

A Service of



Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Aldabbagh, Ahmad M.; Economou, Andreas; Chariton, Christou

## Research Report Forecasting global oil demand: Application of machine learning techniques

**OIES** Paper

**Provided in Cooperation with:** The Oxford Institute for Energy Studies, Oxford

*Suggested Citation:* Aldabbagh, Ahmad M.; Economou, Andreas; Chariton, Christou (2024) : Forecasting global oil demand: Application of machine learning techniques, OIES Paper, The Oxford Institute for Energy Studies, Oxford

This Version is available at: https://hdl.handle.net/10419/306644

#### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

#### Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



## WWW.ECONSTOR.EU



November 2024

# Forecasting Global Oil Demand: Application of Machine Learning Techniques



Ahmad M. Aldabbagh, Visiting Research Fellow OIES Andreas Economou, Head of Oil Research OIES Christou Chariton, Research Associate OIES



## Abstract

This study introduces a novel approach to predicting global oil demand by integrating machine learning (ML) techniques to forecast consumption across seven refined oil products and seven key regions. By aggregating these forecasts, we offer a comprehensive view of global demand trends. The paper examines the efficacy of ML models in providing robust and accurate demand forecasts. It also provides a transparent and repeatable process to forecast oil demand. A comparison between the extreme gradient boosting (XGBoost) model and Neural Hierarchical Interpolation for Time Series Forecasting (N-HiTS) model was conducted to determine which is a more accurate model to forecast demand. Our comparative analysis demonstrates that N-HiTS performs better. The accuracy of global oil demand forecasts is pivotal for economic planning and policy making.

The contents of this paper are the authors' sole responsibility. They do not necessarily represent the views of the Oxford Institute for Energy Studies or any of its Members.



## 1. Introduction

Accurate forecasting of oil demand is critical for strategic planning. Traditional econometric models, while useful, often struggle to capture the complex dynamics influenced by numerous economic indicators and geopolitical factors. These models typically rely on linear assumptions and may not adequately address the non-linear relationships inherent in oil markets. To address these challenges, our paper introduces a methodology that leverages advanced machine learning (ML) techniques, specifically extreme gradient boosting (XGBoost) and Neural Hierarchical Interpolation for Time Series Forecasting (N-HiTS), to enhance the precision and reliability of oil demand forecasts.

Machine learning models have demonstrated superior performance in various forecasting tasks due to their ability to handle large datasets and uncover intricate patterns. Recent studies consistently show that ML techniques outperform traditional econometric methods in time series forecasting by modelling complex, nonlinear relationships and handling large datasets. For instance, Hopp (2022)<sup>1</sup> found that long short-term memory (LSTM) neural networks provided better predictive accuracy than Bayesian vector autoregressions (BVAR) for nowcasting US quarterly GDP growth, especially during economic crises. Deb (2019)<sup>2</sup> highlighted that models exploiting heterogeneity, such as finite mixture models, vielded more accurate healthcare spending forecasts compared to generalized linear models and loglinear regression. Similarly, Lukong et al. (2022)<sup>3</sup> showed that long short-term memory recurrent neural networks (LSTM-RNN) models achieved significantly lower mean absolute percentage error (MAPE) in long-term electricity load forecasting than linear regression models. In financial time series forecasting, Liu et al. (2023)<sup>4</sup> reported that ensemble methods like Random Forest and LSTM outperformed traditional econometric models in both accuracy and interpretability. Additionally, Kontopoulou et al. (2023)<sup>5</sup> reviewed various applications and concluded that ML algorithms generally surpassed autoregressive integrated moving average (ARIMA) models, particularly in capturing intricate data patterns, with hybrid models proving most effective. Furthermore, comparative analyses by Oukhouya and El Himdi (2023)<sup>6</sup> also illustrate the superior performance of support vector regression (SVR), XGBoost, LSTM, and multilaver perceptron (MLP) in stock market forecasting, with ML models generally outperforming their econometric counterparts. This superiority is attributed to ML models' ability to learn from data without relying on pre-defined assumptions, allowing them to capture more nuanced and complex relationships. These findings underscore the enhanced accuracy, efficiency, and flexibility of ML models in time series forecasting across diverse domains.

In the context of oil demand forecasting, the integration of ML techniques has shown significant improvements over traditional methods. Studies by Zhu (2023)<sup>7</sup> and Alkhammash et al. (2022)<sup>8</sup> have validated the effectiveness of ML models in this domain, demonstrating their superior performance in capturing complex patterns in data. Zhu (2023) conducted an Al-based analysis incorporating both endogenous and exogenous factors, finding that machine learning models significantly improve forecasting accuracy compared to traditional models. Similarly, Alkhammash et al. (2022) used optimized multivariate adaptive regression splines (LR-MARS) to predict crude oil demand in Saudi Arabia, showcasing the adaptability and precision of ML models in dynamic environments.

XGBoost<sup>9</sup>, a gradient boosting algorithm, is known for its robustness and efficiency. It has been successfully applied in domains such as temperature forecasting and precipitation prediction, highlighting its versatility and effectiveness across different predictive tasks. For instance, Singh and Rawat (2023)<sup>10</sup> conducted a comparative analysis of XGBoost with other ML models like support vector machine (SVM) and Random Forest in temperature forecasting, emphasizing the importance of model selection based on specific task requirements. Similarly, Dong et al. (2023)<sup>11</sup> utilized XGBoost for short-term precipitation forecasting, demonstrating its capability to enhance forecast accuracy by correcting biases in numerical weather predictions. In the context of oil demand forecasting, XGBoost's capability to capture complex interactions between variables makes it a promising tool. Research by Dezhkam and Manzuri (2023)<sup>12</sup> demonstrated the efficacy of combining XGBoost with the Hilbert-Huang Transform for stock market forecasting, a domain with analogous complexity to oil markets.



Similarly, the N-HiTS model<sup>13</sup>, a neural network-based approach, has shown remarkable effectiveness in time series forecasting. This model leverages hierarchical interpolation and has been recognized for its ability to handle non-linear time series data efficiently. Studies like those by Souza et al. (2023)<sup>14</sup> have employed N-HiTS to predict COVID-19 cases and deaths, showcasing its potential in handling non-linear time series data. The N-HiTS model's architecture allows it to adapt to different scales of data, making it suitable for various forecasting horizons. Although relatively new, N-HiTS promises high accuracy, particularly for long-term forecasts, which are crucial for strategic energy planning. Its ability to dynamically adjust and learn from new data points ensures that the forecasts remain relevant and accurate over time.

This paper employs a top-down approach, using economic and other industry-specific indicators as inputs to forecast global demand across different refined oil products in seven global regions. These regions are strategically chosen to represent the major oil-consuming and producing areas, ensuring a comprehensive analysis. The forecasts are then aggregated to form a comprehensive global view on oil demand. The performance of the ML models, specifically XGBoost and N-HiTS, is assessed using metrics such as Mean Absolute Percentage Error (MAPE) and Mean Squared Error (MSE), targeting a high accuracy with a MAPE around 10% for forecasts extending one year out-of-sample. These metrics are critical for evaluating the models' predictive performance and ensuring that the forecasts are both reliable and actionable.

Our methodology demonstrates its application in three ways: as a short-term forecasting tool (using monthly inputs and forecasting up to 24 months ahead), as a medium-term forecasting tool (using quarterly inputs and forecasting up to 5 years ahead), and as a forecast scenarios analysis tool. These applications highlight the versatility of our approach and its potential to inform strategic decisions across different planning horizons.

By leveraging the successes of ML in various domains, this study seeks to set a new benchmark in oil demand forecasting. The ultimate goal is to enhance the strategic planning capabilities of the energy industry, offering more accurate and timely forecasts to inform strategic decisions. The research highlights the advantages of ML in identifying complex patterns and adapting to new data, making it particularly suitable for the volatile energy sector.

