FORECASTING ENERGY CONSUMPTION IN THE PHILIPPINES USING MACHINE LEARNING ALGORITHMS

PRESENTED TO

Ara Abigail Ambita

PRESENTED BY

Earl James Rentillo

Erru Torculas

OUTLINE

1. Introduction
    - Motivations
2. Methodology
    - Dataset and Preprocessing
    - ML algorithms and setting of hyperparameters
3. Results
    - Highest scoring models by RMSE
4. Conclusion and Recommendations

MOTIVATION

1. **LACK OF DATASETS IN DEVELOPING ECONOMIES TO BE ABLE TO PREDICT FUTURE DEMANDS IN ELECTRICITY**
   Philippine average electricity generation per person peaked at 961 KWh in 2019 (Ritchie et al., 2019).

2. **PROMOTE ENERGY CONSUMPTION TREND AS A FACTOR TO WORK ON BETTER POLICY-MAKING ESPECIALLY ON SUSTAINABLE ENERGY USE**
   Philippine population's access to electricity is at an all-time high at 96.84% (Ritchie et al., 2019).

MOTIVATION

3. **AID FUTURE RESEARCH TO OTHER DEVELOPING COUNTRIES AND FOREGROUND THE GENERATION OF MORE ACCURATE RESULTS**
   Consumption of energy as a resource exhibits a continuous rise to developing economies (Barak & Sadegh, 2016).

4. **FIND OUT THE EFFECT OF THE DISRUPTION CAUSED BY THE COVID-19 PANDEMIC TO MODEL ACCURACY**
   The dataset spans from January 2014 to June 2022, encompassing the pandemic period which may be a factor to model accuracy.

METHODOLOGY

Aim to assess different machine learning algorithms in forecasting
time series of energy consumption in the Philippines and compare the viability and accuracy of each forecasting

DATASET

MONTHLY AND QUARTERLY MARKET ASSESSMENT REPORT OF THE WHOLESALE ELECTRICITY SPOT MARKET (WESM) OF PHILIPPINE ELECTRICITY MARKET CORPORATION (PEMC)

Energy-related Reports from January 2014 to the Second Quarter (June) of 2022

DEPARTMENT OF ENERGY (DOE) 2020 SUMMARY POWER STATISTICS

Basis for Power Consumption by Sector, specifically, the Commercial and Industrial data

PREPROCESSING

Date and **Total Energy Consumption (in GWh)**
Univariate approach for analysis

Period 1: Pre-pandemic
January 2014 to March 2020

Period 2: Pandemic
January 2014 to June 2022

Training Data
January 2014 to June 2018
Testing Data
July 2018 to March 2020

Training Data
January 2014 to February 2020
Testing Data
March 2020 to June 2022

MACHINE LEARNING MODELS

Scikit-learn Python library

Random Forest (RF) Model
XGBoost
Linear Regression
Support Vector Regression (SVR)

MODELS

1	**RandomForestRegressor()** *n_estimators = 10000* *max_features = 4*
2	**XGBRegressor()** *n_estimators = 10000*
3	**LinearRegression()** *No hyperparameters for tuning*
4	**SVR()** *kernel = linear* *C = 10000* *epsilon = 10*

METRICS

1	**Root Mean Squared Error (RMSE)**
2	**Mean Absolute Percentage Error (MAPE)**

Table 2: Interpretation of MAPE Results for Forecasting Accuracy.

MAPE-value	Accuracy of Forecast
Less than 10%	Highly Accurate Forecast
11% to 20%	Good Forecast
21% to 50%	Reasonable Forecast
More than 51%	Inaccurate Forecast

RESULTS

Table 1: Performance of the models by period

		ML Models			
	Metric	RF	XGBoost	Linear Reg	SVR
Period 1	RMSE	$\boxed{422.737}$	366.691	411.578	431.366
	MAPE	0.050	0.044	0.047	0.050
Period 2	RMSE	687.665	$\boxed{692.077}$	935.880	982.202
	MAPE	0.061	0.061	0.123	0.131

