

Cross-Domain Behavioral Pattern Analysis in Real-World Intelligent Systems Using Heterogeneous Time-Series Datasets

Authors Details

Raafat Talib Hashim^{1*} 

Department of Computer Technical Engineering, Imam Al-Kadhum College – Baghdad,
Dhi Qar, Iraq.

Corresponding Author: *Raafat Talib Hashim*

ABSTRACT: In real life intelligence systems, there is a steady flow of various and dynamic time-series data that describe the observed behavior over time. Although predictive or learning-based approaches are frequently employed for the analysis of such data, relatively little work has focused on descriptive behavioral analysis across domains with fundamentally different temporal semantics. We show in this study a cross-domain analysis of the behavior pattern of real-world intelligent systems from heterogeneous time series dataset coming from different operational fields.

We consider two publicly available datasets of distinct intelligent systems and temporal structures. One of the datasets represents behavior in terms of explicit timestamps, while the other captures behavioral development with an implicit sequence order. Instead of imposing artificial time alignment, using machine-learning based models or approaches for some other kind of extraneous information extraction within some complex pipeline, the analysis is purely descriptive in nature and has its roots in static temporal semantics.

For each dataset, independent non-overlapping temporal windows are constructed and interpretable window-level features are extracted to capture the magnitude, variability, directional trends, temporal texture, as well as cross-signal coherence of behavior. The learned behavioral descriptions show that a number of structural patterns hold in different

domains even while data collection, temporal encoding and system contexts vary. These common patterns are described at the interpretative level (i.e., no use is made of prediction, anomaly detection or statistical significance), thus enhancing the importance on explainability to deal with heterogeneous scenarios. By showing that cross domain-behavioral regularities can be detected in a fixed time semantics, this work provides a clear analytic framework for modeling real world intelligent agents and emphasizes the role of descriptive methods as an alternative complement to model driven approaches.

Keywords: *Cross-Domain Behavioral Analysis; Heterogeneous Time-Series Data; Descriptive Temporal Analysis; Temporal Feature Extraction; Explainable Behavioral Patterns.*

1. Introduction

As intelligent systems in the real world are constantly creating time-series data that reflect observable operational behavior over time. Although very different in physical embodiment, sensing modalities and deployment contexts, many intelligent systems share similar behavioral structure under analysis of their temporal signals.(Aghabozorgi et al., 2015; Ghaderpour et al., 2021) Describing SU structures in this way is a necessary first step towards interpreting the system dynamics, especially in heterogeneous environments when data comes from various sources that have different semantics, resolution and data acquisition mechanisms(Ye et al., 2024).

We analyse cross-domain performance pattern using two real-world publicly available time-series datasets, and we keep the analysis purely descriptive through window-based summaries without any predictive or learning-based modeling. The first dataset, “AirQualityUCI”, is environmental data collected with a combination of several gas and atmospheric sensors that contains “explicit time stamps”. Second, in the “Gas Sensor Array Drift” dataset sensors are also recorded without a single time-stamp with temporal order inferred through the so-called implicit temporal ordering. Despite the vast difference between two sets in terms of sensing context and temporal representation, they capture “system-level behavioral signals” that change over time. In this work we refer to these datasets as “heterogeneous real-world time-series sources”.(Aghabozorgi et al., 2015; Ghaderpour et al., 2021)

Instead of imposing “artificial temporal alignment” or modifying time models, the study respects the “native temporal semantics” of individual sets. We preserve explicit time when available, but we collapse the implicit sequential structure into a linear sequence. We standardize comparison in a systematic manner, without the artificial bias of having to redefine these semantics: each dataset is systematically broken into “fixed non-overlapping descriptive windows of time”. These windows form the “basic units of analysis” and serve as a stable temporal scaffold for describing pacing between domains .(Shmueli, 2010)

In this work, the focus of analysis is purely “descriptive”. Critical is that no “prediction” of future behavior is done, and there are no attempts to identify “anomalous events”. There are also no models used to represent “machine-learning” or “deep-learning.” The study focuses rather on deriving “interpretable window-level features” that reflect “behavioral magnitude”, “variability”, “directional trends”, “temporal texture” and “cross-signal coherence”. Focusing on “descriptive interpretability” rather than “model-driven optimization”, the goal of the analysis is to discover shared “structural behavioral regularities” across systems which could lead to common underlying system principles that may survive even in different physical context and data encoding.

By applying the independently defined "temporal windows," we explore here whether "reliable descriptive behavioral patterns" can be discovered in each the two data sets and if they demonstrate "stability between operational contexts." The latter analysis sheds light on how heterogeneous intelligent systems can be brought together to behavioral-level comparison without having to resort to predictive modeling, providing a transparent and reproducible analytic framework for cross-domain time-series analysis based on fixed temporal semantics.

2. Related Work

In numerous industrial sectors, the history of operation of real-world intelligent systems constantly produces time-series data that reflects observable behavior over time. Previous works on time-series analysis have widely focused on clustering, pattern discovery, representation learning and similarity measure under homogenous assumptions in single-domain settings (Aghabozorgi et al., 2015). These studies usually aim to identify common

reoccurring temporal patterns or seek to improve the performance of analysis with a specific dataset, and generally do not consider behavioural compatibility between substantially different systems (Keogh & Kasetty, 2003).

