cluster IV. Execution of MOMENT-small for S1.

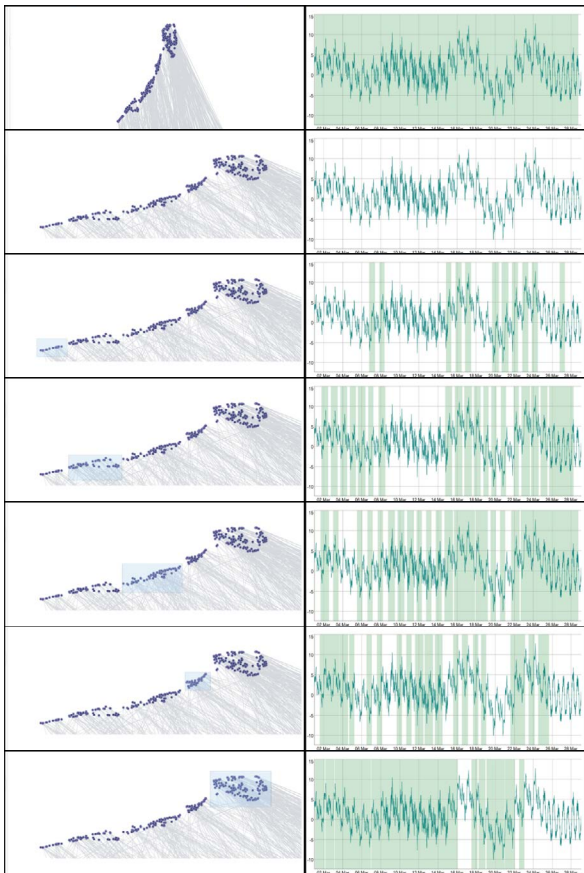


Fig. A17. Cluster V. Execution of MOMENT-small for S1.

Fig. A18, Fig. A19, Fig. A20, show the visualizations for the different clusters of the embedding space when executing MOMENT-small for S1.

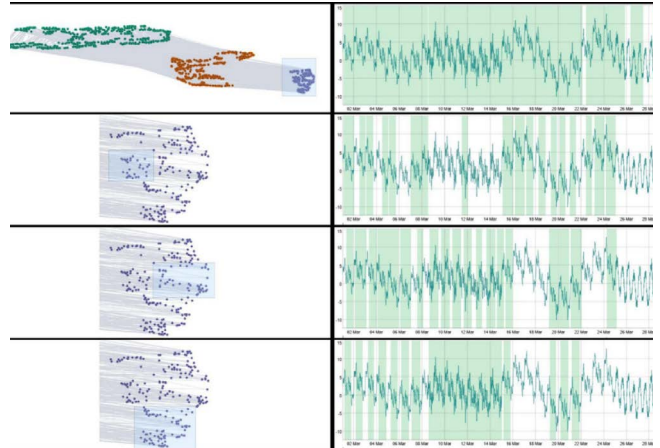


Fig. A18. Cluster I: execution of MOMENT-base for S1.

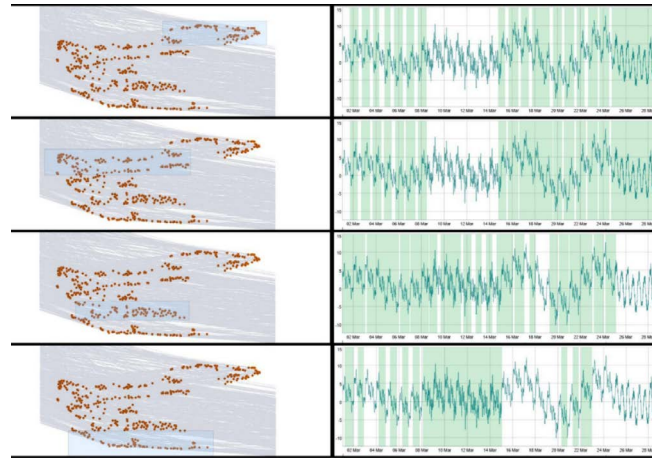


Fig. A19. Cluster II: execution of MOMENT-base for S1.

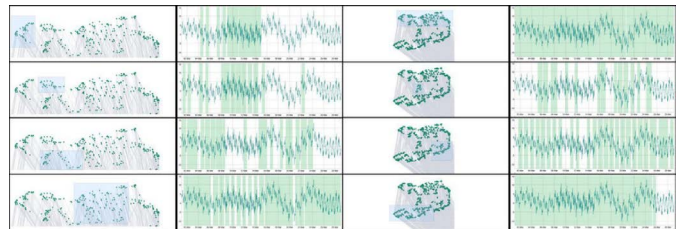


Fig. A20. Cluster III: execution of MOMENT-base for S1.