The remainder of the paper is organized as follows. In Section 2 we review the data processing stage providing the data description and elaborating on the methodologies employed for data cleaning, data transformation and data splitting. Section 3 presents the methodology and framework in terms of the approach used for model selection and training. Section 4 demonstrates the application of our methodology in constructing short-term and medium-term forecasts of global oil demand and examines comparatively information against actuals and consensus forecasts. In Section 5, we demonstrate how these forecasts can be supplemented with various scenarios. These forecast scenarios, for example, can illustrate the sensitivity of the baseline forecasts to changes in assumptions about the global macroeconomy, the adoption of new technologies, and evolving patterns of oil consumption. Section 6 draws the conclusions and discusses areas for future research.



## 2. Data processing

The data processing stage is instrumental in transforming raw data into a format that is amenable to effective modelling. This process encompasses several critical steps: data cleaning, data transformation, and data splitting. Each step is meticulously designed to address specific challenges inherent in time series forecasting, such as handling missing values, ensuring data stationarity, and optimizing the dataset for training and validation purposes. Moreover, the same approach is followed whether it is applied for short-term or mid-term forecast with some minor differences that are highlighted. This section elaborates on the methodologies employed in preparing the dataset, emphasizing the technical strategies and their theoretical underpinnings to ensure the highest data quality for subsequent modelling.

Table 1. Global oil demand forecasting model data structure				
Geography				
<b>Regions / Countries</b>	Description			
OECD				
US	United States			
Other Americas	Includes other OECD Americas, namely, Canada, Chile and Mexico.			
Europe	Includes OECD Europe, namely, Austria, Czech Republic, Denmark, Estonia, France, Germany, Greece, Hungary, Italy, the Netherlands, Norway, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, Turkey and the UK.			
APAC	Includes OECD Asia-Oceania, namely, Australia, Israel, Japan, Korea and New Zealand.			
Non-OECD				
China	China.			
India	India.			
Other non-OECD	Includes rest of world.			
Oil products				
Oil product	Description			
LPG (d1)	Includes all liquefied petroleum gases.			
Naphtha (d2)	Includes naphtha as feedstock to the petrochemical industry and for gasoline production. Excludes naphtha type jet fuel.			
Gasoline (d3)	Includes finished motor gasoline and motor gasoline blending components and additives.			
Jet/Kero (d4)	Includes kerosene-type jet fuel and other kerosene.			
Gasoil/Diesel (d5)	Includes diesel oil, light heating oil and other gas oils.			
Fuel oil (d6)	Includes all residual fuel oils.			
Other products (d7)	Includes other oil products, such as crude oil, other NGL, synthetic crude/fuels, orimulsion, hydrogen, refinery gas, aviation gasoline, naphtha-type jet fuel, white spirit, SBP, lubricants, bitumen, paraffin waxes, petroleum coke, tar, sulphur, aromatics and olefin.			

Source: IEA MODS, OIES



## 2.1. Data description

Our global oil demand forecasting models disaggregate global oil demand into OECD and non-OECD regions, as well as seven primary oil products, namely: LPG (d1), naphtha (d2), gasoline (d3), jet/kero (d4), gasoil/diesel (d5), fuel oil (d6), and other products (d7). Demand for oil products in OECD is further disaggregated into the US, Other Americas, Europe and Asia-Oceania (APAC). Demand for oil products in the non-OECD is further disaggregated into China, India and other non-OECD. OECD data are based on the International Energy Agency's *Monthly Oil Data Service* (IEA MODS) database. For non-OECD data we utilize various sources such as IEA, Argus, China's National Bureau of Statistics (NBS), India's Petroleum Planning and Analysis Cell (PPAC), and other industry sources. The sample period spams from January 1990 to June 2023 for OECD, and from January 2000 to June 2023 for non-OECD countries/regions. All oil data are monthly. **Table 1** summarizes the data structure for the dependent variables.

The forecasting models also utilize four groups of independent determinants covering global prices, global economics, global industry and sector-specific indicators. These are summarized in **Table 2**. The main source of data is Oxford Economics' *Global Databanks*. The significance of these predictors in forecasting global oil demand is analyzed and evaluated in the following sections.

Table 2. Description of selected global oil demand predictors					
Global prices					
rpo	Oil price	Brent price, period average.			
wci	World commodity index (non-fuel)	Average of world food price, world beverages price, world agriculture raw materials price and world metals price.			
Global ec	onomics				
рор	Population	Total population.			
gdp	GDP	GDP, in USD\$ terms.			
срі	CPI	Consumer Price Index: All items.			
inc	Disposable income	Personal disposable income.			
wtr	World trade index	Oxford Economics' global trade index.			
lkq	Li Keqiang index	Index measuring China's economy based only on rail freight volume, electricity production, and bank loans.			
Global in	dustry				
ind	Industrial production	Manufacturing production index.			
ррі	PPI	Producer price index.			
inv	Total investment	Total fixed investment.			
out	Gross output	Total value added.			
pcars	Stock personal cars	Light vehicles for personal use.			
ccars	Stock commercial cars	Light and heavy vehicles for commercial use.			
cbui	Construction of buildings	Industrial production in construction of buildings.			
chem	Sales gross output (chem)	Output chemicals minus Pharmaceuticals.			
egen	Electricity gen by oil	Electricity production from oil sources (% of total).			
Air passenger forecasts					
airf	Air fares index	Average air fares index. Sample starts in 1Q06.			
airp	Air passengers	Total passengers. Sample starts in 1Q06.			
rpk	Revenue passenger km	Revenue passenger kilometers in millions. Sample starts in 1Q06.			
Services Outerd Economics' Clabel Detabanka, OLES					

Source: Oxford Economics' Global Databanks, OIES



## 2.2. Data cleaning

Our short-term demand data is monthly while economic indicators are quarterly. However, for the midterm both the demand and economic indicator data is quarterly and do not require any frequency conversion. For the short-term forecast, any quarterly data must be converted to monthly data to allow for a monthly forecast instead of a quarterly forecast for the short-term. The following sections highlight the required steps prior to model training.

**Converting Quarterly to Monthly Data:** Given the discrepancy in the temporal resolution of demand data (monthly) versus other indicators (quarterly), it was imperative to homogenize the dataset to a monthly scale to maximize observational data points for analysis. To achieve this, the Pandas<sup>15</sup> library's linear interpolation method was employed due to its simplicity and effectiveness in estimating missing values. This approach, particularly suitable for time series data, linearly interpolates missing or NaN values based on the linearly spaced values between known data points. By utilizing the .interpolate(method='linear') function on our dataset, we ensure a smooth transition between quarterly data points, thereby maintaining the trend and variability observed in the original data. This method assumes that the change between two data points is linear, filling in missing values with appropriately spaced estimates that reflect the underlying data pattern.

The interpolation formula used is given by:

$$V_{interpolated} = V_{start} + \frac{(V_{end} - V_{start})}{(T_{end} - T_{start})} \times (T_{interpolated} - T_{start}),$$

where:

Vinterpolated is the value to be interpolated (estimated) for the target time point.

V<sub>start</sub> and V<sub>end</sub> are the known values before and after the point to be interpolated, respectively.

 $T_{start}$  and  $T_{end}$  are the times at which  $V_{start}$  and  $V_{end}$  are observed, respectively.

*T<sub>interpolated</sub>* is the time point at which the value is to be interpolated.

