Fuels Demand Forecasting: Identifying Leading Feature Sets, Prediction Strategy, and Regressors

Jonas Krause, Alexandre C. A. Beiruth, Jean P. Barddal, Alceu S. Britto Jr, Vinicius M. A. Souza
Graduate Program in Informatics (PPGIa), Pontifícia Universidade Católica do Paraná (PUCPR)
Curitiba, Paraná, Brazil
{jonas.krause, alexandre.beiruth, paul.jean, alceu.junior, vinicius.mourao}@pucpr.br

***Abstract*—Fuels are crucial for any country's development and economy, impacting various sectors such as transportation, industry, and electricity generation. Accurate prediction of monthly fuel demand can improve supply chain management, strategic decision-making, and financial planning for businesses while helping governments develop decarbonization policies and estimate pollutant emissions. This paper explores machine learning models to forecast fossil fuels and biofuel demand 12 months ahead, using univariate time series data representing the historical sales of 27 Brazilian states, one of the world's leading producers and consumers of fuels. We evaluate different time series feature sets, machine learning regression models, and prediction strategies to address the complexity of fuel sales influenced by factors such as economic conditions and geopolitical events. Our comprehensive evaluation aims to determine an effective setting for predictive models in the fuel domain. Our results show that popular feature extractors for time series, such as Catch22 and TsFresh, cannot improve the original data representation for most forecasting models. Although focused on Brazil, our findings apply to other countries, since the trained models do not rely on external variables, such as micro and macroeconomic indicators.**

***Index Terms*—Fuel sales, demand forecasting, time series features, multi-step prediction strategies.**

I. Introduction

Fuels play a fundamental role in the development and economy of any country, impacting many sectors (buildings, industrial, transportation systems, and electricity) [1]. The primary fuels are refined petroleum products, including gasoline, kerosene, and diesel oil. Gasoline (or petrol) is the most prominent because of its widespread use in vehicles.

Fossil fuels are responsible for several environmental negative impacts. For example, the substances produced when gasoline is burned (carbon monoxide, carbon dioxide, nitrogen oxides, particulate matter, and unburned hydrocarbons) contribute to air pollution [2]. Thus, alternative fuels derived from renewable sources, such as ethanol, have played an essential role in the fuel chain in the last decades.

Ethanol can be made biologically from various biomass sources such as sugar cane, corn, and cassava root [3]. The use of ethanol is widespread, and more than 98% of gasoline in the United States (the largest gas producer and consumer worldwide) contains some ethanol. In addition to being less pollutant than fossil fuels, ethanol stimulates local rural production and reduces dependence on imported oil.

Predicting the monthly demand for fossil fuels and biofuels in a country or region enables businesses and producers to optimize their supply chains and inventory management, enhancing strategic decision-making. This forecasting allows better financial planning and faster responses to market fluctuations. Governments can leverage this information to develop decarbonization policies, estimate pollutant emissions, and plan transition strategies for renewable sources.

This paper investigates machine learning (ML) models to predict fuel demand 12 months ahead. Many factors, such as fuel type, region of sales, seasonal variations, economic conditions, and geopolitical events, influence fuel sales. This complexity poses a significant challenge for forecasting models, requiring methods to find complex sequential patterns.

We note that most works from the literature do not explore extracting intrinsic information from the univariate time series representing the historical demand to enrich the original representation of the data. Since no work performed a broad evaluation comparing different feature sets, deciding which features and settings lead to accurate models is still challenging.

In a comprehensive experimental evaluation involving a thousand results, our objective is to answer the following research questions for fuel demand forecasting:

- Can ML outperform statistical models such as ARIMA?
- Which feature extractors for time series, such as Catch22, TsFeatures, TSFEL, TsFlex, Cesium, and TsFresh, provide the best data representations?
- What are the best settings considering window size for feature extraction, feature set, regression model, and prediction strategy (e.g., recursive or direct)?

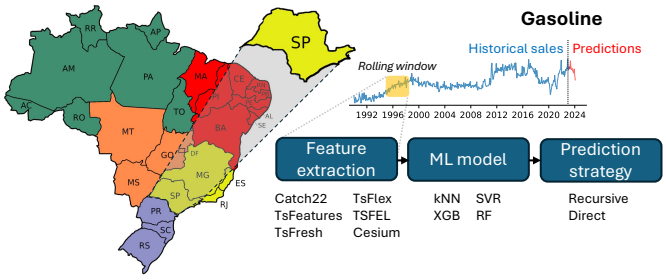


Fig. 1. General view of the fuels demand forecasting and settings evaluated. We consider six time series feature sets, four regressors, and two prediction strategies for forecasting gasoline, diesel oil, and ethanol demand in 27 states.

We conducted our experiments and analysis based on historical sales in 27 Brazilian states of the main types of fuels:

gasoline, diesel oil, and ethanol. In particular, this challenging dataset has test data beginning in March 2023, highly affected by the end of the COVID-19 pandemic phase on May 2023 [4]. Fig 1 presents a general view of our work illustrating the historical gasoline sales in the Sao Paulo state (SP) and the predictions 12 steps ahead.

Brazil is one of the largest fuel users in the world, and it has a substantial infrastructure that encompasses refining, exports, production, and supply to the population. The country is the second-largest producer of biofuels in the world, behind the United States, accounting for a global share of 20% in 2021 [5]. Ethanol represented 16.9% of the country's total energy consumption by the transport sector in 2022, while gasoline represented 27.1% and diesel oil 44.6% [6]. Therefore, predicting the demand for these fuels in Brazil ensures efficient energy management, economic stability, and meeting both domestic and international environmental goals.

Despite the importance of carrying out this prediction task with data from Brazil, we highlight that our findings are helpful in fuel forecasting in any country since our models do not depend on external variables, such as economic indicators. Besides, our findings regarding leading feature sets, prediction strategies, and regressors can guide the building of accurate models on different domains.