2. S2 Auxiliar Plots

Fig. A21 and Fig. A22 show the two clusters of the execution of MOMENT-small in the zero-shot version.

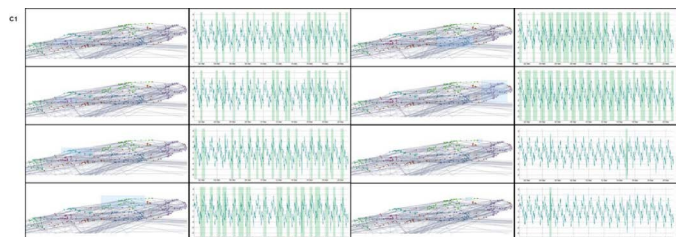


Fig. A21. Cluster I. Analysis of the first cluster of the projection plot for the MOMENT-small zero-shot version of the model for S2.

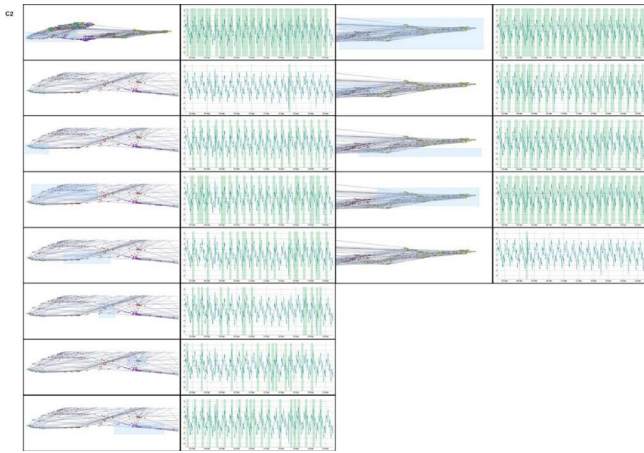


Fig. A22. Cluster II. Analysis of the second cluster of the projection plot for the MOMENT-small zero-shot version of the model for S2.

Fig. A23, Fig. A24, Fig. A25, Fig. A26, Fig. A27 show the five clusters of the execution of MOMENT-large in the zero-shot version.

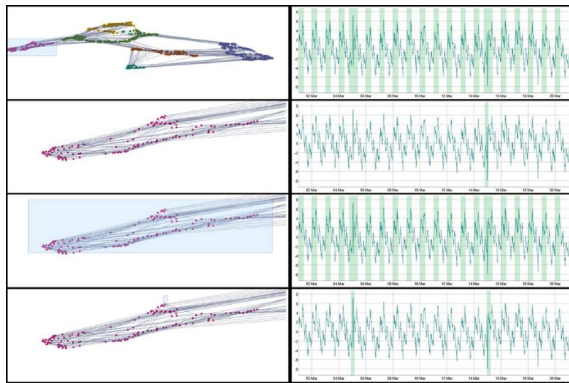


Fig. A23. Cluster I. Analysis of the zero-shot version of MOMENT-large applied to S2. Contains both anomalies.

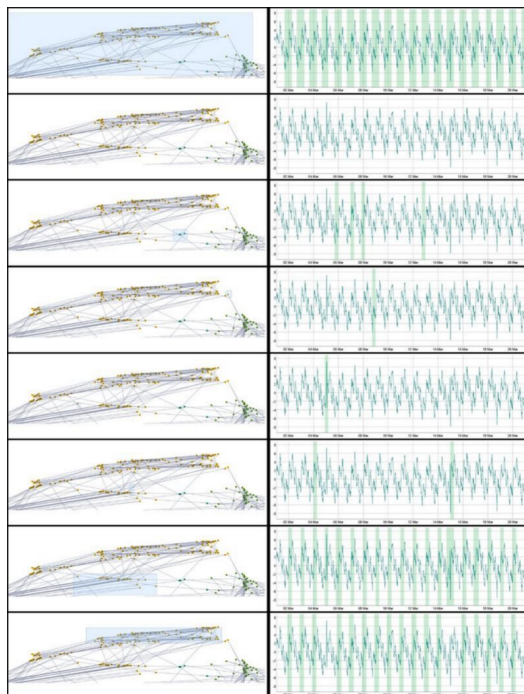


Fig. A24. Cluster II. Analysis of the the embeddings projections of the zero-shot version of MOMENT-large applied to S2. Contains both anomalies.