#### Figure 1: Example of backcasting missing values



Source: Authors' analysis

**Handling Missing Values using Prophet:** To address instances of missing data in our dataset, we integrated the Prophet<sup>16</sup> package for backcasting purposes. Prophet, renowned for its efficacy in discerning underlying trends and seasonal fluctuations within time series data, facilitated the extrapolation of data points backwards. This method allowed us to leverage established patterns to infer and populate missing values, thereby ensuring the dataset's completeness for subsequent analysis. The Prophet package, an open-source tool devised by Facebook, excels in automatic



forecasting of univariate time series data, simplifying the process of selecting optimal hyperparameters to enhance forecast accuracy. The reason why we decided to not drop the rows with missing values is the limited dataset we had. Furthermore, our results show a boost in accuracy when keeping the rows with missing values and imputing as opposed to dropping the rows with missing values. **Figure 1** shows an example of Prophet applied to back two variables with missing values. It is important to note that this step applies exactly to both the short-term and mid-term where the only difference is the frequency of the time-series being processed.

## 2.3. Data transformation

Achieving Stationarity: Differencing has been applied to stabilize the mean and variance of the time series to ensure stationarity. While differencing is a common approach to achieve stationarity, its minimal impact on our results suggests that the ML models selected were robust to non-stationary data, a characteristic advantageous for our forecasting objectives as shown in **Figure 2**.



Figure 2: Effect of differencing using XGBoost

Notes: US d1. Source: Authors' analysis

**Scaling:** Preliminary analysis indicated that scaling the data to standardize feature ranges did not significantly affect the performance of our ML models. Our tests involved using a min-max scaler in XGBoost on our training set and comparing the effect with and without scaling. For N-HiTS, we use a robust scaler<sup>17</sup> which is a method used to standardize features by removing the median and scaling with the mean absolute deviation (MAD) or inter quartile range (IQR). This technique is particularly useful with noisy data where outliers can heavily influence the sample mean and variance in a negative way.



## 2.4. Data splitting

The following methodology applies to both the input data for the short-term and medium-term models with any specific modification except for the frequency parameter (monthly versus quarterly) that needs to be specified in some of the software packages that we used.

**Training, Validation, and Test Sets:** The dataset was divided into training, validation, and testing sets following time series forecasting best practices. This split ensures that the models are trained on historical data, validated to tune hyperparameters with unseen data, and finally, tested to evaluate performance on the most recent, unseen data. Specifically, data up until 2018 was used for training and validation; and 2019 data was used for out of sample testing. The rational for cutting our sample in 2019 was to avoid the impact of the 2020 COVID shock in evaluating our models' performance, as both its significant exogeneity and magnitude weighing on global oil demand driven by the governments' restrictions on mobility, as well as pend-up demand in response to lifting these restrictions in 2021 and 2022, were identified as sources of distortion both for model training and evaluation.

In the next section, we'll delve into Model Development and Evaluation, detailing the methodologies behind model selection, training, evaluation metrics, and the post-processing techniques employed to refine the forecasts.

#### 2.5. Evaluation metrics

To assess the performance of our forecasting models, we employ two primary evaluation metrics: Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE). These metrics are chosen for their ability to measure forecast accuracy in complementary ways, providing a holistic view of model performance.

**Mean Squared Error (MSE):** The MSE is calculated as the average of the squared differences between the actual and predicted values. It is given by the formula:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$
,

where  $y_i$  is the actual value,  $\hat{y}_i$  is the predicted value, and *n* is the number of observations. MSE is sensitive to large errors, making it useful for identifying models that might be prone to significant forecast deviations.

**Mean Absolute Percentage Error (MAPE):** The MAPE measures the average magnitude of errors as a percentage of actual values, offering an intuitive understanding of model accuracy. It is defined as:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| ,$$

MAPE is particularly valuable in contexts where the relative size of the forecast error is more important than the absolute size, providing insights into the model's performance in percentage terms. This metric is generally easier to understand especially when we want to compare the performance across the different products that might have varying error margins.

These metrics together allow for a comprehensive evaluation of forecast accuracy, with MSE highlighting models that may have large errors for debugging purposes; and MAPE offering a percentage-based measure that is easily interpretable and used for reporting and comparing models.



## 3. Model selection and training

The following sections explain the training and model selection approach that was used. This applies to both the short-term and medium-term forecasting models.

## 3.1. XGBoost for time series forecasting

XGBoost (Extreme Gradient Boosting) operates on the principle of boosting, an ensemble technique that combines multiple weak learners (typically decision trees) into a strong learner in a sequential manner. Each tree attempts to correct the errors made by the previous one, with this process guided by the gradient of the loss function.

Decision trees, usually create a model that predict the target by creating trees using the if/else statement, and by using the minimum number of such statements they try to find the probability of having the correct decision. Such trees are used for classification or regression (as in our problem).

The model's capacity to handle various data irregularities, such as non-linearity and missing values, stems from its robust loss function optimization and regularization features. XGBoost introduces a regularization term in the objective function, reducing overfitting by penalizing complex models. This feature, combined with its ability to perform automatic handling of missing values, makes XGBoost particularly well-suited for time series forecasting where such irregularities are common.

Gradient boosting decision trees (GBDT) such as XGBoost are well known for their performance against deep learning models in tabular problems. There are several GBDT packages, however, in our case we have decided to XGBoost as the most used by both, ML researchers and academia. **Figure 3** illustrates the way XGBoost (and other GBDT models) finalize their predictions where the final prediction for a given sample is the sum of predictions from each tree.





Source: Recreated from AWS Sagemaker<sup>18</sup>

The figure depicts the iterative process of training a gradient boosting model. Initially, the data set (X, Y) is used to train the first tree,  $F_1(X)$ . The residuals  $r_1$  are computed to measure the discrepancy between the predicted and actual values. A regularization parameter  $\alpha_1$  is then determined to minimize the loss when  $F_1(X)$  and  $r_1$  are combined.



This process is repeated for each subsequent tree. For the  $i_{th}$  tree, the model  $F_i(X)$  is updated to

 $F_{i-1}(X) + \alpha_i h_i(X, r_{i-1})$ ,

where  $\alpha_i$  and  $r_i$  are the regularization parameters and residuals computed with the  $i_{th}$  tree respectively, and  $h_i$  is a function that is trained to predict residuals  $r_i$  using *X* for the  $i_{th}$  tree.

The objective is to find the optimal set of  $\alpha$  parameters by minimizing the differentiable loss function L(Y, F(X)), which is computed as:

$$\arg\min_{\alpha} = \sum_{i=1}^{m} L(Y_i, F_{i-1}(X_i) + \alpha h_i(X_i, r_{i-1}))$$

Through this iterative optimization, the final model  $F_m(X)$  is a combination of each individual tree's predictions adjusted by their corresponding regularization parameters, aimed at reducing the loss function.

For the short-term forecast application (i.e., 2-years ahead), the *model development process* involved creating a single model for every time step in our prediction horizon for a given region (i.e., 24-steps ahead). This means that for the short term, we need to create 24 models for each product resulting in a total of 167 models to do a short-term forecast for a given region.

To retain the temporal relationship between past demand observations, the feature engineering process was crucial, involving the creation of 12 new features representing lags of the target variable from t-1 to t-12, capturing the historical demand trends and seasonality. Additionally, future lags from t+1 to t+24 were generated to predict demand across various future months, enhancing the model's forecasting capabilities. **Figure 4** illustrates how the input matrix looks like.

#### Figure 4: Input matrix setup for XGBoost



Source: Authors' analysis

As seen in **Figure 4**, for every product d(i) for  $i \in \{1,2,3,4,5,6,7\}$  an input matrix is created. Then, it is mapped to a target column  $d(i)_{t+j}$  where j is the index of the forecasted month. This means that in the short-term j ranges from 1 to 24 since we have 24 months in our forecasting horizon.