In summary, the main contributions of this paper are:

- We built accurate forecasting models for fossil and biofuel demand, a strategic domain that affects the economy and climate-related decisions of any country.
- We provide a comparative analysis between various feature extractors for time series that can guide researchers and practitioners dealing with similar forecasting tasks in different domains.
- We make available 81 univariate time series datasets that can be used for forecasting [7], motif discovery [8], discords [9], joins [10], and other time series tasks.

The paper is structured as follows. Section II discuss the related work; Section III presents the main concepts of time series and forecasting using ML models; Section IV introduces our datasets for fuel demand forecasting; Section V presents the results of our comprehensive evaluation; and Section VI presents our conclusions and future work.

II. Related work

In this section, we discuss the most recent related works for fuel demand prediction using time series data and works that provide an experimental evaluation comparing feature extractors for time series mining.

Ceylan et al. [11] evaluated the performance of ML models for gasoline demand in Turkey during the COVID-19 pandemic. Four models (Gaussian Process Regression, Sequential Minimam Optimization Regression, Multi-Layer Perceptron Regression, and Random Forest) were evaluated to provide daily predictions one year ahead for 2020. The models used different data as input, such as the historical gasoline demand in 2019 and additional data, such as national holidays, gasoline prices, and COVID-19-related factors.

Yu et al. [12] propose a decomposition ensemble model to predict the gasoline demand in China one trimester ahead. The main idea is to decompose a time series on multiple scales to reduce the difficulty of modeling using wavelet decomposition. Then, an SVR model predicts the future values of each component, and then the predictions of each component are summed to reconstruct the time series. The model uses the gasoline consumption data of 30 provincial administrative units as input. The data from the first quarter of 2010 to the fourth quarter of 2016 are used as the training data (28 observations) to determine the model parameters. The data from the first quarter of 2017 to the fourth quarter of 2018 (8 observations) are used to evaluate the performance of the forecasting model.

Rodrigues et al. [13] consider the same data source of our work to predict the consumption of light fuels (fuel used by vehicles with internal combustion engines) in Brazil. The evaluation considers eight methods, from statistical (ARIMA and SARIMA) to ML models (artificial neural networks, support vector regression, and random forest), for a forecast horizon of 3, 6, and 12 months. When using ML models, the authors propose to include two external variables: GDP per capita and the average price of fuels to the final consumer. The experimental evaluation considers a single time series representing the sum of three fuels (gasoline, ethanol, and natural gas) over all regions instead of predicting the demand by fuel type and region, as we did in our work.

All the works discussed above and most of the literature do not explore extracting relevant information from the univariate time series representing the historical sales to enrich the original representation of the data, as performed in similar domains such as signal processing [14]. In addition, since no work performed a broad evaluation comparing different feature sets, deciding which features lead to accurate models in forecasting problems is still a challenge.

Henderson and Fulcher [15] evaluated seven feature extractors for time series: Catch22, hctsa, TsFeatures, TsFresh, Kats, TSFEL, and Feats. However, the authors compared their computational speed, assessed the redundancy of features in each set, and evaluated the overlap and redundancy across different feature sets. Thus, no comparison concerning the impact of these extractors on the predictive performance of different models was discussed, as presented in our work.

III. Background

The historical fuel sales are represented by univariate time series. A time series $T = (t_1, t_2, \ldots, t_n)$ is an ordered sequence of n real-values t_i measured at equal intervals (e.g., monthly), in which t_i represents a value observed at time i.

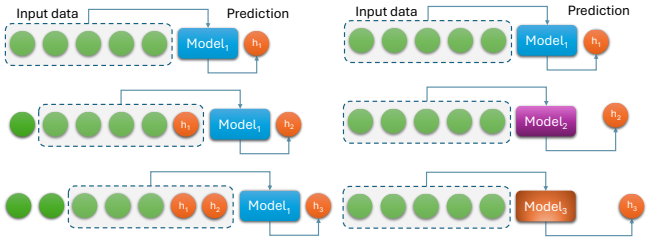
Forecasting models receive a series T as input and aim to predict the following h values as output. In this work, we consider a prediction horizon of 12 months ($h = 12$). When $h > 1$, we can consider different strategies to provide a multi-step prediction using an ML model. In the following, we discuss two of them evaluated in this work.

A. Multi-step prediction strategies

When the goal of a forecasting model is to predict only the next value of a series (i.e., $h = 1$), we call the task of *one-step prediction*. However, most of the time, we are interested in the values of multiple steps ahead. There are two main strategies for multi-step forecasting:

- **Recursive**, in which we call a one-step model multiple times. In this case, the prediction for the prior time step is used as input to predict the following time step;
- **Direct**, in which we train an individual model for each forecast step.

Fig. 2 illustrates both strategies considering a prediction horizon of 3 steps ($h = 3$). In this example, the same one-step ahead model is invoked three times for the recursive strategy (*left*). We call three models trained with the same input variables and varying targets in the direct strategy (*right*). The first model learned to predict the next value; the second model learned to predict the second value, and so forth.



(a) Recursive (b) Direct

Fig. 2. Multi-step prediction strategies considering $h = 3$.

B. Feature extraction

The success of ML algorithms is highly dependent on informative data descriptions to represent a problem. We obtain these descriptions through feature extraction, a process in which a method transforms a weakly representative input space into a strongly representative subspace with relevant information [16].

Due to the inherent characteristics of time series data, numerous software packages extract features considering their particularities. In general, these software extract information that may be relevant for different domains, including summaries of the distribution of values (e.g., Gaussianity, standard deviation, zero-crossing rate, properties of outliers), autocorrelation structure (power spectral measures, mutual information), stationarity (how properties change over time), information theoretic measures of entropy and temporal predictability [17].