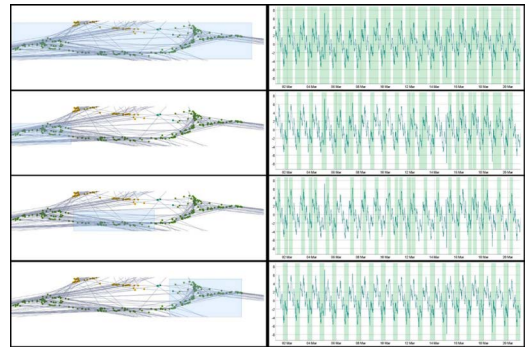


Fig. A25. Cluster III. Analysis of the embeddings projections of the zero-shot version of MOMENT-large applied to S2.

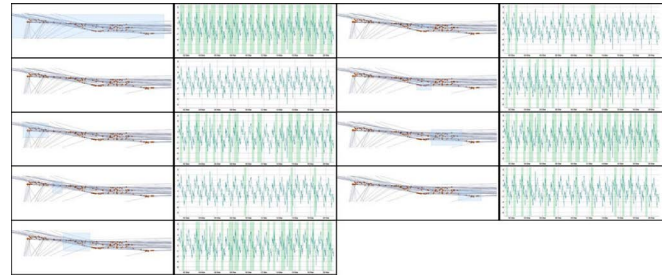


Fig. A26. Cluster IV. Analysis of the embeddings projections of the zero-shot version of MOMENT-large applied to S2.

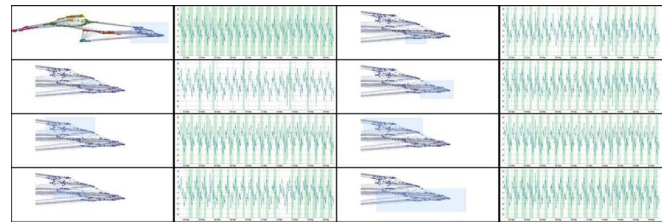


Fig. A27. Cluster V. Analysis of the embeddings projections of the zero-shot version of MOMENT-large applied to S2.

3. S3 Auxiliar Plots

Auxiliar plots for the visual analysis of S3 using the diferent versions of MOMENT.

Fig. A28 and Fig. A29 show the visualization resulting from the inference with zero-shot and fine-tuned versions of MOMENT-base while Fig. A30 contains the plot for the MOMENT-large version.

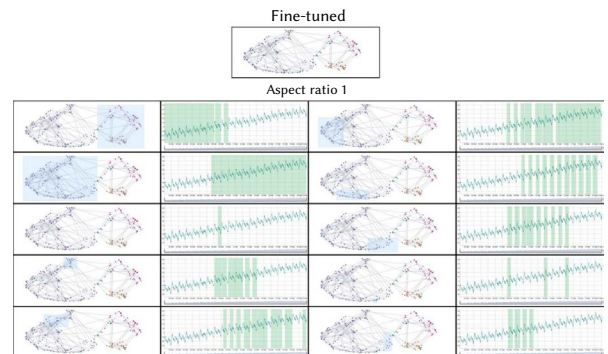


Fig. A28. Embeddings projections of the zero-shot version of MOMENT-base applied to S3.