The hyperparameter tuning strategy employed grid search and cross-validation techniques to optimize parameters such as the learning rate and the number of trees. The learning rate controls the step size at each iteration of the model's optimization, while the number of trees determines the complexity of the model. The objective was to find a balance that minimized forecast error without overfitting to the training data. The mathematical representation of the gradient boosting process can be summarized as follows:

$$\hat{y}_i(t) = \hat{y}_i(t-1) + \eta \cdot ft(x_i),$$

where  $\hat{y}_i(t)$  is the prediction at iteration *t*, *ft* is the decision tree added at iteration *t*,  $x_i$  is the feature vector for instance *i*, and  $\eta$  is the learning rate.



To establish a baseline for XGBoost we used the US and China as the main regions to calculate the model performance on our 2019 out of sample data. MAPE is used to measure the performance by calculating the weighted average of MAPE across products by weighing them with the average demand in 2019. This weighted average is calculated according to the following formula:

 $Weighted Average = \frac{Average_Demand_i \cdot MAPE_i}{\sum_{i=1}^{7} Average_Demand_i}$ 

Where  $Average\_Demand_i$  is the demand for product (i) and  $MAPE_i$  is for the same product.

Table 3. XGBoost performance and out-of-sample data (2019)						
	US		China			
Product	XGBoost	Total oil demand	XGBoost	Total oil demand		
	%	kb/d	%	kb/d		
LPG	10.48	2,638	11.99	2,248		
Naphtha	8.08	207	10.22	1,298		
Gasoline	1.53	9,296	9.01	3,662		
Jet/Kero	3.78	1,760	17.35	975		
Gasoil/Diesel	2.63	4,112	5.13	3,949		
Fuel oil	19.58	306	14.07	604		
Other products	15.43	2,260	24.33	1,406		
Weighted Average	4.95		10.83			

Source: Authors' analysis

**Table 3** shows that the performance of XGBoost is quite promising for two of the largest demand centers. The variance between the US and China's performance might be attributed to weaker correlations between the economic indicators and product demand in the case of China.

## 3.2. N-HiTS architecture and training

N-HiTS (Neural Hierarchical Interpolation for Time Series Forecasting) introduces a novel approach by employing hierarchical interpolation and multi-rate data sampling. This architecture decomposes the time series into components of varying frequencies, allowing the model to capture complex temporal patterns effectively. In principle, N-HiTS perfectly reconstructs harmonic signals from the original combined data (**Figure 5**).



#### Figure 5: Illustration of N-HiTS architecture and signals harmonization



The hierarchical interpolation technique enables the model to integrate information across different time scales, from high-frequency details to low-frequency trends. Multi-rate data sampling further enhances this capability by allowing the model to learn from data points sampled at different intervals, adapting to the inherent temporal dynamics of the dataset.

The advantages of N-HiTS over traditional and other ML-based forecasting methods include its superior ability to capture long-term dependencies and its flexibility in handling time series with complex seasonal patterns. These features make N-HiTS particularly effective for long-horizon forecasting tasks. Unlike previous efforts, N-HiTS uses neural network models specifically designed for time series forecasting. N-HiTS, allows an interpretable non-linear decomposition which is quite important in high-stake applications as in our problem.

Using Nixtla's<sup>19</sup> implementation of AutoNHITS, we initialized the model with various types of exogenous variables to enrich the forecasting model with additional context. The inclusion of future exogenous variables was facilitated by specifying these under the **futr\_exog\_list** parameter during model initialization.

The training process involved scaling both the target and exogenous variables using a robust scaler to ensure consistency in the data range, crucial for the model's performance. The model was then fitted to the dataset, with the training duration and computational resources carefully managed to optimize performance without compromising efficiency.

Challenges encountered during the training process, such as model convergence and overfitting, were addressed through careful tuning of the model's architecture and parameters. Early stopping and regularization were employed as key strategies to ensure that the model generalizes well to unseen data.

By leveraging the hierarchical and multi-rate capabilities of N-HiTS, combined with a thoughtful approach to model training and exogenous variable inclusion, we were able to develop a forecasting model that offers both high accuracy and efficiency.

**Table 4** shows that N-HiTS outperforms XGBoost in most of the highest demand regions including the US, Europe, APAC and China. The main advantage of N-HiTS over XGBoost is model maintenance as N-HiTS builds a single model per region while XGBoost builds 167 models per region without an advantage in performance. Furthermore, N-HiTS can be easily adapted to run longer term forecasts. Based on this, our methodology utilizes N-HiTS as the main algorithm to forecast demand. Moreover, N-HiTS can also be expanded to medium-term forecasting<sup>1</sup> using the same features and architecture. Additionally, the same methodology can be employed for scenario analysis as discussed in the following sections.

Table 4. N-HiTS vs XGBoost average MAPE by region						
	N-HiTS	XGBoost				
Region	Weighted average MAPE	Weighted average MAPE				
	%	%				
US	3.51	4.95				
Other OECD Americas	10.13	6.30				
OECD Europe	3.22	4.33				
OECD APAC	4.20	5.57				
China	10.35	10.82				
India	9.8	8.37				
Other non-OECD	5.85	3.25				

Source: Authors' analysis

<sup>1</sup> Next 8-10 years on a quarterly basis



## 3.3. Post-processing predictions

In both our XGBoost and N-HiTS predictions, it is crucial to address potential extreme values that diverge significantly from historical patterns. This refinement process involves two main steps: establishing bounds based on historical percent changes and applying these bounds to the forecasts. This refinement process is applied to both the short-term and long-term forecasts.

#### **Establishing Bounds**

**1. Calculating Historical Percent Changes:** For each time series, corresponding to a product in a specific region, we calculate the month-to-month percent changes historically observed. The percent change between two consecutive months,  $P_{change}$ , is calculated as:

$$P_{change} = \left(\frac{V_{current} - V_{previous}}{V_{previous}}\right) \times 100\% \text{ ,}$$

where  $V_{current}$  is the value of the current month, and  $V_{previous}$  is the value of the preceding month.

**2. Creating a Distribution:** We construct a distribution of historical percent changes for each month. This involves aggregating all percent changes observed in the same month across different years to form a comprehensive view of historical variability.

**3. Determining Bounds**: Based on the distribution of percent changes, we establish upper and lower bounds for acceptable changes. These bounds are determined by selecting a confidence interval, say the 90th percentile, which represents a statistically significant threshold. For example, if the chance of observing more than a 29% change in January is less than 5%, we set our bounds at the 5th and 95th percentiles of the January percent change distribution.

#### **Applying Bounds to Forecasts**

**1. Calculating Forecast Percent Changes:** Similar to historical calculations, we determine the percent change between forecasted values for consecutive months.

**2. Applying Bounds:** If a forecasted percent change exceeds the established bounds, it is adjusted to the nearest bound. This ensures that our forecasts remain within a historically plausible range, enhancing their reliability and accuracy

$$\begin{cases} P_{lower} & if & P_{forecast} < P_{lower} \\ P_{forecast} & if & P_{lower} \leq P_{forecast} \leq P_{upper} \\ P_{upper} & if & P_{forecast} > P_{upper} \end{cases}$$

Where  $P_{forecast}$  is the percent change of the forecast, and  $P_{lower}$ ,  $P_{upper}$  are the determined lower and upper bounds, respectively.

This post-processing step ensures that extreme forecasts are refined to align with historical trends, maintaining the integrity of our predictions and bolstering confidence in their applicability for strategic decision-making. It also preservers the structure of the data and inter-month changes unlike a simple moving average which might over-smooth such changes. **Figure 6** shows the difference between the proposed methodology and a simple moving average.