Choosing the extractor that leads to the best feature set is challenging and depends on the domain and the learning task. We evaluated six popular software packages to identify which provides accurate forecasting models for our domain and whether these features can outperform the raw data without feature extraction. Table 1 shows the number of features obtained by each extractor.

TABLE I
NUMBER OF FEATURES GENERATED BY THE FEATURE EXTRACTORS.

Feature extractor	Number of features
Catch22[1] [18]	22
TsFeatures[2] [19]	42
TSFEL[3] [20]	140
TsFresh[4] [21]	10,178
TsFlex[5] [22]	14
Cesium[6] [23]	118

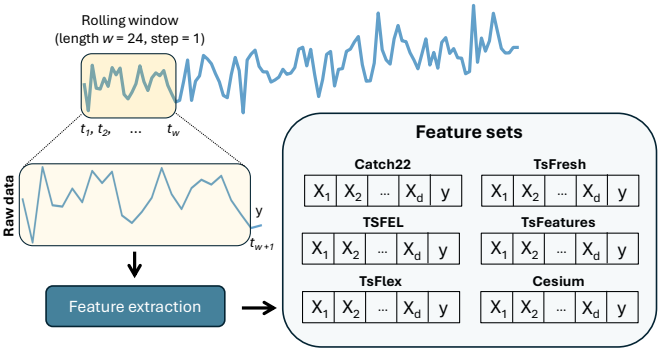


Fig. 3. General feature extraction process from time series data using a rolling window and different extractors. These feature sets are used as input for training ML regressors.

Given a time series with length n as input, we slide a rolling window with a reduced length w (e.g., 12, 24, or 36) to extract features from the subsequence within the window. This window slides through all the historical sales data, generating a new example $X = (X_1, X_2, \ldots, X_d, y)$ with d features and a y value representing a future observation. The number of features d depends on the extractor, as shown in Table I. In each step, the window moves towards the following observation (i.e., step = 1), repeating the feature extraction in the updated window. Fig. 3 illustrates the entire process.

In addition, we also consider training the ML regressors with the subsequence of observations in the window as features, i.e., before the extraction. We call this feature set *raw data* throughout the paper. In all cases, we normalize the extracted features to have zero mean and unit standard deviation (z-normalization) [24].

IV. Data Collection, Preprocessing and Analysis

To carry out the experiments and analyses, we collected publicly available data from the Brazilian National Agency for Petroleum, Natural Gas and Biofuels (ANP)[7]. The ANP portal provides historical data on monthly sales of different petroleum derivatives and biofuels over the last 34 years (from January 1990 to February 2024) in 27 Brazilian states.

[1]https://github.com/DynamicsAndNeuralSystems/catch22
[2]https://pypi.org/project/tsfeatures/
[3]https://tsfel.readthedocs.io/
[4]https://tsfresh.readthedocs.io/
[5]https://github.com/predict-idlab/tsflex/
[6]https://cesium-ml.org/
[7]https://www.gov.br/anp/pt-br/centrais-de-conteudo/dados-abertos/vendas-de-derivados-de-petroleo-e-biocombustiveis

We selected data from fuels that most impact the country's economy (diesel oil, gasoline, and ethanol), forming a dataset with 81-time series with 410 observations. We remove duplicated entries and outlier observations (3σ rule [25]) and impute missing values (spline interpolation [26]). The cleaned and selected datasets are available on our supporting website [27].

In Fig 4, we show the time series representing the historical sales since 1990 of the three fuels investigated in this work. This series represents sales in Sao Paulo, one of the most relevant regions in consumption and sales.

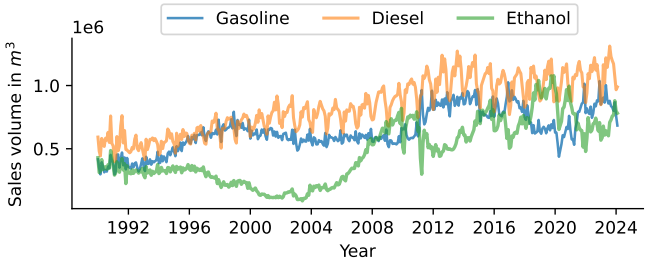


Fig. 4. Monthly fuel sales volume in the state of Sao Paulo since 1990.

The seasonality of sales varies depending on the fuel. In all cases, they have a pattern within 12 months. In Fig. 5, we show the seasonal component of each fuel in the last five years obtained by the series decomposition process into components. We note a similar behavior between gasoline and ethanol, with well-defined peaks in December, October, and March and valleys mainly in February, January, and November. This correlation between both fuels is expected since the gasoline sold in Brazil has 27% ethanol [28] and the country owns the largest market of flex-fuel vehicles capable of running on gasoline and ethanol in any proportion [29].

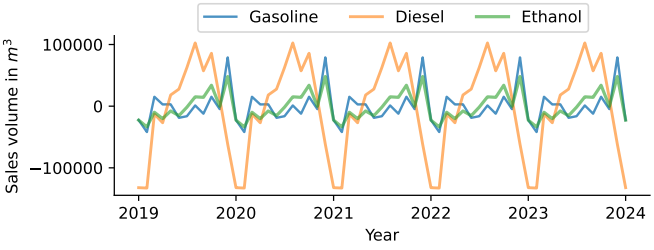


Fig. 5. Seasonal fuel sales behavior in a period of five years.

For diesel, peaks are observed in August and October, whereas a significant reduction is observed in December and January. In this case, the peaks and valleys are justified by the use of diesel-powered machinery by rural producers, such as tractors and planters, during intense production months and off-season. It should be noted that in Brazil, passenger cars must be fueled only by gasoline or ethanol since diesel engines are not allowed due to environmental regulations.