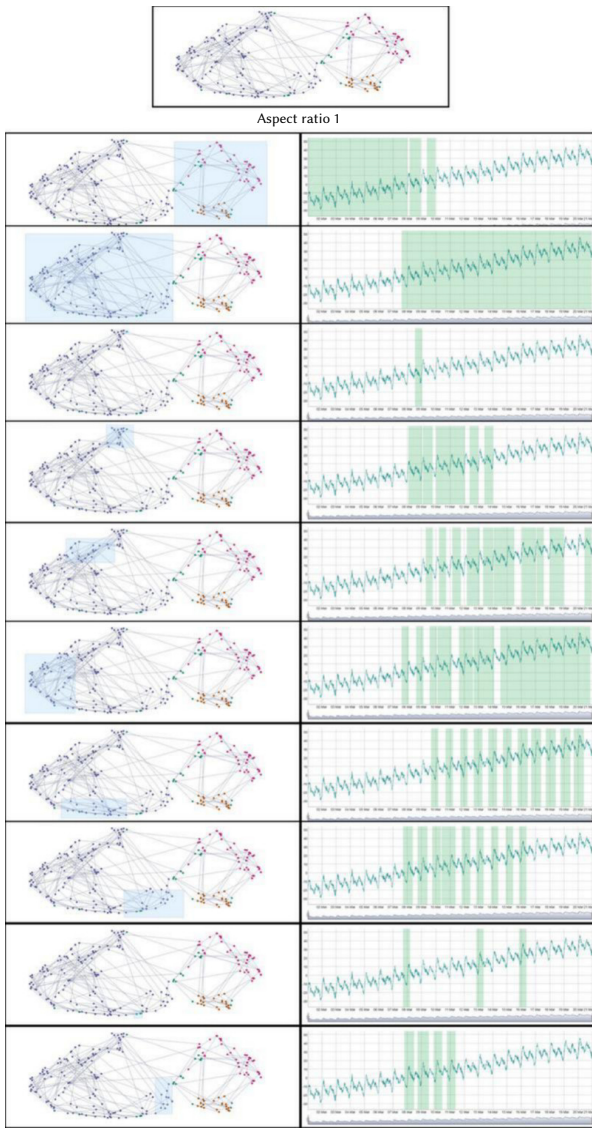


Fig. A29. Embeddings projections of the fine-tuned version of MOMENT-base applied to S3.

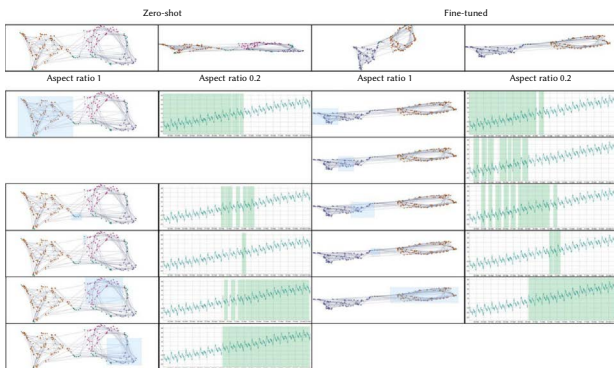


Fig. A30. Embeddings projections of the fine-tuned version of MOMENT-large applied to S3.

4. M-Toy Auxiliar Plots

Fig. A31 and Fig. A32 show the embedding projections for the MOMENT-base and MOMENT-large versions applied to M-Toy.

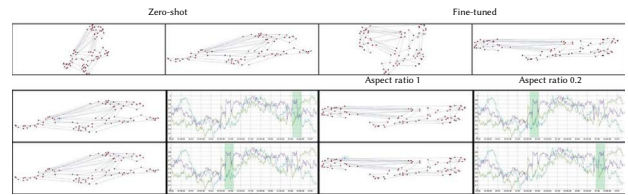


Fig. A31. Embeddings projections of MOMENT-base applied to M-Toy.

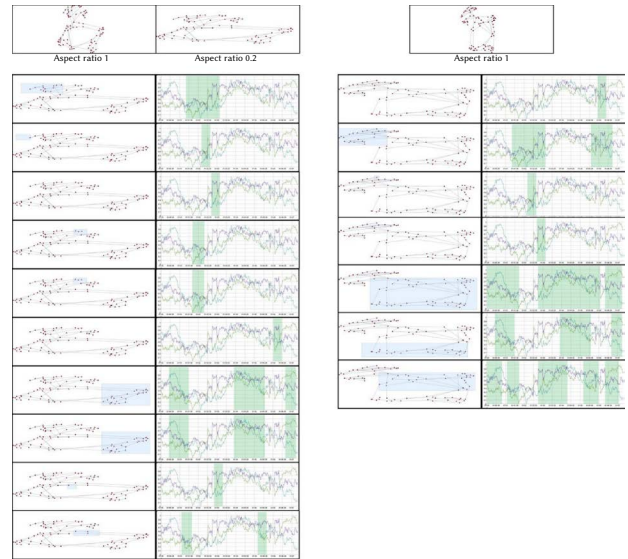


Fig. A32. Embeddings projections of MOMENT-large applied to M-Toy.

5. Kohl's Auxiliar Plots

Fig. A33 and Fig. A34 show the visual analysis of the zero-shot executions for MOMENT-base and MOMENT-small for the Kohl's dataset.

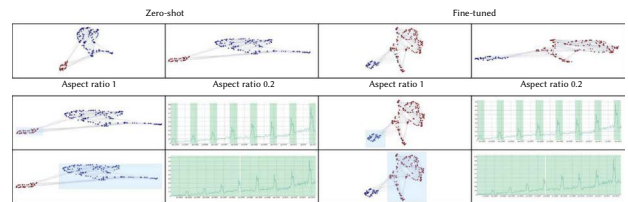


Fig. A33. Embeddings projections of MOMENT-base applied to Kohl's.