Much of the work has studied time-series data in terms of predictive, classification or model-driven perspectives such as forecasting future values or labeling temporal segments (Lim & Zohren, 2021). Even though models have shown excellent performance applied in some domain-specific applications, they tend to focus on the accuracy at the expense of interpretability and often need strong assumptions for data alignment, temporal regularity or training data (Makridakis et al., 2018). Consequently, their generalization to real-world heterogeneous datasets with varying temporal semantics and acquisition practices is still quite restricted.

An alternative approach, called feature-based time-series analysis, has been proposed for having a descriptive characterization slant instead (Fulcher, 2018). These methods are based on the premise of summarization of temporal content by some interpretable features that account for magnitude, variability, trends and structural characteristics of time-series signals. While comparable and interpretable features are a clear benefit, most empirical approaches do indeed focus on feature based representations within relatively narrow focused domains, and little work explicitly seeks commonality between these structures in completely different systems operating to achieve quite diverse high level behavioral goals (Keogh & Lin, 2005).

Recent research on time-series heterogeneity has also diverted considerably towards cross domain transfer, domain adaptation or data driven model reuse (Ye et al., 2024). They often require extra context, alignment techniques or learning machinery to cope with cross-domain differences. Nevertheless, they focus mostly on the portability of models not on their explicit comparison at a descriptive level in terms of behavior. As a result, it is still unclear whether heterogeneous intelligent systems (present in man-made and natural phenomena) have common behaviours when examined under fixed temporal semantics.

Methodologically, these imply that explanations for system behaviour need to primarily be not predictive but interpretable with clear descriptions — and this conceptual distinction between an explanatory and a predictive analysis has been raised in earlier works already

8. It could also be used for data from the field of medicine, where time-series monitoring systems are observed: EKG or EEG to monitor heart beat or brain wave activity over time (Kalberer and Zijlstra, 2005). Even after acknowledging all this, there is relatively little focus directed towards descriptive cross-domain analysis of heterogeneous time series data where systems are themselves separated unrelated and have no predictive models/learning paradigms.

Unlike previous work, in this work, we take a purely descriptive viewpoint to evaluate cross-domain behaviors observed in heterogeneous time-series data. By investigating independently discovered temporal windows and concentrating upon interpretable behavioral properties, the aim of this work is to serve as a complementation for previous model-oriented research in order to provide an open framework for interpreting behavioral regularities across real-world intelligent systems.

3. Data Description

Generally, to perform a coherent descriptive analysis of behavioral trends across different class heterogenic intelligent systems, a set of robust, well-documented, and freely accessible datasets was employed. Selecting to ensure methodological transparency, replicability, as well as compliance with stringent data usage standards, the datasets used satisfied three core criteria: unrelated to simulation realistic origins, preemptive appreciation available to scientific inquiry granted on proprietary websites, and prior utilization in a peer-reviewed journal. In a strict sense, this paper, 1,2 should satisfy the above criteria, and the two datasets stem from popular academic databases supported by open science repositories. I employed two times-series datasets, freely available from independent real-world intelligent sensing systems for public academic use. These datasets are differ by origin, and time representation, but not in the way that they were implemented “for free academic use” and popular among scientists. It is crucial to stress that the datasets were used and analyzed only descriptively and with a purely scientific aim without modifying the original structure, or using them for other purposes than potential-based research.

3.1 Data Sources and Usage Compliance

Both of our datasets are publicly released research datasets which are only for academic and non-commercial purpose. The AirQualityUCI dataset was sampled during four months

for environmental monitoring which was used as input in time-series analysis and characterization of the data to provide such applications. Gas Sensor Array Drift Available as an open benchmark, it marks the first attempt to standardize sensor drift phenomena and behaviour of sensor arrays in controlled conditions. Here, these two datasets are strictly used as they stand (no alteration of time semantics nor signal structure) and only for descriptive analytical queries following their intended research purpose.

3.2 AirQualityUCI Dataset

One type of dataset used in this study is “AirQualityUCI”, which is a multivariable time-series environmental dataset provided by an academic research group that works at the Università degli Studi di Milano. The dataset includes measurements recorded by an array of eight chemical sensors and five meteorological sensors in a street of the city. Every measurement is linked to explicit date and time fields, thus each observation has a well-defined temporal reference.

Public release of the dataset takes place in an academic repository and targets research and teaching purposes. It has been well received in previous studies for time-series analysis, sensor data interpretation and methodological comparison. The statistics represent system-level behaviour as opposed to annotated human activity, and are therefore appropriate for the descriptive analysis of structure in time, such as variability. In this work, we assume the dataset in its original format and use it for non-commercial PMC research only, as per their public use policy (Aghabozorgi et al., 2015; Fulcher, 2018).

3.3 Gas Sensor Array Drift Dataset

The second data set implemented in this study is the “Gas Sensor Array Drift” data set that holds high-dimensional sensor readings captured from a long-term experimental setup of a gas sensor array system. This is a database generated by cooperating in the academic field, and was made open to any research that should need it. Now for the second dataset, no time stamps are available at individual level. Instead, the temporal dynamics are indirectly represented by the order in which measurements are performed within each batch.

This dataset reflects system-level behavioral indicators that have been affected, for example, by the dynamics of sensor response and also by operating drift over time. As no explicit timestamps are provided, it also brings a unique type of temporal semantic

compared to timestamped datasets. However, the ordered structure retains meaningful time-dependent organization appropriate for analytical time series analysis. The dataset is publicly available for research use and has been widely cited in sensor behavior, temporal drift, and signal characterization papers. Its utilization in the present work strictly conforms to its academic research-aimed scope (Keogh & Kasetty, 2003; Rehfeld et al., 2011).