Figure 6: Proposed bounding function vs moving average



Source: Authors' analysis

## 4. Estimation results

Our objective throughout this section and the next is twofold. First, to present the application of our ML model in forecasting global oil demand by region and by product in both the short-term using monthly inputs and the medium-term using quarterly inputs, highlighting the versatility of our approach and its potential as an informative forecasting tool across different horizons, as well as for forecasting scenario analysis. Second, to contrast our medium-term forecasting results with the consensus oil demand forecasts available by the International Energy Agency (IEA), the U.S. Energy Information Administration (US EIA) and the Organisation of the Petroleum Exporting Countries (OPEC) at the time of our analysis, as well as to compare the forecasting performance of the ML model against conventional econometric forecasting models. For expository purposes, in the analysis below we focus on the N-HiTS algorithm to derive our forecasts that has been shown to perform better than the XGBoost and provides a flexible methodological framework in terms of complexity, training time and scalability.

## 4.1. Short-term oil demand forecasts

For our short-term forecasting analysis, we compare the model predictions for global oil demand with the actual estimates of global oil demand in 2019 and 2022 and investigate the forecasting performance of the model against the real outcomes. The forecast horizon is 12 months, and the model is trained based on monthly data up to December of the previous year. We purposely refrain from including in the out-of-sample analysis the period 2020-2021 as this was the period most severely impacted by the COVID shock and by the pandemic impact on mobility and economic activity, albeit related exogenous elements affecting the model's forecasting performance persist well into 2022 as the global pandemic was officially declared over by the World Health Organization only on 5 May 2023.

**Figure 7** shows the model predictions against the actual outcome of global oil demand in 2019. The model forecasts annual global oil demand in 2019 averaging 100.77 mb/d from 100.1 mb/d in 2018 and versus the actual 100.86 mb/d, growing y/y by 670 kb/d versus the actual growth of 760 kb/d. That is a difference between actual and predicted growth for global oil demand of less than 100 kb/d or 0.1% of annualized global oil demand.





Figure 7: Actual vs predicted global oil demand in 2019

#### Source: OIES

**Figure 8** breaks down the comparison by region and by product. In terms of regions, the forecasting model appears to underestimate OECD demand growth in 2019 by 410 kb/d to -370 kb/d compared to the actual 40 kb/d, underperforming mainly in other Americas and Europe. On the other hand, it overestimates the growth projections for non-OECD oil demand by 320 kb/d to 1 mb/d versus the actual 720 kb/d overstating demand growth mainly in other non-OECD. In terms of products, the forecast performs relatively better particularly across the main transportation fuels overstating growth in naphtha demand and fuel oil at the expense of predicted demand growth for the group category of other petroleum products. The main disagreement in the OECD forecasts versus the actuals can be observed on LPG and diesel/gasoil predicted demand growth, both of which are underestimated by 200 kb/d y/y, but the difference was mainly offset by overstating the predicted diesel/gasoil demand growth in non-OECD by 340 kb/d. By comparison, IEA's *Oil Market Report* in January 2019 was predicting global oil demand growth of 1.4 mb/d for the year, while OPEC's *Monthly Oil Market Report* had its oil demand growth forecast for 2019 at 1.3 mb/d versus our predicted growth of 670 kb/d and the actual 760 kb/d.









Figure 8 (cont.): Actual vs predicted global oil demand growth in 2019

#### Source: OIES

Moving to our out-of-sample forecast predictions for 2022, this year was very challenging with the oil market being subject to a series of large shocks such as the Russian invasion of Ukraine and the ensuing sanctions, embargoes and price caps on Russian oil imports that followed as a response, the coordinated response by oil-consuming nations led by the US to control prices by a massive release of strategic stocks, the recessionary and inflationary pressures weighing on the global economy and China's demand shocks from its strict zero-COVID policy as the rest of the world was still recovering from the COVID shock. **Figure 9** plots the predicted global oil demand versus the actual demand data in 2022. The forecasting model predicts annual global oil demand growth of 2.6 mb/d with global oil demand averaging 100.06 mb/d for the year, from 97.44 mb/d in 2021, 290 kb/d above actual global oil demand growth of 2.3 mb/d that averaged 99.77 mb/d.





Source: OIES



As **Figure 10** shows, again the forecasting model appears to understate the predicted OECD demand growth in 2022 by 1.2 mb/d with the Americas accounting for 90% of the forecast error, while it overpredicts oil demand growth in the non-OECD by 1.5 mb/d with China alone accounting for 88% of the error. In terms of products, the weakest forecast can be seen in jet/kero demand with growth for the year underestimated by 610 kb/d followed by LPG demand growth that is seen 355 kb/d below the actuals. These were offset by the forecast overpredicting demand for naphtha, gasoline and diesel/gasoil by a combined 1.2 mb/d. By comparison, IEA's *Oil Market Report* in January 2022 was predicting global oil demand growth of 3.3 mb/d for the year and OPEC's *Monthly Oil Market Report* 4.15 mb/d versus our predicted growth of 2.6 mb/d and the actual 2.3 mb/d. In both the 2019 and 2022 out-of-sample forecast horizons and despite some limitations, our forecasting models outperform the consensus at the time suggestive of satisfactory predictive performance.



Figure 10: Actual vs predicted global oil demand growth in 2022

By product



Source: OIES



Another important aspect to examine is the predictive performance of the ML model against conventional econometric forecasting models. For this exercise we utilize a 12-step ahead vector autoregressive with exogenous variables (VAR-X) forecasting model of global oil demand by region and oil product. This purpose-built forecasting model first developed in 2021 under our OIES Oil Monthly series to forecast the two-year ahead impact of the COVID pandemic on global oil demand, employs global prices, macroeconomic and other industry variables along with exogenous COVID-specific and other indices (e.g., mobility indices, flight schedules, etc.) to predict out-of-sample global oil demand for 2022 and compare the results with the ML forecast. Figure 11 shows the performance analysis. The VAR-X model predicts global oil demand growth of 3.23 mb/d for 2022, versus the predicted growth of 2.62 mb/d by the ML model and the actual 2.33 mb/d. However, considering the limitations of the ML predictions stated above and the fact that the VAR-X forecasting model was specifically designed to capture the dynamic behaviour of the relationship between endogenous and exogenous variables related to COVID that dictated performance at the time, the forecast accuracy of the VAR-X model outperforms that of the ML model under all forecast accuracy metrics used for the performance analysis. The VAR-X achieved a lower Mean Absolute Percentage Error (MAPE) of 1.2% versus 1.6% of the ML forecast, as well as Root Mean Square Error (RMSE) of 1.5 mb/d versus 1.8 mb/d, respectively, in absolute levels. That said, differences between the two forecasts remain comparatively small and the ML forecasting framework remains superior due to its flexibility, modularity and adaptability, and overall model management compared to conventional econometric forecasting models such as the VAR-X.



Figure 11: Actual vs predicted global oil demand in 2022 using ML and VAR-X models

#### Source: OIES

To further examine the role of externalities particularly in 2022 in the ML forecasting framework we conduct a sensitivity analysis investigating whether taking into account some of these externalities can improve our forecast's accuracy. We incorporate in the US model two additional variables and run the forecast results for 2022 against our baseline and the actuals. The first variable is a composite measure of nine policy response metrics to the COVID pandemic namely *stringency index* available by the Blatvantik School of Government of the University of Oxford (**Figure 12**). The second is the stock of excess savings in the US available by the U.S. Federal Reserve in order to capture the sensitivity of the US oil demand forecast to consumers' savings during the pandemic, defined as savings accumulated above the trend of household savings rate (**Figure 13**).