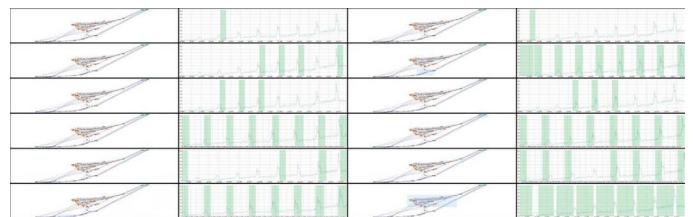


Fig. A34. Embeddings projections of MOMENT-large applied to Kohl's.

Fig. A35 shows the execution of the fine-tuned version of MOMENT large for Kohl's.

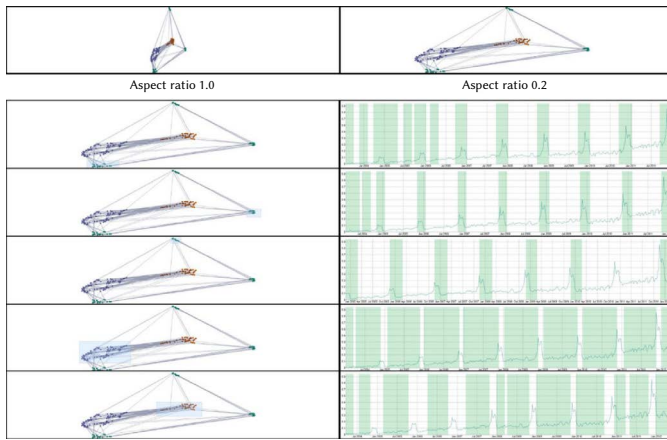


Fig. A35. Embeddings projections of the fine-tuned version of MOMENT-large applied to Kohl's.

CREDiT AUTHORSHIP CONTRIBUTION STATEMENT

Inmaculada Santamaria-Valenzuela: conceptualization, investigation, methodology, software, validation, visualization, writing (original draft), writing (review & editing).

Victor Rodriguez-Fernandez: conceptualization, investigation, methodology, software, supervision, writing (original draft), writing (review & editing).

Javier Huertas-Tato: methodology, software, supervision, validation, visualization, writing (original draft), writing (review & editing).

Jonh Hyuk Park: supervision, writing (review & editing), funding acquisition.

David Camacho: conceptualization, methodology, project administration, supervision, writing (review & editing), funding acquisition.

DATA STATEMENT

The data supporting the findings of this study are available from the following sources:

Datasets S1, S2, and S3: Generated using the code and methods described in this paper, with the source notebook available at: https://github.com/vrodriguezf/deepvats/blob/master/nbs/_synthetic_data.ipynb.

M-Toy Dataset: Publicly available on Zenodo at: <https://zenodo.org/record/4328047/files/toy.csv?download=1>

Kohls Dataset: Publicly available on Zenodo at: https://zenodo.org/record/4276348/files/Time_Series_Chains_Kohls_data.csv

The GitHub repository containing the reproduction code is available at: <https://github.com/vrodriguezf/deepvats>.

DECLARATION OF CONFLICTS OF INTEREST

We have no conflict of interest to declare.

ACKNOWLEDGMENT

This work has been supported by the following projects: H2020 TMA-MSCA-DN TUAI project "Towards an Understanding of Artificial Intelligence via a transparent, open and explainable perspective" (HORIZON-MSCA-2023-DN-01-01, Grant agreement n°: 101168344); by Strategic Networking and Development Program

funded by the Ministry of Science and ICT through the National Research Foundation of Korea (RS-2023-00267476); and by Comunidad Autonoma de Madrid, CIRMA-CAM Project (TEC-2024/COM-404).