RESULTS

RANDOM FOREST (RF) MODEL

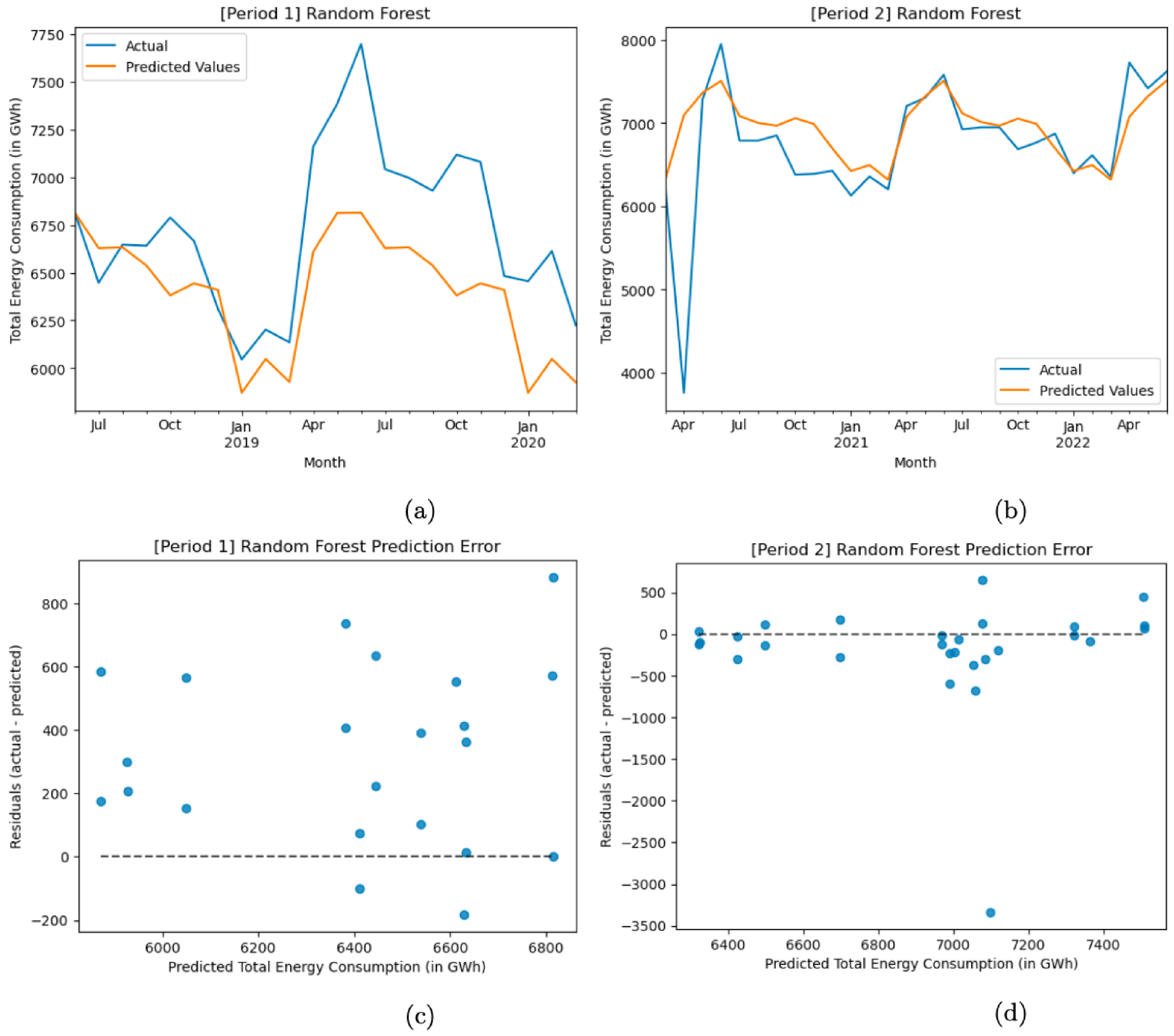


Fig. 5: Random forest model comparison by period with prediction error.