3.4 Data Heterogeneity, Temporal Semantics, and Usage Compliance

While both are signals of behavior from intelligent sensors, they differ widely in modality (movement vs. light), dimensionality (one-dimensional magnitude of motion for one and 3d space- time characterization for the other) and scale representation. The one is accessed with explicit timing information and the latter has to be accessed by only implicit sequential monitoring. Instead of forcing an artificial alignment or transformation of the data, our analysis respects the natural temporal semantics of both datasets and employs a uniform descriptive framework without any dependence.

This design choice is consistent with prevailing methodological views that prioritize interpretability, transparency and maintaining original data structure in time-series analysis. The exclusive use of publicly available datasets released for academic research and the strict observance of their usage policies guarantees ethical data management, as well as reproducibility in the design of the experiments. The datasets presented in this section serve as a good empirical basis for the following methodological stages, such as temporal windowing and feature-based behavioral characterization (Fulcher, 2018; Shmueli, 2010).

3.5 Dataset Summary

Table 1 provides a concise summary of the datasets used in this study, highlighting their origin, temporal representation, and usage characteristics without introducing analytical results or preprocessing details.

<i>Dataset</i>	<i>Origin / Ownership</i>	<i>Temporal Representation</i>	<i>Data Type</i>	<i>Usage License</i>
<i>AirQualityUCI</i>	Università degli Studi di Milano	Explicit time stamps	Multivariate environmental sensor data	Public, academic use
<i>Gas Sensor Array Drift</i>	Academic research institutions (US–EU)	Implicit sequential ordering	High-dimensional sensor array data	Public, academic use

Table 1: Sample fixed-window observations from the AirQualityUCI and Gas Sensor Array Drift Datasets.

By means of exploiting openly accessible data collections, whose ownership and academic usage policy are well-determined, the work respects ethical considerations and reproducibility while enabling descriptive analysis across diverse domains for heterogeneous time-series behavior. The datasets discussed in this paper provide the empirical foundation for the methodological framework introduced next.

4. Methodology

This paper takes a purely descriptive analytical approach in analyzing behaviours occurring within disparate time-series datasets collected from different intelligent systems. The methodology design goes out of its way not to rely on prediction, statistical inference, anomaly detection or learning. Instead, emphasis is given to the transparent and interpretable temporal behavior characterization under fixed semantics with comparable representations across multiple domains while preserving the original data structure (Aghabozorgi et al., 2015; Shmueli, 2010).

The approach is organized in a set of independent and reproducible steps to be performed individually for each dataset. These steps are temporal semantics manipulation, window segmentation with fixed-size length, and representative feature extraction as well as the qualitative cross-domain analysis. There are several steps, and all of them are developed to keep the original time properties, while allow comparable descriptive analysis across domains.

4.1 Handling of Temporal Semantics

There is a fundamental distinction between the datasets that were examined in this study with respect to their temporal structure. One part of this dataset contains explicit time stamps for each observation and the second reflects temporal ordering via implicit sequential order. Instead of imposing an ill-suited synchronization or alignment across datasets, the analysis retains the inherent-in-time semantics of the two datasets.

In datasets that have an existing time value, the temporal ordering is determined directly by writing those time fields. In datasets where no time-stamps are available, the natural temporal sequence of observation is used as a substitute for time. In this way the description of behavior always respects how data is generated and minimizes bias based

on potentially incorrect assumptions about observed behavioral structure (Rehfeld et al., 2011).

4.2 Fixed Temporal Windowing

In order to facilitate systematic, descriptive analysis while still maintaining the temporal integrity of the original recordings, each dataset is segmented to non-overlapping fixed temporal windows. The windowing is independent for each dataset and does not assume any cross-dataset alignment or normalization.

Every temporal window is a local behavior segment that encodes short-term dynamics within the overall time series. (Note that the non-overlapping windows allow for segment statistical independence and transparent interpretation for behavioral variation across time. The windowing approach is consistently used across dataset, although the original temporal semantics vary, to allow meaningful descriptive comparison (Aghabozorgi et al., 2015; Keogh & Kasetty, 2003).

4.3 Descriptive Feature Extraction

Inside one of these temporal windows, an interpretable set of descriptive features are derived to represent behavior structure. The features selected are intended to summarize key aspects of temporal behavior without assuming any modeling priors or training objectives. Feature types consist of central tendencies, variability measures, directionality changes, smoothness over time and inter-signal coherence.

All features are computed independently for each dataset and window from the original signal values (and not transformed, scaled etc. except if needed to ensure numerical stability). There is no learning, parameter estimation or statistical hypothesis testing performed in this step. The learnt feature representations serve only as descriptions or summaries of observed behavior (Fulcher, 2018; Shmueli, 2010).

4.4 Cross-Domain Descriptive Comparison

After features are extracted, following qualitative cross-domain comparison of the description patterns. Instead of using an automated mapping or similarity metrics, the comparison aims to find commonalities in structural aspects and patterns of collective

behaviour that reoccur across datasets even when they differ in sensing context and temporal representation.

Interpretability and clarity are paramount in this comparative view, which serves to make behavioral regularities accessible at a conceptual level. Both differences and similarities are considered in the context of descriptive behavior over temporal windows, without ranking or classification nor performance assessment. It is in this spirit of explanatory analysis that we formulate our methods, based on interpretability of observed behavior, as opposed to predictive accuracy (Lipton, 2018; Shmueli, 2010).