REFERENCES

- [1] Y. Wu, S. Gao, S. Zhang, X. Dou, X. Wang, Q. Li, "From requirement to solution: Unveiling problem-driven design patterns in visual analytics," *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–18, 2025, doi: <https://doi.org/10.1109/TVCG.2025.3538768>.
- [2] X. Zhang, X. Yang, H. Hu, H. Qin, "Visual causal analysis of multivariate time series," *Journal of Visualization*, vol. 28, no. 2, pp. 323–339, 2025, doi: <https://doi.org/10.1007/s12650-024-01041-6>.
- [3] TensorFlow, "Embedding projector - visualization of high-dimensional data," 2023. [Online]. Available: <https://projector.tensorflow.org/>, Accessed: 2024-07-02.
- [4] L. McInnes, "Interactive visualizations," 2018. [Online]. Available: https://umap-learn.readthedocs.io/en/latest/interactive_viz.html, Accessed: 2024-06-26.
- [5] V. Rodriguez-Fernandez, D. Montalvo-Garcia, F. Piccialli, G. J. Nalepa, D. Camacho, "Deepvats: Deep visual analytics for time series," *Knowledge-Based Systems*, vol. 277, p. 110793, 2023, doi: <https://doi.org/10.1016/j.knsys.2023.110793>.
- [6] G. Woo, C. Liu, A. Kumar, D. Sahoo, "Pushing the limits of pre-training for time series forecasting in the cloudops domain," *arXiv preprint arXiv:2310.05063*, 2023.
- [7] Z. Zuo, Y. Huang, Z. Li, Y. Jiang, C. Liu, "Mixed contrastive transfer learning for few-shot workload prediction in the cloud," *Computing*, vol. 107, no. 1, p. 5, 2025.
- [8] A. G. Meyer, F. Lu, L. Clemente, M. Santillana, "A prospective real-time transfer learning approach to estimate influenza hospitalizations with limited data," *Epidemics*, vol. 50, p. 100816, 2025, doi: <https://doi.org/10.1016/j.epidem.2025.100816>.
- [9] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, P.-A. Muller, "Transfer learning for time series classification," in *2018 IEEE international conference on big data (Big Data)*, 2018, pp. 1367–1376, IEEE.
- [10] M. Goswami, K. Szafer, A. Choudhry, Y. Cai, S. Li, A. Dubrawski, "Moment: A family of open time-series foundation models," 2024.
- [11] A. F. Ansari, L. Stella, C. Turkmen, X. Zhang, P. Mercado, H. Shen, O. Shchur, S. Rangapuram, S. P. Arango, S. Kapoor, J. Zschiegner, D. M. Robinson, H. Wang, M. Mahoney, K. Torkkola, A. Wilson, M. Bohlke-Schneider, Y. B. Wang, "Chronos: Learning the language of time series," 2024. [Online]. Available: <https://arxiv.org/abs/2403.07815>.
- [12] G. Woo, C. Liu, A. Kumar, C. Xiong, S. Savarese, D. Sahoo, "Unified training of universal time series forecasting transformers," in *41 International Conference on Machine Learning (ICML 2024)*, 2024, pp. 38–44, PMLR.
- [13] I. Santamaria-Valenzuela, V. Rodriguez-Fernandez, D. Camacho, "On the integration of large-scale time series distance matrices into deep visual analytic tools," *Cognitive Computation*, vol. 17, no. 29, 2025, doi: <https://doi.org/10.1007/s12559-024-10394-x>.
- [14] I. Santamaria-Valenzuela, V. Rodriguez-Fernandez, D. Camacho, "Exploring scalability in large-scale time series in deepvats framework," in *Proceedings of the 13th Conference on Information Technology and its Applications (CITA 2024)*, vol. 882 of *Lecture Notes in Networks and Systems*, Cham, 2024, pp. 244–255, Vietnam - Korea University, Springer Nature Switzerland.
- [15] M. Shahcheraghi, R. Mercer, J. M. d. A. Rodrigues, A. Der, H. F. S. Gamboa, Z. Zimmerman, K. Mauck, E. Keogh, "Introducing mplots: scaling time series recurrence plots to massive datasets," *Journal of Big Data*, vol. 11, no. 1, p. 96, 2024, doi: <https://doi.org/10.1186/s40537-024-00954-1>.
- [16] L. Su, X. Zuo, R. Li, X. Wang, H. Zhao, B. Huang, "A systematic review for transformer-based long-term series forecasting," *Artificial Intelligence Review*, vol. 58, no. 3, p. 80, 2025.
- [17] A. Zeng, M. Chen, L. Zhang, Q. Xu, "Are transformers effective for time series forecasting?," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, 2023, pp. 11121–11128.
- [18] H. Wu, J. Xu, J. Wang, M. Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," *Advances in neural information processing systems*, vol. 34, pp. 22419–22430, 2021.