Figure 13: Stock of excess savings in the US

Source: US Federal Reserve

**Figure 14** shows the forecast scenario for US oil demand in 2022 against our baseline forecast and the actuals. Clearly, the inclusion of the two variables aiming to capture the externalities in the year associated with COVID and consumer behaviour greatly improve the US forecast with growth under the forecast scenario corrected to 290 kb/d compared to the actual 410 kb/d and the growth prediction under the baseline of -120 kb/d. This goes to show that oil demand forecasts are inherently susceptible, not least, to external economic and policy factors, shifts in consumer behaviour and disruptive events, but also on a more technical level to data complexity and availability. In this case, including only these two exogenous variables more than halved the Mean Absolute Percentage Error (MAPE) of our US oil demand forecast in 2022, down from 3.9% to 1.9%.



#### Figure 14: Forecast scenario analysis of US oil demand in 2022

Source: Blavatnik School of Government, Uni. of Oxford So

Source: OIES



## 4.2. Medium-term oil demand forecasts

The forecast horizon for our medium-term forecasting analysis is 30 quarters, spanning from 3Q 2023 to 4Q 2028. The model is trained based on quarterly data between 1Q 2000 and 2Q 2023 as described in Section 2. We forecast global oil demand by product (namely LPG, naphtha, gasoline, jet/kero, diesel/gasoil, fuel oil and other products) across US, other OECD Americas, OECD Europe and OECD Asia-Pacific (APAC) to derive total OECD oil demand, as well as China, India and other non-OECD for total non-OECD oil demand. The forecasts are then aggregated to represent our baseline global oil demand outlook. The forecasting models also utilize four groups of independent determinants covering global prices, global economics, global industry and sector-specific indicators (see Table 2, p. 5), and they are driven by modelled projections of these predictors to 4Q 2028 obtained by Oxford Economics' *Global Databanks* serving as our baseline assumptions.



Figure 15: Medium-term global oil demand outlook

Global cumulative oil demand growth, 2022-2028



Notes: Consensus averages IEA Oil 2024, OPEC World Oil Outlook 2024 forecasts. Source: OIES

**Figure 15** shows our baseline global oil demand forecast to 2028 on annual terms. Global oil demand growth is forecast to progressively slow over the 2024-2028 forecast period, but to remain within the 1-1.5 mb/d range. Growth decelerates from 2.3 mb/d in 2023 to 1.7 mb/d in 2024 and 1.3 mb/d in 2025,



before settling to 1.1 mb/d in 2026/27 and fall to 900 kb/d in 2028. This results in a net increase of 8.5 mb/d during the 2022-2028 period, compared to 8.4 mb/d based on consensus forecasts. According to our baseline forecast, global demand growth is dominated by the non-OECD, especially China and India, with OECD demand growth shifting to small contractions from 2025-onwards. In absolute terms, after surpassing its 2019 pre-COVID level of 100.7 mb/d in 2023 by 960 kb/d averaging 101.6 mb/d, global oil demand is set to grow to 107.8 mb/d by 2028.







Global cumulative oil demand growth by region, 2022-2028

Notes: Consensus averages IEA Oil 2024, OPEC World Oil Outlook 2024 forecasts. Source: OIES

As **Figure 16** shows, non-OECD dominate the global oil demand growth outlook to 2028, with China leading gains but annual growth settling around 300 kb/d after 2024, a notable slowdown from 1.7 mb/d in 2023. This sees India that maintains a steady pace of growth around 220 kb/d closing the gap with China by 2027 and becoming the main source of global demand growth in 2028 rising y/y by 300 kb/d. Most of incremental oil demand however is projected to originate from other non-OECD accounting for 2.6 mb/d between 2024 and 2028, more than half of total non-OECD demand increment of 4.9 mb/d in the same period. By contrast, OECD oil demand is forecast to post a 400 kb/d net decline during the 2024-2028 period, led by Europe and APAC, with growth in OECD Americas narrowly avoiding a



contraction up until 2027 and US demand failing to grow in 2027 before reversing to a 70 kb/d y/y contraction in 2028.

In terms of products, **Figure 17** shows that petrochemicals demand (LPG/ethane and naphtha) is projected to drive global growth accounting for 3.3 mb/d of incremental demand between 2022 and 2028, largely in line with the consensus, followed by jet/kero fuel demand at 2.6 mb/d but growth pace is expected to progressively slowdown. This is also the case with our global gasoline outlook with demand growth expected to slow progressively and remain only marginally positive by 2028, but we fail to see global gasoline demand growth contracting or even plateauing before 2029. Diesel/gasoil demand growth is expected to gain pace from its 2023/2024 lows in 2025 and 2026 as the global economy picks-up before settling around 200 kb/d y/y in the remainder of the forecast horizon. Lastly, fuel oil demand remains supported to 2026, largely due to marine demand growth, before shifting to marginal y/y contractions in 2027/2028.



Figure 17: Medium-term global oil demand outlook by product

Global cumulative oil demand growth by product, 2022-2028



Notes: Consensus averages IEA Oil 2024, OPEC World Oil Outlook 2024 forecasts. Source: OIES



The most important difference between the consensus oil demand forecasts, not least, for the mediumterm is associated with the gasoline demand outlook. Between 2022 and 2030, IEA expects gasoline demand to contract by 900 kb/d with global gasoline demand essentially plateauing in 2024/2025 at 27.2 mb/d before gradually declining to 25.4 mb/d in 2030. Conversely, OPEC forecasts gasoline demand growing by 2.9 mb/d by 2030 reaching 29.2 mb/d. Our baseline projections for gasoline demand fall on the upper side of the consensus range, with global gasoline demand expected to grow by 1.7 mb/d between 2022-2028 to 27.9 mb/d (Figure 18). The main disagreement between IEA and OPEC on gasoline demand is associated with their assumptions about the pace of EV penetration to global fleet with IEA assuming a higher pace of penetration reaching 16% by 2030, while OPEC estimates the share of EVs to global fleet not to exceed 10% by the end of the decade, from close to 2% in 2022 in both cases. Efficiency improvements and mobility transformations (e.g., teleworking) are also factored in their respective forecasts. For our baseline outlook, we do not make any explicit assumptions about the composition of global fleet, efficiency gains or mobility transformations but rather we allow the algorithm to capture the evolution of the relationship between global fleet and motor fuels demand based on its ability to handle non-linear relations across different time scales as these emerge in the training sample.



Notes: Consensus averages IEA Oil 2024, OPEC World Oil Outlook 2024 forecasts. Source: OIES

Another important debate following the COVID pandemic is the pace of the recovery of jet fuel demand as it remains the only major fuel that as of 2023 failed to recover to its 2019 pre-pandemic levels. As **Figure 19** shows, our baseline forecast expects global jet/kero fuel demand on annual terms to exceed its 2019 level of 7.9 mb/d by 2025 in line with the OPEC forecast, while IEA projects a return to pre-pandemic levels no sooner than 2027. As before, IEA attributes this slow recovery on aircraft fuel efficiencies, operational improvements and optimisation in flight planning, while noting the decoupling between jet fuel demand and air travel (measured in revenue-passenger kilometers) that is expected to make a full recovery this year.

Summing up, our medium-term oil demand outlook to 2028 is slightly above the consensus, with oil demand growing by 8.5 mb/d between 2022 and 2028, versus the consensus estimate of 7.5 mb/d in the same period and reaching 107.8 mb/d in 2028, 640 kb/d above the consensus (see **Figure 20**). This difference can be attributed to the relaxed assumptions about new technologies and efficiency gains that can greatly impact our modelled projections as, for example, in the case of IEA. The low/high range based on all global oil demand forecasts ranges between 104.9 mb/d and 111 mb/d in 2028, with the divergence reaching 6.1 mb/d. In terms of global oil demand growth, the baseline forecast remains



close to or slightly above the consensus between 2024 and 2026 but begins to diverge higher from 2027-onwards for the same reasons. This is suggestive that the large divergence between consensus oil demand forecasts in recent years not least for the medium-term, can be attributed to divergent views about the speed of the energy transition and governments' decarbonization policies adding an important layer of uncertainty with important implications on oil investment decisions for the years to come.