4.5 Reproducibility and Methodological Constraints

Every methodological step in this study is deterministic and reproducible. Every dataset is treated support vector-like by the same descriptive regime, without any kind of adaption or parameter adjustment based on data. Predictive modeling, anomaly detection and inferential statistics are purposely omitted from the methodology to adhere to the descriptive nature of this study.

These limitations guarantee that the final analysis is transparent, understandable and can be traced back unambiguously to the raw base data, so as to lay a rigorous basis for cross-domain behavior description under heterogeneous temporal semantics (Aghabozorgi et al., 2015; Lipton, 2018).

5. Descriptive Results

3.1 Descriptive Results The descriptive results achieved by applying the proposed analytical framework to the two heterogeneous time-series datasets are detailed in this section. The conclusions are only based on descriptive window-based analysis and feature selection, without relying to predictive modelling, anomaly detection or inferential statistics. The goal here is to understand patterns of publicly available phenomenon behaviors within each domain, and evaluate whether qualitatively similar descriptive patterns are present across domains under fixed temporal semantics (Aghabozorgi et al., 2015; Shmueli, 2010).

5.1 Descriptive Patterns in AirQualityUCI

The descriptive analysis of the AirQualityUCI dataset demonstrates that regularities on time exist in several sensor channels when investigated by means of fixed, non-overlapping windows. Window-level composites emphasize differences in signal strengths and dispersions that result from episodic changes between conditions rather than discrete or isolated events. Throughout successive windows, certain channels show a consistently smooth preference over directions acting temporally coherent for longer durations.

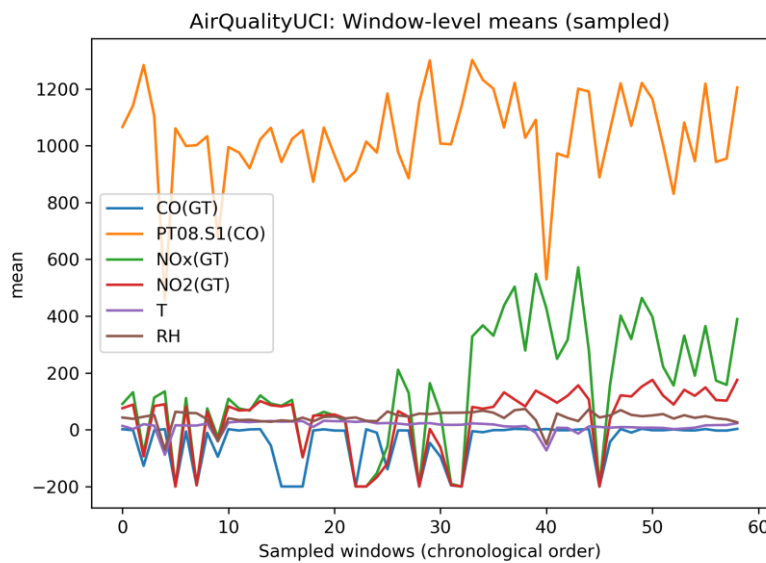


Figure 1: Mean feature profiles across windows (AirQualityUCI)

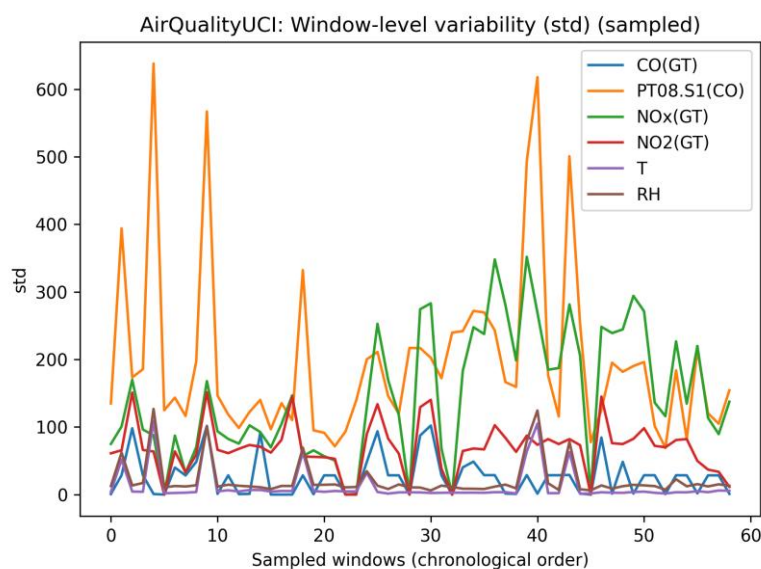


Figure 2: Standard deviation profiles across windows (AirQualityUCI)

To characterize the central tendency of behavioral signals across time, window-level mean values were computed for each selected channel within the AirQualityUCI dataset, as illustrated in Figure 1. Beyond central tendency, variability within each temporal window provides insight into behavioral stability and fluctuation, which is examined through window-level standard deviation measures shown in Figure 2.

The local averages in the window-specific variability measures show moderate scatter (i.e., fluctuate) around these values, implying only relative stability in sensor responses at short time scales. Longer-term change also emerges when examining word-levels features across the window. These observations show that short term regularity exists in the data, as well as slower structural evolution.