- [19] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
- [20] Y. Wang, H. Wu, J. Dong, G. Qin, H. Zhang, Y. Liu, Y. Qiu, J. Wang, M. Long, "Timexer: Empowering transformers for time series forecasting with exogenous variables," *Advances in Neural Information Processing Systems*, vol. 37, pp. 469–498, 2025.
- [21] Y. Wang, N. Huang, T. Li, Y. Yan, X. Zhang, "Medformer: A multi-granularity patching transformer for medical time-series classification," *Advances in Neural Information Processing Systems*, vol. 37, pp. 36314–36341, 2025.
- [22] S. G. Dhekane, T. Ploetz, "Transfer learning in sensor-based human activity recognition: A survey," *ACM Computing Surveys*, vol. 57, Mar. 2025, doi: <https://doi.org/10.1145/3717608>.
- [23] M. Awais, M. Naseer, S. Khan, R. M. Anwer, H. Cholakkal, M. Shah, M.-H. Yang, F. S. Khan, "Foundation models defining a new era in vision: a survey and outlook," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 4, pp. 2245–2264, 2025, doi: <https://doi.org/10.1109/TPAMI.2024.3506283>.
- [24] A. Rupa, C. Swapna, "Advancing domain-specific adaptations of large language models through transfer learning and fine-tuning techniques: An analytical study," in *Sentiment Analysis Unveiled*, CRC Press, 2025, pp. 58–73, 1 ed.
- [25] Y. Yang, J. Zhou, X. Ding, T. Huai, S. Liu, Q. Chen, Y. Xie, L. He, "Recent advances of foundation language models-based continual learning: A survey," *ACM Computing Surveys*, vol. 57, no. 5, pp. 1–38, 2025.
- [26] W. Dai, P. Chen, M. Lu, D. Li, H. Wei, H. Cui, P. P. Liang, "Climb: Data foundations for large scale multimodal clinical foundation models," 2025. [Online]. Available: <https://arxiv.org/abs/2503.07667>.
- [27] A. Garza, C. Challu, M. Mergenthaler-Canseco, "Timegpt-1," *arXiv preprint arXiv:2310.03589*, 2023.
- [28] K. Rasul, A. Ashok, A. R. Williams, A. Khorasani, Adamopoulos, R. Bhagwatkar, M. Biloš, Ghonia, N. Hassen, A. Schneider, et al., "Lag-llama: Towards foundation models for time series forecasting." Poster presented at the R0-FoMo Workshop, NeurIPS 2023, 2023. [Online]. Available: <https://openreview.net/forum?id=8R3VyopG16>.
- [29] L. Van der Maaten, G. Hinton, "Visualizing data using t-sne," *Journal of machine learning research*, vol. 9, no. 11, 2008.
- [30] L. McInnes, J. Healy, J. Melville, "Umap: Uniform manifold approximation and projection for dimension reduction," 2018. [Online]. Available: <https://arxiv.org/abs/1802.03426>.
- [31] Y. Wang, H. Huang, C. Rudin, Y. Shaposhnik, "Understanding how dimension reduction tools work: An empirical approach to deciphering t-sne, umap, trimap, and pacmap for data visualization," *Journal of Machine Learning Research*, vol. 22, no. 201, pp. 1–73, 2021.
- [32] S. Bhatnagar, A. Singh, V. Awasthi, S. SK, "Combining data mining and visualization: Umap on rope," in *2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS)*, 2024, pp. 1623–1629, IEEE.
- [33] J. Stym-Popper, N. Painchaud, P.-Y. COURAND, C. Rambour, N. Thome, O. Bernard, "Dafted: Decoupled asymmetric fusion of tabular and echocardiographic data for cardiac hypertension diagnosis," in *Proceedings of the Medical Imaging with Deep Learning Conference (MIDL 2025)*, San Francisco, CA, USA, 2025, pp. 1–15, PMLR.
- [34] M. Ali, M. W. Jones, X. Xie, M. Williams, "Timecluster: dimension reduction applied to temporal data for visual analytics," *The Visual Computer*, vol. 35, no. 6–8, pp. 1013–1026, 2019.
- [35] N. Andrienko, G. Andrienko, A. Artikis, P. Mantenoglou, S. Rinzivillo, "Human-in-the-loop: visual analytics for building models recognising behavioural patterns in time series," *IEEE Computer Graphics and Applications*, vol. 44, no. 3, pp. 14–29, 2024, doi: <https://doi.org/10.1109/MCG.2024.3379851>.
- [36] I. Oguiza, "tsai: A state-of-the-art deep learning library for time series and sequential data," 2023. [Online]. Available: <https://github.com/timeseriesAI/tsai>, Accessed: 2025-06-26.
- [37] H. Ismail Fawaz, B. Lucas, G. Forestier, C. Pelletier, D. F. Schmidt, J. Weber, G. I. Webb, L. Idoumghar, P.-A. Muller, F. Petitjean, "Inceptiontime: Finding alexnet for time series classification," *Data Mining and Knowledge Discovery*, vol. 34, no. 6, pp. 1936–1962, 2020.
- [38] S. M. Law, "STUMPY: A Powerful and Scalable Python Library for Time Series Data Mining," *The Journal of Open Source Software*, vol. 4, no. 39, p. 1504, 2019, doi: <https://doi.org/10.21105/joss.01504>.
- [39] Y. Matsubara, Y. Sakurai, C. Faloutsos, "The web as a jungle: Non-linear dynamical systems for co-evolving online activities," in *Proceedings of the 24th international conference on world wide web*, 2015, pp. 721–731.
- [40] M. Middlehurst, A. Ismail-Fawaz, A. Guillaume, C. Holder, D. Guijorubio, G. Bulatova, L. Tsaprounis, L. Mentel, M. Walter, P. Schäfer, et al., "aeon: a python toolkit for learning from time series," *Journal of Machine Learning Research*, vol. 25, no. 289, pp. 1–10, 2024.
- [41] M. Cuturi, M. Blondel, "Soft-dtw: a differentiable loss function for time-series," in *Proceedings of the 34th International Conference on Machine Learning - Volume 70, ICML'17*, 2017, pp. 894–903, PMLR, JMLR.org.