V. Experimental Evaluation

Our analysis is based on $13,608$ results (27 states $\times$ 3 fuels $\times$ 4 regressors $\times$ 7 feature sets $\times$ 2 prediction strategies $\times$ 3 window sizes). Additional results, figures, and datasets are available on our supporting website [27].

A. Setup

Each time series with 410 observations representing the fuel sales of a state is split into training (first 398 observations) and test (last 12 observations), i.e., a holdout sample validation.

Since the range of sales values varies significantly between states (even for the same fuel), we consider error measures independent of units for a fair comparison. We use two measures: mean absolute percentage error (MAPE) and prediction of change in direction (POCID), defined according to Eq. 1 and 2, respectively. In these equations, y_i represents the actual value of the i-th observation of the test series, $\hat{y}_i$ the predicted value, and h the number of predictions (i.e., a horizon of 12 observations). We consider a small ϵ value to avoid division by zero in MAPE.

$$MAPE = \frac{1}{h}\sum_{i=1}^{h}\frac{|y_i - \hat{y}_i|}{max(\epsilon, |y_i|)} \tag{1}$$

$$POCID = \frac{\sum_{i=2}^{h} D_i}{h-1} \times 100, \quad D_i = \begin{cases} 1 : (\hat{y}_i - \hat{y}_{i-1})(y_i - y_{i-1}) > 0 \\ 0 : \text{otherwise} \end{cases} \tag{2}$$

While MAPE indicates the error rate, POCID measures the percentage of times the model correctly predicts the forecast direction (i.e., increase, decrease, or stable) compared to the previous time step. Hence, precise models show MAPE values close to 0 and POCID values close to 100.

We evaluated the performance of four base learners considering direct and recursive strategies: Random Forest Regressor (RF), k-Nearest Neighbors (kNN), eXtreme Gradient Boosting (XGB), and Support Vector Regressor (SVR). For all learners, we consider the default values for the hyperparameters, as described in Table II.

TABLE II
Machine learning regressors and parameters.

ID	Regressor	Parameters
RF	Random Forest	100 trees with unlimited depth
kNN	k-Nearest Neighbour	Euclidean dist., $k = 3$
XGB	eXtreme Gradient Boosting	DART, $\eta = 0.3, \gamma = 0$
SVR	Support Vector Regression	RBF, $C = 1, d = 3, \epsilon = 0.1$

The results of ML models are compared against the statistical model ARIMA ($p = 12, d = 1, q = 6$), defined according to the ACF and PACF analysis.

B. Best machine learning settings by fuel type

The forecast results depend on four main variables: i) feature set, ii) ML regressor, iii) prediction strategy, and iv) window size for feature extraction. Table III presents a ranking with the best MAPE results to predict gasoline demand that

TABLE III
TOP-5 BEST SETTINGS FOR GASOLINE DEMAND FORECASTING.

Rank	Feature set	Regressor	Strategy	Window	MAPE	POCID
1	Raw data	SVR	Recursive	36	0.0452	77.77
2	Raw data	SVR	Recursive	24	0.0489	72.72
3	Raw data	kNN	Recursive	24	0.0534	74.74
4	Raw data	RF	Recursive	12	0.0543	68.01
5	TsFlex	SVR	Recursive	12	0.0579	54.20
				ARIMA	0.0636	67.67

helps to choose the values of these variables. The results show the average of MAPE and POCID in 27 states.

Table IV presents the results for diesel oil demand forecast. For both gasoline and diesel oil, raw data as a feature set, a recursive prediction strategy, and a window with 36 observations lead to the best results. An SVR model presents the best gasoline results, while an RF model is the most accurate for diesel oil. The ML approach outperforms the statistical baseline (ARIMA) in MAPE and POCID.

TABLE IV
TOP-5 BEST SETTINGS FOR DIESEL OIL DEMAND FORECASTING.

Rank	Feature set	Regressor	Strategy	Window	MAPE	POCID
1	Raw data	RF	Recursive	36	0.0742	72.05
2	Raw data	kNN	Recursive	36	0.0762	73.40
3	Raw data	XGB	Recursive	12	0.0767	66.32
4	Raw data	RF	Recursive	12	0.0783	69.36
5	Raw data	RF	Recursive	24	0.0789	73.06
				ARIMA	0.0758	71.38

Table V presents the results for ethanol. In this case, a different setting presents the best results, with TSFEL as a feature set and direct prediction strategy. When considering a window with 36 observations and an SVR model, the results outperform ARIMA for MAPE and POCID measures.

TABLE V
TOP-5 BEST SETTINGS FOR ETHANOL DEMAND FORECASTING.

Rank	Feature set	Regressor	Strategy	Window	MAPE	POCID
1	TSFEL	RF	Direct	24	0.2598	47.47
2	TSFEL	SVR	Direct	36	0.2614	63.63
3	TSFEL	kNN	Direct	36	0.2633	57.23
4	TsFlex	SVR	Direct	36	0.2634	57.91
5	TSFEL	XGB	Direct	36	0.2635	53.87
				ARIMA	0.2787	56.90

When analyzing the results for each fuel type, we note that ethanol is the most challenging, with MAPE around 0.26 and POCID around 55%, while we can obtain a MAPE of 0.04 and POCID of 77% for gasoline. The drastic change in the sales volume of ethanol justifies this difficulty. Such a change occurred mainly due to the price decrease of around 15% in this period [30]. For example, 15 of the 27 Brazilian states had ethanol as the most economically favorable fuel, influencing the decision of a consumer who owns a flex-fuel car.

Fig. 6 shows the predictions obtained by the best models in the Sao Paulo state for ethanol demand, a major producer and one of the largest consumers of biofuels in the world [31]. In this figure, we also show the historical sales one year before the prediction time (dotted vertical line), in which we can note a significant increase in the sales volume and sales behavior quite differently from the last year. In this case, the models with the best MAPE predicted a higher sales volume with slight variations, which causes them to have a lower POCID.