RESULTS

XGBOOST MODEL

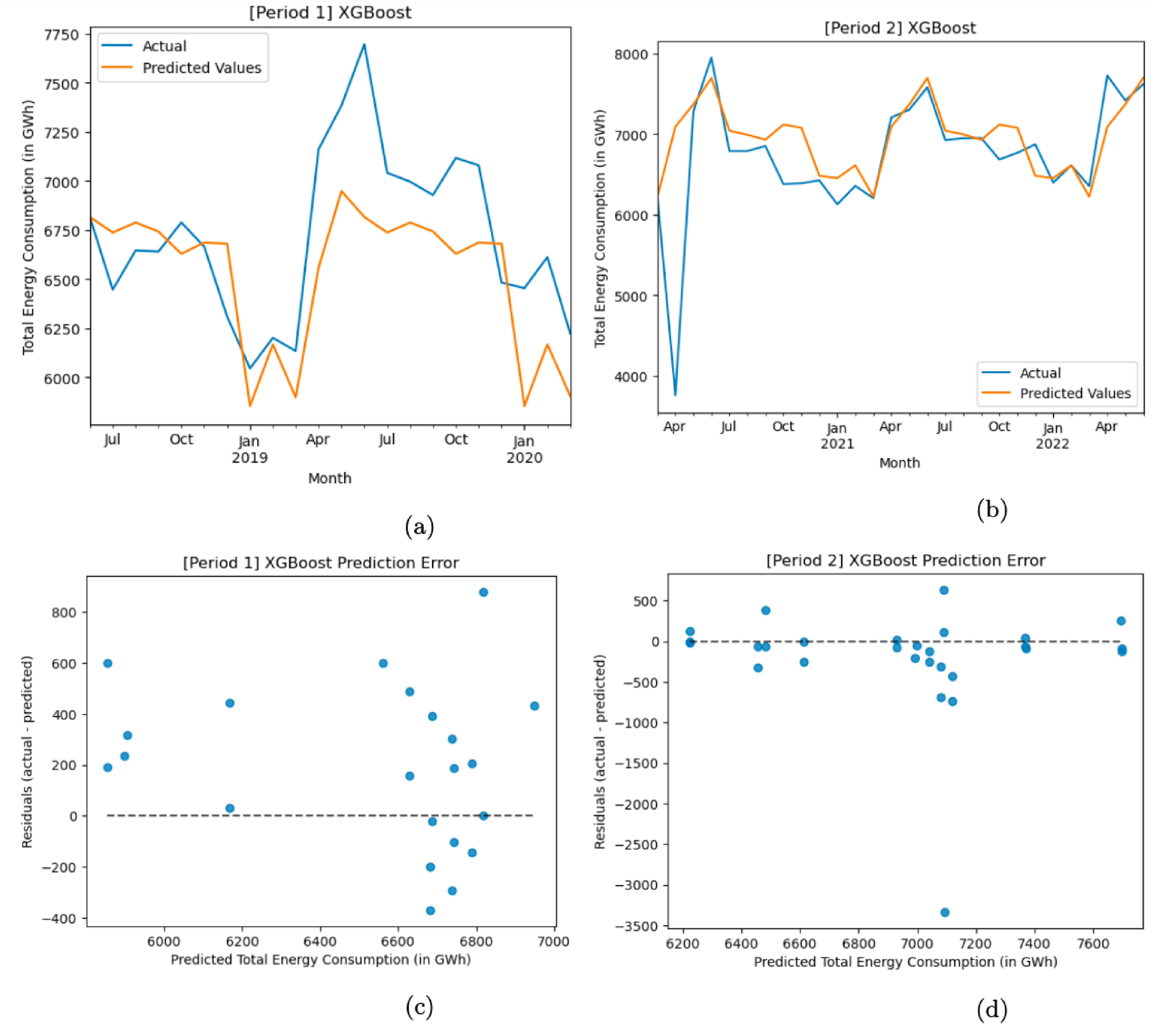


Fig. 6: XGBoost model comparison by period with prediction error.