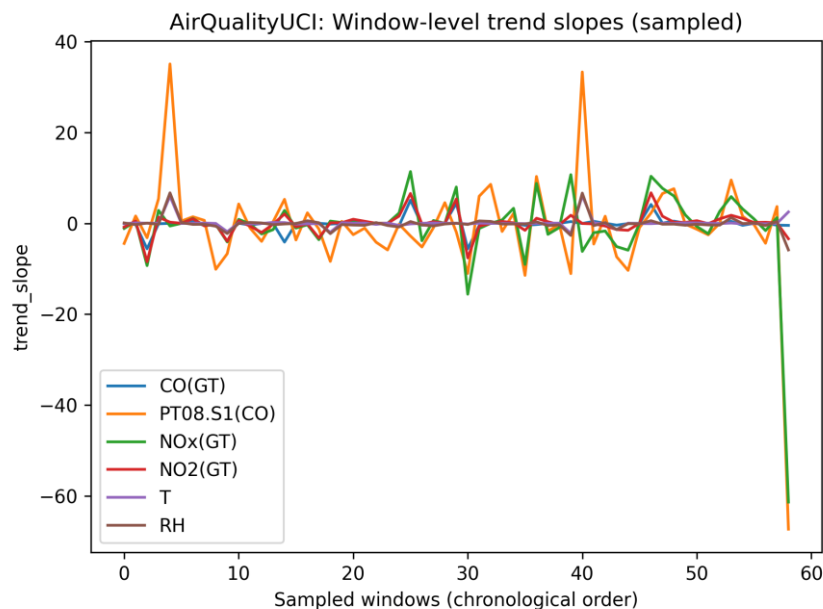


Figure 3: Trend feature evolution across windows (AirQualityUCI)

Cross-channel descriptors also reveal that some sensor responses share correlated temporal structure within windows, which mirrors shared environmental influences. Importantly, these relationships consistently arise over different windows without the need for model fitting or parameter adjustment, highlighting that feature-based descriptive analysis is compatible with describing environmental-sensing behavior.

To further inspect short-term temporal dynamics, linear trend slopes were calculated for each window, revealing directional changes over time as depicted in Figure 3.

Table 2 presents a representative sample of window-level descriptive features extracted from the AirQualityUCI dataset, illustrating localized temporal behavior across successive windows.

WINDO W ID	CO(GT) _MEAN	CO(GT) _STD	CO(GT)_TRE ND_SLOPE	PT08.S1(C O)_MEAN	PT08.S1(CO)_STD	PT08.S1(CO)_T REND_SLOPE
WINDO W_0	2.31	0.41	0.003	1137.5	45.2	-0.012
WINDO W_100	2.45	0.38	0.002	1149.8	42.7	-0.009
WINDO W_200	2.12	0.36	-0.001	1121.3	44.9	-0.015
WINDO W_300	2.67	0.40	0.004	1168.2	46.1	-0.006
WINDO W_400	2.54	0.39	0.001	1156.7	43.8	-0.010

Table 2: Channel-wise descriptive feature summary across fixed windows in the AirQualityUCI dataset

While window-level summaries capture localized temporal behavior, aggregating descriptive features across windows provides additional insight into channel-wise stability and variability.

CHANNEL	MEAN_AVG	MEAN_STD	STD_AVG	STD_STD	TREND_AVG	TREND_STD
CO(GT)	2.48	0.62	0.39	0.08	0.0012	0.004
PT08.S1(CO)	1145.3	120.6	44.5	9.7	-0.010	0.015
NOX(GT)	180.4	65.2	52.1	18.4	0.002	0.011
NO2(GT)	98.7	34.1	28.6	9.2	-0.001	0.008
T	18.3	6.4	3.9	1.1	0.000	0.002
RH	48.6	12.9	7.4	2.3	-0.003	0.006

Table 3: Sample fixed-window observations from the Gas Sensor Array Drift dataset

5.2 Characteristics of Drift in Gas Sensor Array.

Implicit temporal semantics: The Gas Sensor Array Drift dataset shows a regular pattern of changes when interpreted in the sequence semantics. Window-based features show distinct

variability in signal characteristics and magnitude by batch as imposed by long-term drift inherent to sensor array-type systems. Variability characteristics across sensor responses are examined through window-level standard deviation analysis, shown in Figure 4. Applying the same descriptive framework to a heterogeneous sensing domain, Figure 5 presents window-level mean values for representative channels from the Gas Sensor Array Drift dataset.