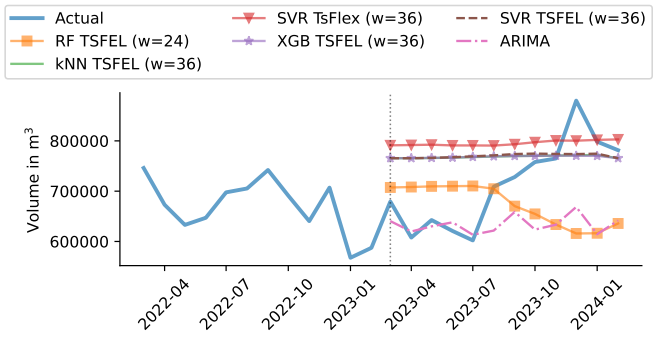


Fig. 6. Predictions of ethanol demand in the Sao Paulo state. For this fuel, all forecasters have difficulty due to the drastic change in the sales volume.

Fig. 7 shows the predictions of the most accurate forecaster for gasoline demand in the state of Sao Paulo. In this case, most models can reproduce the expected time series shape. However, we remember that the end of the COVID-19 pandemic was declared in May 2023 [4], which may have affected the sales behavior of all fuels in this period.

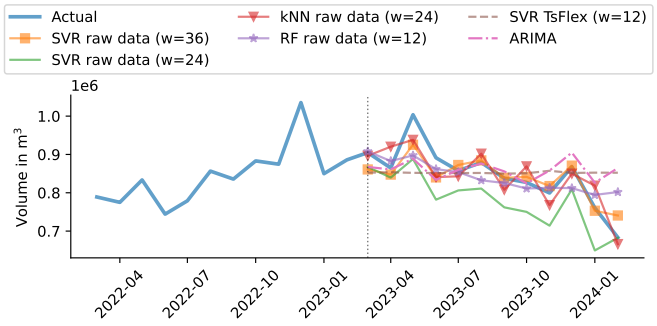


Fig. 7. Predictions of gasoline demand in the Sao Paulo state.

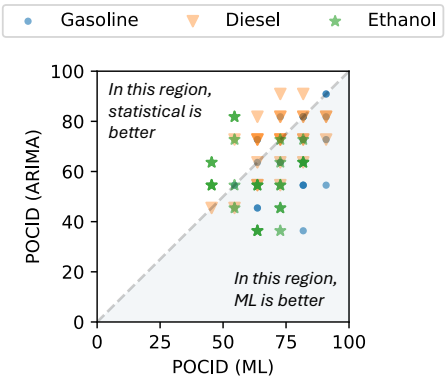


Fig. 8. Comparison between ML and ARIMA results.

Fig. 8 compares the best ML settings against the statistical baseline, i.e., an SVR model using raw data and recursive prediction ($w = 36$) for gasoline, an RF model using raw data and recursive prediction ($w = 36$) for diesel oil, and an SVR model using TSFEL and direct prediction ($w = 36$) for ethanol. In this figure, each point represents the POCID result of the ML and statistical model in the same state. The points below the main diagonal represent cases where the ML approach outperformed the baseline statistical model.

The comparative reveals that from the 81 results (27 states and three fuels), ML won 43 cases (53%), the statistical method won 20 (25%), and there were 18 draws (22%).

C. Impact of window size

An essential parameter of the ML pipeline based on feature extraction is the window size w in which we extract the features. While small windows can extract local patterns observed in sequential months, they might miss broader contextual information, such as correlations over the years. Therefore, we evaluated rolling windows with small ($w = 12$), medium ($w = 24$), and large ($w = 36$) sizes. Fig. 9 presents the distribution of MAPE results in the 27 states. We choose the best regressor and prediction strategy for each fuel.

From Fig. 9, we note that a window size with 12 observations leads to the best MAPE results for most feature sets obtained by extractors, mainly for gasoline and diesel oil. When using the raw data as features, a larger window with 36 observations leads to the best results for gasoline that outperforms all feature sets. A window with 24 observations for ethanol forecast leads to the best average results for most feature sets.

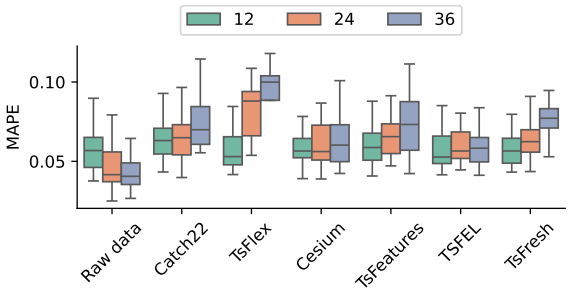
D. Impact of multi-step strategies and regressors

Table VI presents the average MAPE and POCID results of direct and recursive prediction strategies. This table compares the seven feature sets and four ML regressors for gasoline demand forecasts considering $w = 12$. Given a feature set and a regressor, the best result between the direct and recursive strategy is highlighted in bold.

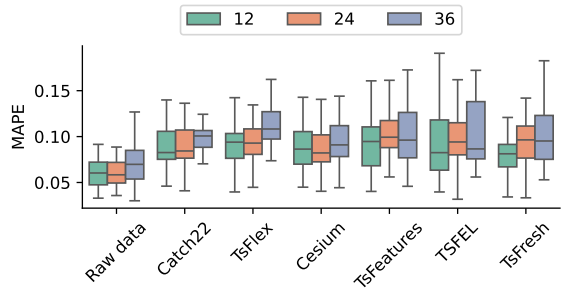
Although Table VI shows superior MAPE results obtained by the recursive strategy for most regressors and feature sets, it is important to note that such behavior is only evident for gasoline. For example, Fig. 10 compares recursive and direct prediction strategies in the MAPE results for diesel oil and ethanol demand.