RESULTS

LINEAR REGRESSION

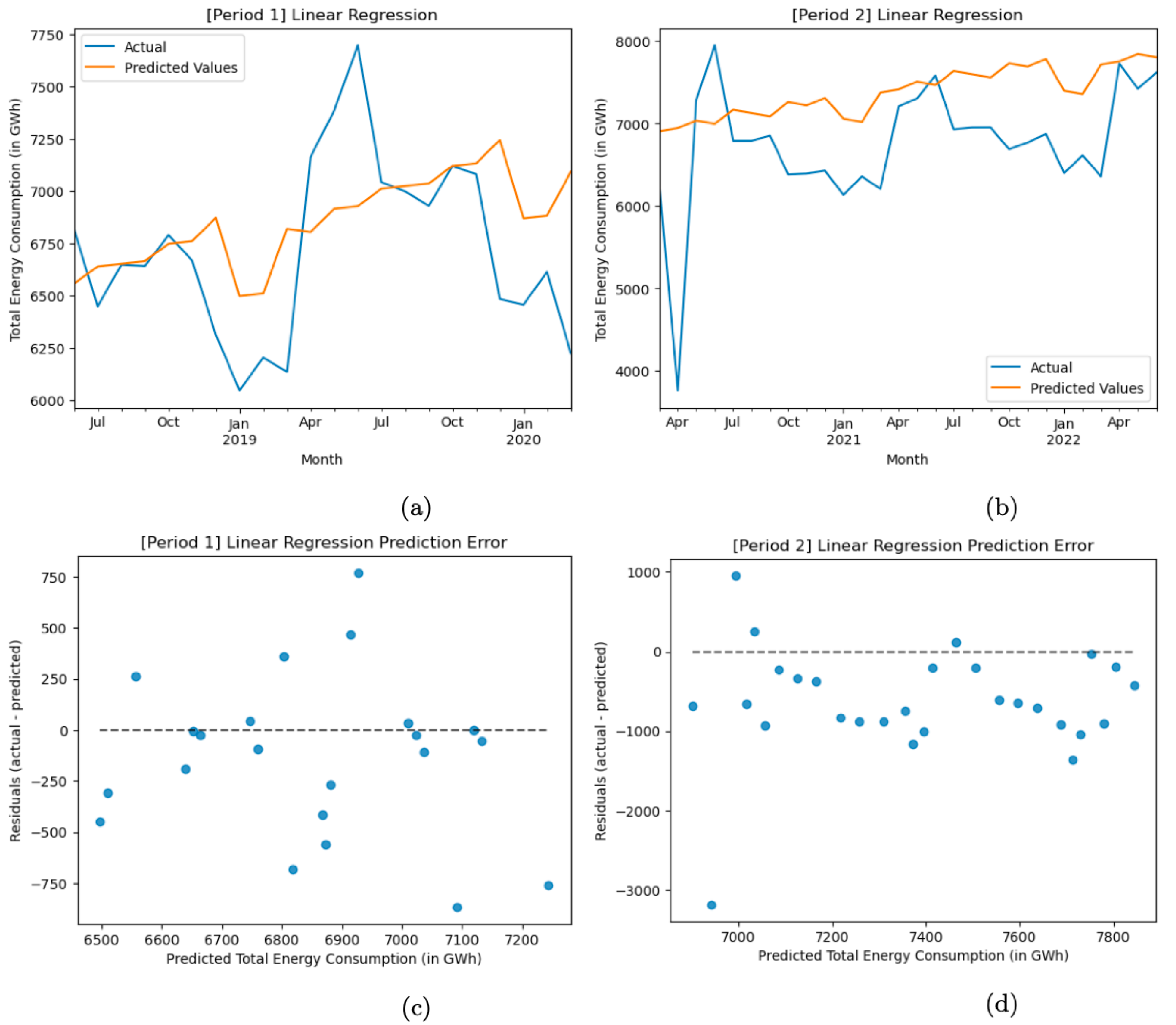


Fig. 7: Linear Regression model comparison by period with prediction error.