Global oil demand growth

Notes: Consensus averages IEA Oil 2024, US EIA International Energy Outlook 2023, OPEC World Oil Outlook 2024 forecasts. Source: OIES



## 5. Forecast scenarios

In this section, we leverage additional exogenous variables to explore the feasibility of expanding our forecasting methodology for comprehensive forecasting scenarios analysis. These variables are specifically curated to assess the impact of various economic, technological and sector-specific changes on global oil demand by creating a purpose-built dataset for each scenario. Each dataset, representing a unique scenario, introduces new variables to examine different potential futures against our baseline. This nuanced approach enables us to test the adaptability and effectiveness of our methodology in implementing risk analysis using forecast scenarios and capturing the variances across diverse scenarios, reflecting the uncertainties inherent to our baseline outlook. For expository purposes, the scenario analysis presented in this section is maintained at a global level for the period 1Q 2024 to 4Q 2028, but our methodology is easily transferable at a regional and country-specific level, while it can utilize an inexhaustive list of scenarios conditional to data availability.

## 5.1. Implementation

Our forecasting scenarios analysis spans across two exemplary categories, namely: Macro scenarios and Sector-specific scenarios: Aviation. However, the initial pass through these scenarios revealed an unexpected challenge: the lack of significant differences in outcomes across scenarios. This uniformity was traced back to high correlations among several variables, which essentially echoed the baseline scenario information, masking the unique impacts of the scenarios.

For instance, consider analyzing two scenarios regarding a variable of interest—demand (D). The input variables under consideration include three input factors: X1, X2, and X3. The challenge arises when X1 and X2 vary across scenarios, but X3 remains constant, and X3 has a strong correlation with demand (D). As a result, even though X1 and X2 change, the constant nature of X3—combined with its high correlation with demand—dampens the observable variation in D across different scenarios during inference.

By applying a feature selection process, we decide to focus solely on X1 and X2, excluding X3 from the model due to its constant value across scenarios and its high correlation with demand. This adjustment allows the model to more clearly distinguish between the scenarios, directly attributing differences in predicted demand (D) to the variation in X1 and X2, without the confounding influence of X3. This simplified approach highlights the importance of feature selection in enhancing model clarity and effectiveness in scenario analysis.

To ensure the model focuses on the most impactful variables, we turned to a feature selection strategy that uses correlation analysis. By calculating Pearson correlation coefficients (r), as shown in the formula below, we pinpointed variables that moved in tandem above a certain threshold (0.9), indicating a strong linear relationship.

$$r = \frac{\sum_{i=1}^{n} (X_i - \overline{X}) (Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2} \sqrt{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}}$$

where:

n is the number of samples,

 $X_i$  and  $Y_i$  are the individual sample points,

 $\overline{X}$  and  $\overline{Y}$  are the means of the X and Y samples respectively.

The correlation coefficient ranges from -1 to +1, with +1 meaning a perfect positive linear correlation, -1 meaning a perfect negative linear correlation, and 0 indicating no linear correlation.



While this allowed us to identify variables for potential omission, we ensured the retention of certain key variables critical to each scenario's narrative:

**Macro scenarios:** In scenarios assessing the macroeconomic landscape ("baseline", "macro\_high", "macro\_low"), core variables including 'wci' (world commodity index), 'ind' (industrial output), and 'gdp' were preserved to anchor the forecasts in solid economic foundations.

Aviation scenarios: For scenarios like "air\_high" and "air\_low," variables such as "airp" and "rpk" were deemed essential for their direct relevance to air travel demand, an area expected to fluctuate significantly between high and low air travel scenarios.

**Figure 21** below shows an example of how many of the non-scenario variables in the air scenario are positively correlated with the air scenario variables of interest such as wtr, gdp and ind. Meaning that changing the air scenario variables without changing these variables might not show any differences in demand or it may affect unrelated products.

This methodology, designed to reduce multicollinearity, not only enhanced model interpretability but also its responsiveness to the unique contours of each scenario. Particularly in comparing "air\_high" to "air\_low", we noticed a discernible divergence in jet fuel demand forecasts, indicating the model's heightened sensitivity to changes in scenario-specific variables.



#### Figure 21: Correlation matrix heatmap for the Aviation scenario variables

Source: Authors' analysis, OIES

## 5.2. Macro scenarios

The macroeconomic scenarios reflect low/high risks to the medium-term global economic growth prospects as of 4Q 2023. The downside scenario considers a case of higher for longer interest rates weighing on financial and housing markets, resulting in tighter credit conditions and several years of sub-par growth (*low case*). The upside scenario considers a case of global economic resilience driven by a near-term improvement in the inflation outlook and a consumer-led recovery strengthening the global economy (*high case*). **Figure 22** illustrates the underlying GDP growth assumptions of these two scenarios, along with our baseline global growth assumptions. Clearly, medium-term risks to global growth prospects are tilted to the downside with world GDP well below the baseline in the near-term on increased inflation uncertainty before gradually recovering from 2026-onwards.





Figure 22: Global growth scenarios

Notes: Global GDP in US\$ weighted terms. Source: Oxford Economics Global Scenarios Service

**Figure 23** shows the macro scenarios results. Downside risks to global growth are projected to trim nearly 1 mb/d of the cumulative global oil demand growth between 2024-2028 under our baseline case, with the oil demand outlook hit harder in the near-term (i.e., 2024-2025). Global oil demand is projected to grow at 106.7 mb/d by 2028 compared to our baseline 107.8 mb/d, from 101.6 mb/d in 2023. That said, global oil demand growth between 2024-2028 remains 160 kb/d higher than the consensus. Upside risks to the global economy provide further support to the baseline outlook, lifting the cumulative global oil demand growth between 2024-2028 by 320 kb/d with global oil demand reaching 108.1 mb/d by 2028 versus the base 107.8 mb/d.



#### Figure 23: Macro scenarios

Notes: Consensus averages IEA Oil 2024, OPEC World Oil Outlook 2024 forecasts. Source: OIES



## 5.3. Aviation sector scenarios

For our second set of exemplary scenarios, we illustrate a sector-specific scenario pertaining to aviation and jet fuel demand. The COVID pandemic exerted its heaviest toll on jet fuel demand than any other product, declining y/y by 40% in 2020 to 4.8 mb/d from 7.9 mb/d in 2019. On average, 2023 jet fuel demand was 91% of 2019 levels at 7.2 mb/d, remaining the only major oil product that did not recover to its pre-pandemic levels. At the same time, measures of global air travel activity such as the number of air passengers recovered to their 2019 levels in 2023, while others such as the revenue-passenger kilometers (RPK) have fully recovered by H1 2024 (**Figure 24**). Our scenarios assess whether jet fuel demand will maintain its post-COVID momentum to recover and exceed its 2019 levels aligning with the rebound in air travel activity or if fuel and operational efficiencies mean that the global jet fuel rebound has run its course settling to a slower growth pace in the remainder of the forecast horizon.



#### Figure 24: Air travel assumptions

Source: Oxford Economics Air Passenger Forecasts

**Figure 25** shows the scenario results suggesting that risks to the medium-term jet fuel demand outlook are skewed to the downside, while our base case that projects global jet fuel demand rebounding and exceeding 2019 levels by 2025 is close to the upper range of the balance. Following a remarkable growth of 1 mb/d in 2023, jet fuel demand under our high case is expected to average 330 kb/d between 2024-2028 versus the baseline 300 kb/d. The low case however sees jet fuel demand growth slowing to 200 kb/d on average at the same period, with global jet fuel demand rebounding above its 2019 level only in 2027. In terms of total demand, the low jet fuel demand case drags lower global demand growth between 2024 and 2028 by 400 kb/d to 107.4 mb/d versus our base 107.8 mb/d. Under the high case global demand gains another 100 kb/d reaching 107.9 mb/d in 2028.