A recursive strategy is recommended for diesel oil when using raw data or TsFresh as feature sets. For other feature sets, the direct prediction strategy leads to better results. For ethanol, a direct strategy leads to better general results, in which a model using Catch22, TsFeatures, TSFEL, and TsFresh as feature sets outperform the results of the recursive strategy.

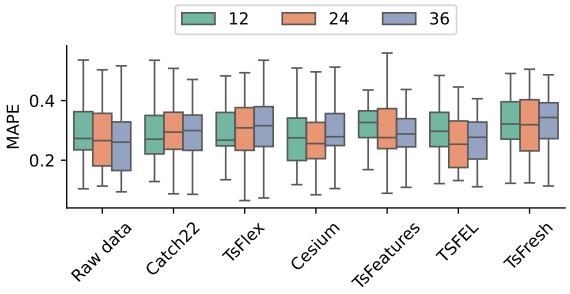
In Fig. 11, we compare the performance of multiple regressors for diesel oil forecast considering the recursive and direct prediction strategies using various feature sets ($w = 12$). In general, SVR and RF were the most accurate regressors.



(a) Gasoline (SVR model and recursive prediction)



(b) Diesel oil (RF model and recursive prediction)



(c) Ethanol (RF model and direct prediction)

Fig. 9. Average MAPE results for gasoline, diesel oil, and ethanol forecasting considering window sizes of 12, 24, and 36 observations.

E. Comparing feature sets from extractors

In order to identify the feature extractors that lead to the most accurate models, we performed a pairwise comparison between the results obtained by each feature extractor overall settings of our experimental evaluation, i.e., we compare the results of a feature set A with a feature set B considering the results obtained in the 27 states, three fuels, three window sizes, two prediction strategies and four regressors (1,944 results). Given the same setting and two different feature sets, we count the number of wins of each extractor. Table VII shows the number of wins/losses when comparing each pair of features.

From Table VII, we note that raw data obtained more wins than other feature sets in all paired comparisons. Taking into account only the six feature extractors, Cesium obtained more wins than all other extractors, followed by TsFlex and Catch22.

TABLE VI
COMPARISON OF DIRECT AND RECURSIVE PREDICTION STRATEGIES FOR GASOLINE DEMAND PREDICTION ($w = 12$).

Feat.	Regr.	MAPE		POCID	
		Recursive	Direct	Recursive	Direct
Raw data	RF	**0.0543**	0.0791	**68.01**	39.73
	kNN	**0.0624**	0.0876	**74.07**	40.06
	XGB	**0.0665**	0.0876	**61.27**	40.74
	SVR	**0.0647**	0.0792	**71.04**	39.39
Catch22	RF	**0.0765**	0.0782	35.69	**50.16**
	kNN	0.0841	**0.0819**	42.08	**50.50**
	XGB	**0.0773**	0.0835	44.78	**51.51**
	SVR	**0.0674**	0.0758	36.02	**54.20**
TsFresh	RF	**0.0591**	0.0875	**35.35**	32.32
	kNN	**0.0585**	0.0927	24.57	**44.44**
	XGB	**0.0640**	0.0909	20.53	**31.31**
	SVR	**0.0590**	0.0813	**60.60**	46.80
TSFEL	RF	**0.0639**	0.0847	**49.83**	42.08
	kNN	**0.0711**	0.0884	30.97	**45.45**
	XGB	**0.0822**	0.0898	**49.15**	43.09
	SVR	**0.0580**	0.0789	**42.08**	39.05
TsFeats	RF	**0.0656**	0.0788	45.45	**54.20**
	kNN	**0.0716**	0.0841	**52.18**	49.83
	XGB	**0.0747**	0.0796	51.17	**55.89**
	SVR	**0.0612**	0.0792	**54.20**	45.79
TsFlex	RF	**0.0645**	0.0743	45.11	**58.24**
	kNN	**0.0645**	0.0807	45.11	**59.93**
	XGB	**0.0742**	0.0784	43.09	**54.20**
	SVR	**0.0579**	0.0788	**54.20**	39.05
Cesium	RF	**0.0617**	0.0778	44.10	**46.46**
	kNN	**0.0800**	0.0871	32.32	**49.15**
	XGB	**0.0738**	0.0845	36.02	**44.78**
	SVR	**0.0594**	0.0785	45.79	**47.47**

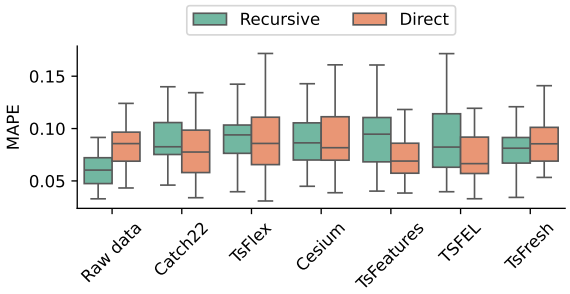
TABLE VII
ONE-VS-ONE COMPARISON BETWEEN FEATURE SETS. THE VALUES REPRESENT THE NUMBER OF WINS/LOSSES GIVEN THE MAPE RESULTS.