Inmaculada Santamaria-Valenzuela

Inmaculada Santamaria-Valenzuela is an Assistant Professor at the School of Computer Systems Engineering, Universidad Politécnica de Madrid (UPM). Her research focuses on deep learning and visual analytics for time series analysis, with particular interest in model interpretability and scalability. She has participated in the development of DeepVATS, a framework that combines deep learning with interactive visualisation tools to explore latent representations. She has presented her work at international conferences and journals and is an active member of the Applied Intelligence & Data Analysis research group at UPM.



Victor Rodriguez-Fernandez

Dr. Victor Rodriguez-Fernandez is an Associate Professor at the Universidad Politécnica de Madrid (UPM) and a core member of the Applied Intelligence and Data Analysis (AIDA) group. His research focuses on deep learning applications for time series analysis, emphasizing scalability and interpretability. He is the lead author of DeepVATS, a visual analytics framework for exploring latent representations in time series data. Dr. Rodríguez-Fernández has collaborated with institutions like MIT and ESA, contributing to projects on space traffic management, solar storm forecasting, and autonomous spacecraft control using large language models.



Javier Huertas-Tato

Dr. Javier Huertas-Tato is an Assistant Professor at Universidad Politécnica de Madrid (UPM) and a researcher in the Naturla Language Processing and Deep Learning (NLP-DL) group, specializing in machine learning, disinformation detection, and social media analysis. With expertise in deep learning techniques like transformers and adversarial robustness, he has contributed to projects such as FightDIS and IBERIFIER, developing AI-driven solutions for misinformation mitigation. His work spans text and multimodal content verification, cybersecurity, and environmental applications. His latest contributions include representation learning transformer models to detect writing styles in social media, and innovations in multimodal content verification to combat manipulated audiovisual data.



Jhon Hyuk Park

Dr. James J. (Jong Hyuk) Park is a professor at Seoul National University of Science and Technology (SeoulTech), Korea, with dual Ph.D. degrees from Korea University and Waseda University. He has an extensive academic and research background, having published around 400 papers in international journals and conferences. He actively contributes to the global research community through leadership roles in conferences and editorial positions in various journals. Dr. Park has received multiple awards for his research and leadership and was recognized among the World's Top 2% Scientists by Stanford University in 2021. His research focuses on areas such as IoT, Cloud Computing, Blockchain, Quantum Information, Information Security, and the Metaverse.



David Camacho-Fernandez

David Camacho (Senior Member, IEEE) received the Ph.D. degree (with Hons.) in computer science from Universidad Carlos III de Madrid, Getafe, Spain, in 2001. He is currently a Full Professor with Computer Systems Engineering Department, Universidad Politécnica de Madrid (UPM), Madrid, Spain, and the Head of the Applied Intelligence and Data Analysis research Group, UPM. He has authored or coauthored more than 300 journals, books, and conference papers. His research interests include machine learning (clustering/deep learning), computational intelligence (evolutionary computation, swarm intelligence), social network analysis, fake news and disinformation analysis. He has participated/led more than 50 research projects (Spanish and European: H2020, DG Justice, ISFP, and Erasmus+), related to the design and application of artificial intelligence methods for data mining and optimization for problems emerging in industrial scenarios, aeronautics, aerospace engineering, cybercrime/cyber intelligence, social networks applications, or video games among others. He was the recipient of the best thesis award in Computer Science for his Ph.D. degree.