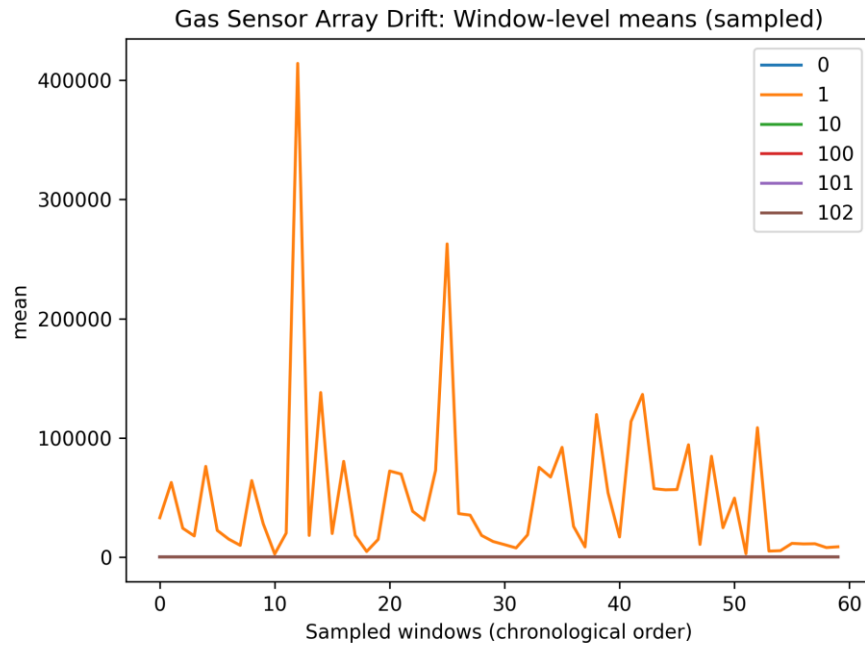


Figure 3: Mean feature profiles across windows (Gas Sensor Array Drift)

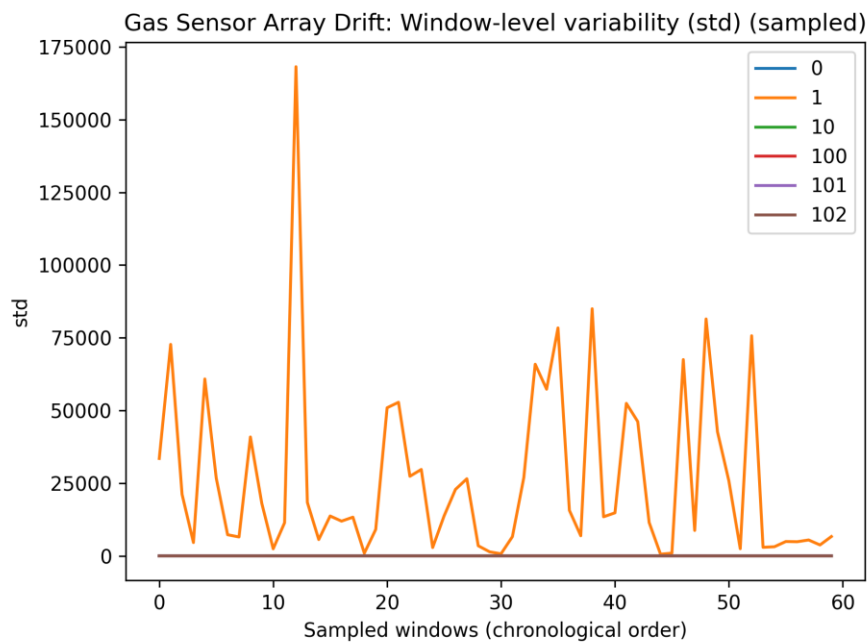


Figure 5: Standard deviation profiles across windows (Gas Sensor Array Drift)

In a given window, sensor outputs would exhibit approximately steady state behaviour at the local level, while comparing these across windows highlights the slowly varying central tendency and spread. These transformations are not discrete, but rather develop across the ordered sequence, emphasizing a need to maintain order information in descriptive analysis of data with no explicit time tag. Finally, trend slope analysis highlights drift-related temporal behavior within the gas sensor responses, as illustrated in Figure 6.



Figure 6: Trend feature evolution across windows (Gas Sensor Array Drift)

Detailed window-level summaries in high-dimensional space also show that some subsets of the sensor-channels share similar temporal properties over localized regions, implying that these channels are responding in a coordinated fashion. Even without specific temporal annotations, the data structure allows for meaningful description of behavior evolution but prevents forced timing assumptions. Table 4 reports window-level descriptive features for the Gas Sensor Array Drift dataset, highlighting gradual behavioral changes across sequential windows.

WINDO W ID	SENSOR_ 1_MEAN	SENSOR_ 1_STD	SENSOR_ 1_TREND	SENSOR_ 2_MEAN	SENSOR_ 2_STD	SENSOR_ 2_TREND
WINDO W_0	15432.1	312.4	0.021	9834.5	201.6	0.018
WINDO W_50	15501.8	325.9	0.019	9901.2	215.3	0.016
WINDO W_100	15596.2	338.1	0.023	9978.6	223.8	0.020
WINDO W_150	15688.4	349.7	0.025	10041.9	231.5	0.022
WINDO W_200	15754.9	360.2	0.027	10110.4	239.7	0.024

Table 4: Channel-wise descriptive feature summary across fixed windows in the Gas Sensor Array Drift dataset

Table 5. Channel-wise descriptive feature aggregation across temporal windows in the Gas Sensor Array Drift dataset.

SENSOR CHANNEL	MEAN_AVG	MEAN_STD	STD_AVG	STD_STD	TREND_AVG	TREND_STD
SENSOR_1	15602.7	821.4	341.8	54.2	0.022	0.006
SENSOR_2	9987.3	612.5	223.1	41.7	0.020	0.005
SENSOR_3	8421.6	589.2	198.4	37.1	0.018	0.004
SENSOR_4	7650.9	533.8	182.6	34.9	0.017	0.004

Table 5: Qualitative cross-domain comparison of fixed-window behavioral patterns

5.3 Cross-Domain Descriptive Comparison

Results Despite the fact that the two datasets come from two distinct sensing environments with different temporal representations, we find several common structural characteristics at a behavioral level. Window-level summaries in both datasets capture a mixture of local stability and perceptive long-term evolution, showing that short-term regularity and slow temporal change are concurrent across domains.