Feature set	Catch22	TsFeats	TSFEL	TsFresh	TsFlex	Cesium
Raw data	1554/876	1557/873	1571/859	1606/824	1432/998	1427/1003
Catch22	–	1234/1196	1218/1212	1225/1205	1137/1293	1041/1398
TsFeats	–	–	1195/1235	1217/1213	1118/1312	1041/1398
TSFEL	–	–	–	1240/1190	1309/1121	1093/1337
TsFresh	–	–	–	–	1163/1267	1075/1355
TsFlex	–	–	–	–	–	1186/1244

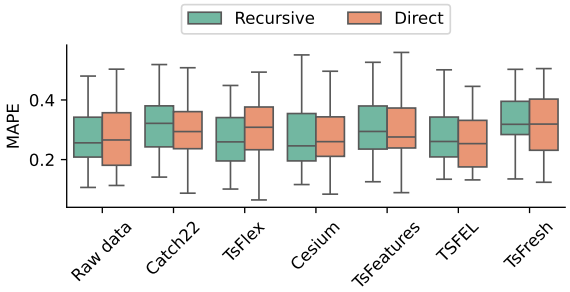
F. Key findings

The main findings of our comprehensive experimental evaluation are summarized in the following:

- Ethanol was the most challenging fuel to predict future demand, being highly affected by gasoline prices and the end of the COVID-19 pandemic phase.
- ML models outperformed the statistical ARIMA model for all fuels evaluated.
- Using raw data with a larger window generally leads to more accurate models than those trained with features extracted by software packages.
- For most feature extractors, using a small or medium window (e.g., 12 or 24 observations) to extract the features leads to a better data representation than oversized windows with 36 observations.
- Of the six feature extractors evaluated, Cesium was the most effective, followed by TsFlex and Catch22, considering the different settings regarding window size, prediction strategy, and regressors.
- A recursive prediction strategy is recommended for gasoline, and a direct prediction strategy is recommended for forecasting diesel oil while using features extracted by software packages. For ethanol, the difference between

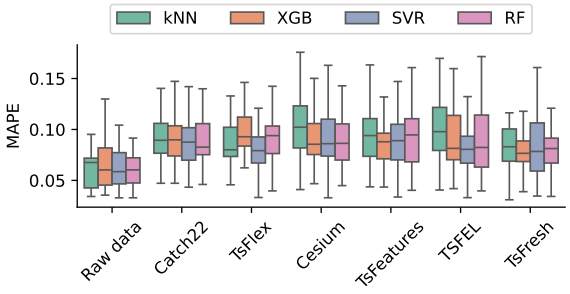


(a) Diesel oil (RF, $w = 12$)

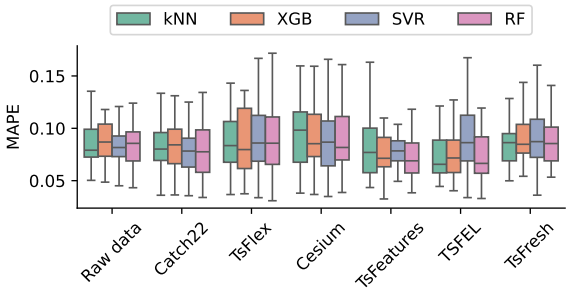


(b) Ethanol (RF, $w = 24$)

Fig. 10. Comparison between recursive and direct prediction strategies.



(a) Recursive



(b) Direct

Fig. 11. Comparison between ML regressors for diesel oil.

the prediction strategy is not significant.
- SVR and RF were the most accurate of the four ML regressors evaluated.

VI. Conclusions and Future work

We investigated univariate ML models for time series forecasting in the complex fuel demand domain. Fuels are essential for any country's development and economy, and forecasting their demand benefits the government, energy industry, and society. However, fuel sales are affected by economic variables, seasonal variations, and geopolitical events. Specifically, we consider a challenger dataset with test data beginning in March 2023, affected by the end of the COVID-19 pandemic phase, declared by the World Health Organization (WHO) on 5 May 2023 [4].

Our evaluation shows that univariate ML models obtain accurate results, not depending on external information (e.g., economic indexes or political event monitoring) that impact fuel demand. These models trained with a simple feature set as raw data outperformed a strong statistical baseline (ARIMA). Not depending on external information makes it possible to build and use these models in a production environment efficiently and at a low cost. If the user wants to extract intrinsic features from the time series, we suggest beginning to evaluate Cesium and TsFlex.

In future work, we plan to extend our evaluation to other strategic energy sources, such as kerosene, fuel oil, and liquefied petroleum gas. We will also investigate subsets of features of the extractors that outperform the current results.

References

[1] S. Bilgen, "Structure and environmental impact of global energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 38, pp. 890–902, 2014.

[2] H. Herzog, B. Eliasson, and O. Kaarstad, "Capturing greenhouse gases," *Scientific American*, vol. 282, no. 2, pp. 72–79, 2000.

[3] P. Iodice, A. Senatore, G. Langella, and A. Amoresano, "Advantages of ethanol–gasoline blends as fuel substitute for last generation si engines," *Environmental Progress & Sustainable Energy*, vol. 36, no. 4, pp. 1173–1179, 2017.

[4] R. Sarker, A. Roknuzzaman, Nazmunnahar, M. Shahriar, M. J. Hossain, and M. R. Islam, "The who has declared the end of pandemic phase of covid-19: Way to come back in the normal life," *Health science reports*, vol. 6, no. 9, p. e1544, 2023.

[5] U.S. Energy Information Administration (EIA), "Biofuels production 2022," Accessed June 12, 2024, https://www.eia.gov/international/rankings/country/BRA?pa=28&u=2&f=A&v=none&y=01%2F01%2F2021&ev=false.

[6] Energy Research Office (EPE), "National energy balance summary report," 2023. [Online]. Available: https://www.gov.br/mme/pt-br/assuntos/secretarias/sntep/publicacoes/balanco-energetico-nacional/ben-2023/relatorio-sintese

[7] A. R. S. Parmezan, V. M. A. Souza, and G. Batista, "Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model," *Information sciences*, vol. 484, pp. 302–337, 2019.

[8] A. Mueen, E. Keogh, Q. Zhu, S. Cash, and B. Westover, "Exact discovery of time series motifs," in *SIAM international conference on data mining*. SIAM, 2009, pp. 473–484.