#### Figure 25: Aviation scenarios

Notes: Consensus averages IEA Oil 2024, OPEC World Oil Outlook 2024 forecasts. Source: OIES

## 6. Conclusion

This paper illustrated the application of machine learning (ML) techniques, specifically Extreme Gradient Boosting (XGBoost) and Neural Hierarchical Interpolation for Time Series Forecasting (N-HiTS), to enhance the precision and reliability of global oil demand forecasts. Our analysis included seven refined oil products and seven major regions, providing a comprehensive perspective on global demand trends. The comparative evaluation demonstrated that the N-HiTS model surpasses XGBoost.

Several key factors influenced our preference for N-HiTS over XGBoost. Firstly, N-HiTS required significantly fewer models to make a prediction, with a single model per region compared to XGBoost's 167 models per region for short-term forecasts. This reduction in model complexity simplifies maintenance and updates. Secondly, N-HiTS consistently achieved lower Mean Squared Error (MSE) on average across regions, highlighting its superior performance in capturing and predicting the intricate patterns of oil demand.

Our findings highlight the considerable potential of ML models in the energy sector, particularly in enhancing strategic planning and decision-making processes. The integration of advanced ML techniques enables the capture of complex, non-linear relationships and provides a transparent, repeatable forecasting process. By achieving a Mean Absolute Percentage Error (MAPE) of less than 10% on average for monthly short-term forecasts extending one year out-of-sample, our methodology balances complexity (top down vs. bottom up) and accuracy in oil demand forecasting.

Furthermore, techniques such as linear interpolation for converting quarterly to monthly data and the use of Prophet for handling missing values ensured the completeness and robustness of our datasets. By backcasting and utilizing Prophet, we maximized the number of observations in our training set, eliminating the need to drop any rows.

Post-processing techniques, including the establishment and application of bounds based on historical percent changes, ensured that our forecasts remained within a historically plausible range, enhancing their reliability. Furthermore, our feature selection strategy, grounded in correlation analysis, effectively reduced multicollinearity, improving model interpretability and responsiveness to scenario-specific variables.



Moreover, our exploration of medium-term forecast scenarios underscored the versatility of our approach in capturing variances across diverse scenarios, enriching our understanding of global oil demand dynamics. The feature selection process enabled us to focus on the most impactful variables, providing a clearer insight into how specific economic and industry-specific factors might shape immediate and future demand trajectories.

Looking forward, several areas for further enhancement and exploration are clear. Enhanced feature engineering could provide additional relevant predictors and improve model accuracy. Dimensionality reduction techniques, such as Principal Component Analysis (PCA), could further mitigate multicollinearity and enhance model efficiency. Additionally, exploring additional advanced ML models and hybrid approaches could yield incremental gains in forecasting performance.

In conclusion, integrating ML techniques into global oil demand forecasting represents a significant advancement, offering enhanced accuracy, efficiency, and flexibility. Our methodology not only addresses the limitations of traditional econometric models but also provides a simple top-down approach instead of a complex bottom-up approach that looks at each demand center by product by region. By leveraging the strengths of ML models, we aim to contribute to economic stability and energy security, paving the way for more informed and effective decision-making in the face of an ever-evolving energy landscape.

### References

<sup>1</sup> Hopp, D. (2022). Benchmarking Econometric and Machine Learning Methodologies in Nowcasting. <sup>2</sup> Deb, P. (2019). Forecasting Health Care Spending: A Comparison of Nonlinear Econometric and Machine Learning Methods.

<sup>3</sup> Lukong, T. K., Tanyu, D. N., Tatietse, T., & Schulz, D. (2022). Long Term Electricity Load Forecast Based on Machine Learning for Cameroon's Power System.

<sup>4</sup> Liu, S., Wu, K., Jiang, C., Huang, B., & Ma, D. (2023). Financial Time-Series Forecasting: Towards Synergizing Performance And Interpretability Within a Hybrid Machine Learning Approach.

<sup>5</sup> Kontopoulou, V. I., Panagopoulos, A., Kakkos, I., & Matsopoulos, G. (2023). A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecast

<sup>6</sup> Oukhouya, H., & El Himdi, K. (2023). Comparing Machine Learning Methods—SVR, XGBoost, LSTM, and MLP— For Forecasting the Moroccan Stock Market. IOCMA 2023.

<sup>7</sup> Zhu, H. (2023). Oil Demand Forecasting in Importing and Exporting Countries: AI-Based Analysis of Endogenous and Exogenous Factors. Sustainability.

<sup>8</sup> Alkhammash, E. H., Kamel, A. F., Al-Fattah, S., & Elshewey, A. M. (2022). Optimized Multivariate Adaptive Regression Splines for Predicting Crude Oil Demand in Saudi Arabia. Discrete Dynamics in Nature and Society.

<sup>9</sup> Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016.

<sup>10</sup> Singh, D., & Rawat, N. (2023). Machine Learning for Weather Forecasting: XGBoost vs SVM vs Random Forest in Predicting Temperature for Visakhapatnam. International Journal of Intelligent Systems and Applications.

<sup>11</sup> Dong, J., Zeng, W., Wu, L., Huang, J., Gaiser, T., & Srivastava, A. (2023). Enhancing short-term forecasting of daily precipitation using numerical weather prediction bias correcting with XGBoost in different regions of China. Engineering Applications of Artificial Intelligence.

<sup>12</sup> Dezhkam, A., & Manzuri, M. T. (2023). Forecasting stock market for an efficient portfolio by combining XGBoost and Hilbert-Huang transform. Engineering Applications of Artificial Intelligence.
<sup>13</sup> Challu, Cristian, et al. "NHITS: Neural Hierarchical Interpolation for Time Series Forecasting." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 37. No. 6. 2023.

<sup>14</sup> Souza, G. N., Mendes, A. G. B., Costa, J. D. S., Oliveira, M. S., Lima, P. V. C., Moraes, V. N., Silva, D. C. C., Rocha, J. E. C., Botelho, M. N., Araujo, F. A., Fernandes, R. S., Souza, D. L., & Braga, M. (2023). Deep learning framework for epidemiological forecasting: A study on COVID-19 cases and deaths in the Amazon state of Pará, Brazil. PLOS ONE.

<sup>15</sup> The Pandas Development Team. (2023). Pandas: Python data analysis library (Version 2.2.1) [Software]. Available from https://pandas.pydata.org/

<sup>16</sup> Taylor, S. J., & Letham, B. (2018). Forecasting at scale. The American Statistician, 72(1), 37-45. For Prophet package version details, see the Prophet website: https://facebook.github.io/prophet/

<sup>17</sup> Amorim, L.B., Cavalcanti, G.D., & Cruz, R.M. (2022). The choice of scaling technique matters for classification performance. Appl. Soft Comput., 133, 109924.

<sup>18</sup> How XGBoost works - amazon sagemaker. (n.d.).

https://docs.aws.amazon.com/sagemaker/latest/dg/xgboost-HowItWorks.html

<sup>19</sup> Olivares, K. G., Challú, C., Garza, F., Mergenthaler Canseco, M., & Dubrawski, A. (2022).

NeuralForecast: User friendly state-of-the-art neural forecasting models. PyCon Salt Lake City, Utah, US 2022. Retrieved from https://github.com/Nixtla/neuralforecast