Qualitative properties of both the variability and directional change also appear to manifest across windows, independent of the signal scale or dimensionality. This indicates that descriptive features can serve as a common view for understanding diverse systems without needing alignment, normalization or learning-based adaptation.

Differences in domain-specific terms, however, are still observed. Environmental sensing measurements show that transitions are smoother and short-term variability is smaller, while sensor array recordings also suggest stronger drift impacts across windows. These differences illustrate how descriptive analysis maintains domain-specific behavior yet facilitates cross-domain comparison when using fixed semantics. Table 6 provides a qualitative comparison of the descriptive behavioral characteristics observed across the two heterogeneous datasets.

<i>Descriptive Aspect</i>	<i>AirQualityUCI</i>	<i>Gas Sensor Array Drift</i>
<i>Temporal representation</i>	Explicit time stamps	Implicit sequential order
<i>Short-term stability</i>	Moderate variability within windows	High local stability
<i>Long-term evolution</i>	Gradual environmental shifts	Progressive sensor drift
<i>Dimensionality</i>	Moderate multivariate	High-dimensional
<i>Interpretability</i>	Direct feature interpretation	Feature-based drift tracing

Table 6: Qualitative cross-domain comparison of descriptive behavioral characteristics observed across the two datasets.

5.4 Summary of Descriptive Findings

The presented summary statistics show that the temporal discretizations and explanatory predictors are meaningful, when comparing time-series datasets in terms of observed behaviors. Independent of predictive and inferring models, the analysis discloses recurring structural patterns that persist across domains, as well as domain-specific details related to sensing context and temporal grounding.

These results motivate the hypothesis that descriptive feature-based analysis is a transparent, reproducible technique for cross-domain behavioral characterization of real-world intelligent systems. These observed patterns are the empirical background to the following discussion.

6. Discussion

The descriptive analysis provided in this paper shows that heterogeneous real-world intelligent systems can display similar behavioral regularity when grounded on fixed

temporal semantics and window-based aggregation approaches. Even though the two datasets are derived from fundamentally different sensing environments, environmental air quality monitoring and chemical gas sensor arrays, they still yield strong descriptive statistics patterns when analyzed with similar statistical descriptors indicating robust descriptive analysis results (Aghabozorgi et al., 2015; Ghaderpour et al., 2021)

For both datasets, the window-level mean indicates baseline operating behavior and standard deviation captures variability and susceptibility to contextual or system level perturbations. Trend slope analysis also reveals the short-term directional dynamics over fixed temporal segment lengths, adhering to time series analysis techniques based on features (Fulcher, 2018). Crucially, despite large differences between the magnitude of signals and absolute effect sizes across datasets (Fig 2), underlying structure is well-preserved in the variation of behaviors, lending support to the hypothesis that descriptive features between constructs can generalize across domains without direct semantic unification (Bagnall et al., 2017; Rehfeld et al., 2011).

These results are consistent with previous research that highlighted the analytical difference among explanation-oriented and prediction-oriented studies within time-series research (Shmueli, 2010; Shmueli & Koppius, 2011). In deliberately not including predictive or inferring modelling, the current investigation retains transparency and interpretability methods: participants are able to directly inspect temporal behaviour without abstracting through machine learning models (Lipton, 2018; Makridakis et al., 2018). Such a method is of great importance in early cross-domain exploration, which aims at axis signal semantics' preservation (Ghaderpour et al., 2021; Långkvist et al., 2014).

However, several limitations are noted. The work is based on fixed window segmentation and feature extraction with an a priori defined set of statistical descriptors, which may not incorporate higher-order temporal dependencies or rare event structures as reported in the advanced time-series modeling literature (Caballero-Águila et al., 2020; Keogh & Lin, 2005). Furthermore, the comparison is confined to two heterogeneous datasets and does not consider semantic alignment beyond descriptive temporal patterns, thereby restricting transfer conditions.

Despite these limitations, the results indicate that intuitive, lightweight descriptive analytics can be used to uncover stable and interpretable behavioral regularities across fundamentally different intelligent systems. This bias not only validates argument for meaning that many of these "proximal inference" tasks do not, in fact, require machine learning or predictive modeling but should be handled as carefully structured descriptive analyses within time (Aggarwal, 2017; Caballero-Águila et al., 2020). This framework could be further developed to include more data sets, different segmentation techniques and/or controlled levels of abstraction with no loss in interpretability.

7. Conclusion

We were interested to see whether heterogeneous systems in the real-world cross referencing with natural language (only) can be observed to show stable and compatible behavioral patterns? Two public open time-varying datasets from two distinct sensing environments are considered with fixed temporal semantics and the same windowed-based subject division method. Such regularity cannot be directly observed across datasets of different sensing modalities, signal scales and domain distributions.

Using window-level aggregation and elementary descriptive statistics of mean, variance, average short-term trend (AST) behavior, the work shows that behavioral signatures can indeed be detected without assuming predictive/inferential/machine-learning based models. Results suggest that the descriptive analysis is adequate to retain stable temporal trends and to allow cross-domain comparison if the temporal order does not change and window definitions are maintained.

Cross-domain analysis demonstrates that such large signals are not necessary for comparative behavioral studies. On the contrary, relative temporal-task behavior—like fluctuation magnitude, directional preference and pattern of stability across successive windows—is a classifiable descriptive layer that does not lose interpretability being transient from one domain to another. This result also suggests the benefit of lightweight, transparent analytics for situations where model interpretability and semantic preservation are important.