[9] D. Yankov, E. Keogh, and U. Rebbapragada, "Disk aware discord discovery: Finding unusual time series in terabyte sized datasets," *Knowledge and Information Systems*, vol. 17, pp. 241–262, 2008.

[10] M. P. Mollah, V. M. A. Souza, and A. Mueen, "Multi-way time series join on multi-length patterns," in *IEEE international conference on data mining*. IEEE, 2021, pp. 429–438.

[11] Z. Ceylan, D. Akbulut, and E. Baytürk, "Forecasting gasoline consumption using machine learning algorithms during covid-19 pandemic," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, pp. 1–19, 2022.

[12] L. Yu, Y. Ma, and M. Ma, "An effective rolling decomposition-ensemble model for gasoline consumption forecasting," *Energy*, vol. 222, p. 119869, 2021.

[13] L. Rodrigues, L. Rodrigues, and M. Bacchi, "Performance evaluation of forecasting models based on time series and machine learning techniques: an application to light fuel consumption in brazil," *International Journal of Energy Sector Management*, vol. 16, no. 4, pp. 636–658, 2022.

[14] V. M. A. Souza, E. A. Cherman, R. G. Rossi, and R. A. Souza, "Towards automatic evaluation of asphalt irregularity using smartphone's sensors," in *International Symposium on Intelligent Data Analysis*. Springer, 2017, pp. 322–333.

[15] T. Henderson and B. D. Fulcher, "An empirical evaluation of time-series feature sets," in *International Conference on Data Mining Workshops*. IEEE, 2021, pp. 1032–1038.

[16] N. Chumerin and M. M. Van Hulle, "Comparison of two feature extraction methods based on maximization of mutual information," in *IEEE signal processing society workshop on machine learning for signal processing*. IEEE, 2006, pp. 343–348.

[17] B. D. Fulcher and N. S. Jones, "hctsa: A computational framework for automated time-series phenotyping using massive feature extraction," *Cell systems*, vol. 5, no. 5, pp. 527–531, 2017.

[18] C. H. Lubba, S. S. Sethi, P. Knaute, S. R. Schultz, B. D. Fulcher, and N. S. Jones, "catch22: Canonical time-series characteristics: Selected through highly comparative time-series analysis," *Data Mining and Knowledge Discovery*, vol. 33, no. 6, pp. 1821–1852, 2019.

[19] R. J. Hyndman, E. Wang, and N. Laptev, "Large-scale unusual time series detection," in *IEEE international conference on data mining workshop*. IEEE, 2015, pp. 1616–1619.

[20] M. Barandas, D. Folgado, L. Fernandes, S. Santos, M. Abreu, P. Bota, H. Liu, T. Schultz, and H. Gamboa, "Tsfel: Time series feature extraction library," *SoftwareX*, vol. 11, p. 100456, 2020.

[21] M. Christ, N. Braun, J. Neuffer, and A. W. Kempa-Liehr, "Time series feature extraction on basis of scalable hypothesis tests (tsfresh–a python package)," *Neurocomputing*, vol. 307, pp. 72–77, 2018.

[22] J. Van Der Donckt, J. Van Der Donckt, E. Deprost, and S. Van Hoecke, "tsflex: Flexible time series processing & feature extraction," *SoftwareX*, vol. 17, p. 100971, 2022.

[23] B. Naul, S. van der Walt, A. Crellin-Quick, J. S. Bloom, and F. Pérez, "cesium: Open-source platform for time-series inference," in *SciPy*, 2016.

[24] F. T. Lima and V. M. A. Souza, "A large comparison of normalization methods on time series," *Big Data Research*, vol. 34, p. 100407, 2023.

[25] A. Blázquez-García, A. Conde, U. Mori, and J. A. Lozano, "A review on outlier/anomaly detection in time series data," *ACM computing surveys*, vol. 54, no. 3, pp. 1–33, 2021.

[26] M. Lepot, J.-B. Aubin, and F. H. Clemens, "Interpolation in time series: An introductive overview of existing methods, their performance criteria and uncertainty assessment," *Water*, vol. 9, no. 10, p. 796, 2017.

[27] anonymous, "Supporting website," July 2024, https://sites.google.com/view/featuresets.

[28] N. A. Policarpo, C. Silva, T. F. A. Lopes, R. dos Santos Araújo, F. S. Á. Cavalcante, C. S. Pitombo, and M. L. M. de Oliveira, "Road vehicle emission inventory of a brazilian metropolitan area and insights for other emerging economies," *Transportation Research Part D: Transport and Environment*, vol. 58, pp. 172–185, 2018.

[29] F. Frutuoso, C. Alves, S. Araújo, D. Serra, A. Barros, F. Cavalcante, R. Araújo, N. Policarpo, and M. Oliveira, "Assessing light flex-fuel vehicle emissions with ethanol/gasoline blends along an urban corridor: a case of fortaleza/brazil," *International Journal of Transportation Science and Technology*, vol. 12, no. 2, pp. 447–459, 2023.

[30] Ministry of Mines and Energy, "Monthly Energy Bulletin Brazil - Edition December 23," Accessed July 23, 2024, 2023, https://www.gov.br/mme/pt-br/assuntos/secretarias/sntep/publicacoes/boletins-mensais-de-energia/boletins/2023-1/ingles/brazilian-monthly-energy-bulletin-december-2023/@@download/file.

[31] D. H. Capitani, L. E. Gaio, F. L. Mattos, R. Franco da Silveira, and J. C. Cruz, "Corn ethanol expansion in brazil: Are volatility interconnectedness changing?" in *2024 Annual Meeting, July 28-30, New Orleans, LA*, no. 343612. Agricultural and Applied Economics Association, 2024.