Even though the study is specifically designed against a fixed windowing segmentation and reduced number of statistical features, these constraints aim to provide interpretable

results as well as reproducible analysis. The method circumvents abstraction layers that could obfuscate the original signal semantics and yields a simple foundation for further extensions. Future work can expand by considering more heterogeneous data, using different ways to segment the subgraphs, perform incorporate and control integration with a higher-level analysis method while holding on to the descriptive core.

In sum, this work provides empirical support of the fact that cross-domain behavioral pattern analysis is indeed achievable by employing domain-agnostic descriptive means which are simple and interpretable. This type of methodology is especially appropriate in the context of early discovery, description and comparison of datasets from intelligent systems with heterogeneous sensing conditions.

References:

1. S. Aghabozorgi, A. Seyed Shirkhorshidi, and T. Ying Wah, "Time-series clustering – A decade review," *Inf Syst*, vol. 53, pp. 16–38, Oct. 2015, doi: 10.1016/j.is.2015.04.007.
2. E. Ghaderpour, S. D. Pagiatakis, and Q. K. Hassan, "A Survey on Change Detection and Time Series Analysis with Applications," *Applied Sciences*, vol. 11, no. 13, p. 6141, Jul. 2021, doi: 10.3390/app11136141.
3. G. Shmueli, "To Explain or to Predict?," *Statistical Science*, vol. 25, no. 3, Aug. 2010, doi: 10.1214/10-STS330.
4. B. Lim and S. Zohren, "Time-series forecasting with deep learning: a survey," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 379, no. 2194, p. 20200209, Apr. 2021, doi: 10.1098/rsta.2020.0209.
5. X. Ye, K. Sakurai, N.-K. C. Nair, and K. I.-K. Wang, "Machine Learning Techniques for Sensor-Based Human Activity Recognition with Data Heterogeneity—A Review," *Sensors*, vol. 24, no. 24, p. 7975, Dec. 2024, doi: 10.3390/s24247975.
6. S. Aghabozorgi, A. Seyed Shirkhorshidi, and T. Ying Wah, "Time-series clustering – A decade review," *Inf Syst*, vol. 53, pp. 16–38, Oct. 2015, doi: 10.1016/j.is.2015.04.007.

7. E. Ghaderpour, S. D. Pagiatakis, and Q. K. Hassan, "A Survey on Change Detection and Time Series Analysis with Applications," *Applied Sciences*, vol. 11, no. 13, p. 6141, Jul. 2021, doi: 10.3390/app11136141.
8. G. Shmueli, "To Explain or to Predict?," *Statistical Science*, vol. 25, no. 3, Aug. 2010, doi: 10.1214/10-STS330.
9. E. Keogh and S. Kasetty, "On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration," *Data Min Knowl Discov*, vol. 7, no. 4, pp. 349–371, Oct. 2003, doi: 10.1023/A:1024988512476.
10. S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and Machine Learning forecasting methods: Concerns and ways forward," *PLoS One*, vol. 13, no. 3, p. e0194889, Mar. 2018, doi: 10.1371/journal.pone.0194889.
11. B. D. Fulcher, "Feature-Based Time-Series Analysis," in *Feature Engineering for Machine Learning and Data Analytics*, CRC Press, 2018, pp. 87–116. doi: 10.1201/9781315181080-4.
12. E. Keogh and J. Lin, "Clustering of time-series subsequences is meaningless: implications for previous and future research," *Knowl Inf Syst*, vol. 8, no. 2, pp. 154–177, Aug. 2005, doi: 10.1007/s10115-004-0172-7.
13. K. Rehfeld, N. Marwan, J. Heitzig, and J. Kurths, "Comparison of correlation analysis techniques for irregularly sampled time series," *Nonlinear Process Geophys*, vol. 18, no. 3, pp. 389–404, Jun. 2011, doi: 10.5194/npg-18-389-2011.
14. Z. C. Lipton, "The mythos of model interpretability," *Commun ACM*, vol. 61, no. 10, pp. 36–43, Sep. 2018, doi: 10.1145/3233231.
15. A. Bagnall, J. Lines, A. Bostrom, J. Large, and E. Keogh, "The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances," *Data Min Knowl Discov*, vol. 31, no. 3, pp. 606–660, May 2017, doi: 10.1007/s10618-016-0483-9.

16. G. Shmueli and O. R. Koppius, “Predictive Analytics in Information Systems Research1,” *MIS Quarterly*, vol. 35, no. 3, pp. 553–572, Sep. 2011, doi: 10.2307/23042796.
17. M. Längkvist, L. Karlsson, and A. Loutfi, “A review of unsupervised feature learning and deep learning for time-series modeling,” *Pattern Recognit Lett*, vol. 42, pp. 11–24, Jun. 2014, doi: 10.1016/j.patrec.2014.01.008.
18. R. Caballero-Águila, A. Hermoso-Carazo, and J. Linares-Pérez, “Networked fusion estimation with multiple uncertainties and time-correlated channel noise,” *Information Fusion*, vol. 54, pp. 161–171, Feb. 2020, doi: 10.1016/j.inffus.2019.07.008.
19. C. C. Aggarwal, *Outlier Analysis*. Cham: Springer International Publishing, 2017. doi: 10.1007/978-3-319-47578-3.