RESULTS

SUPPORT VECTOR REGRESSION (SVR) MODEL

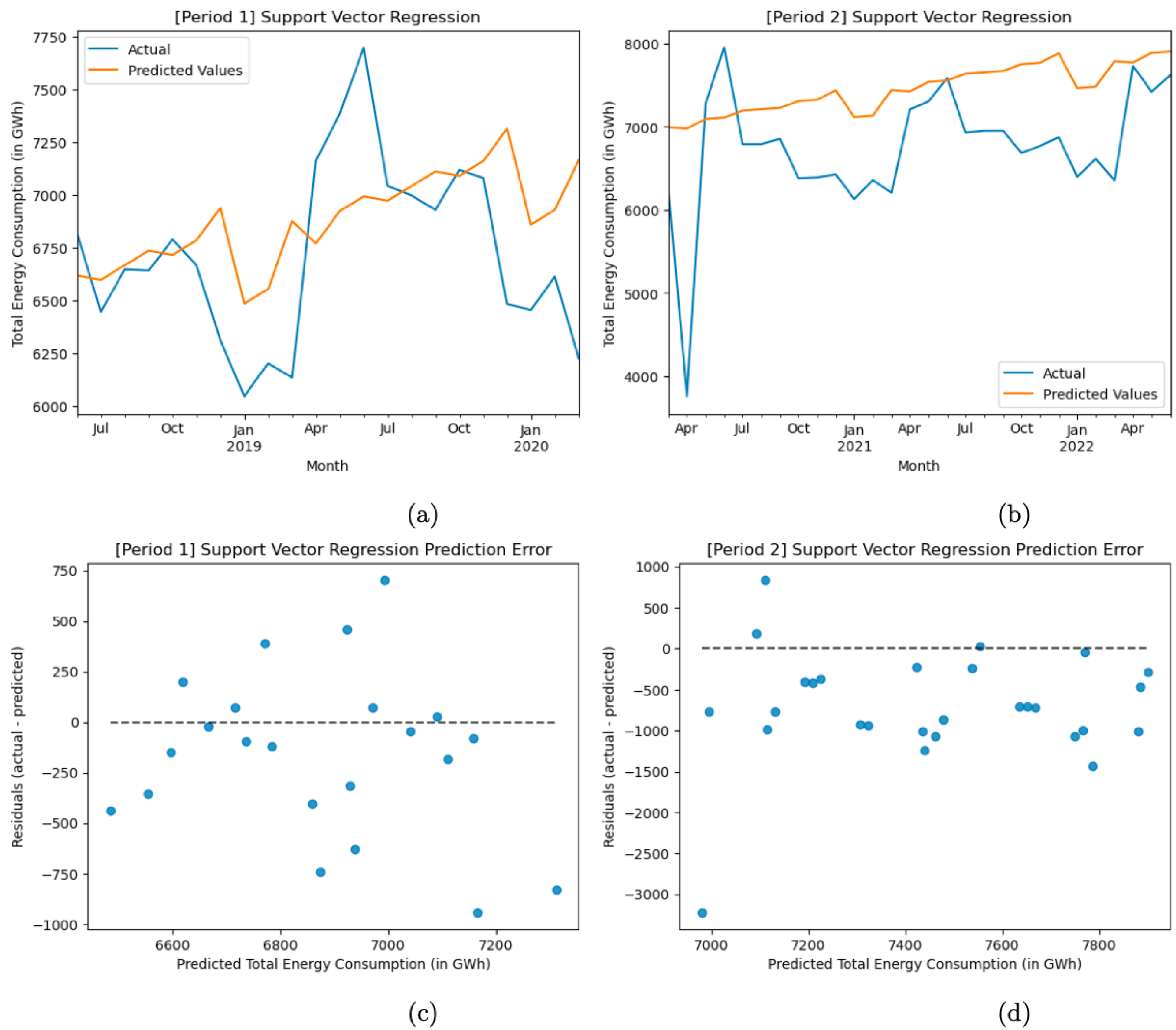
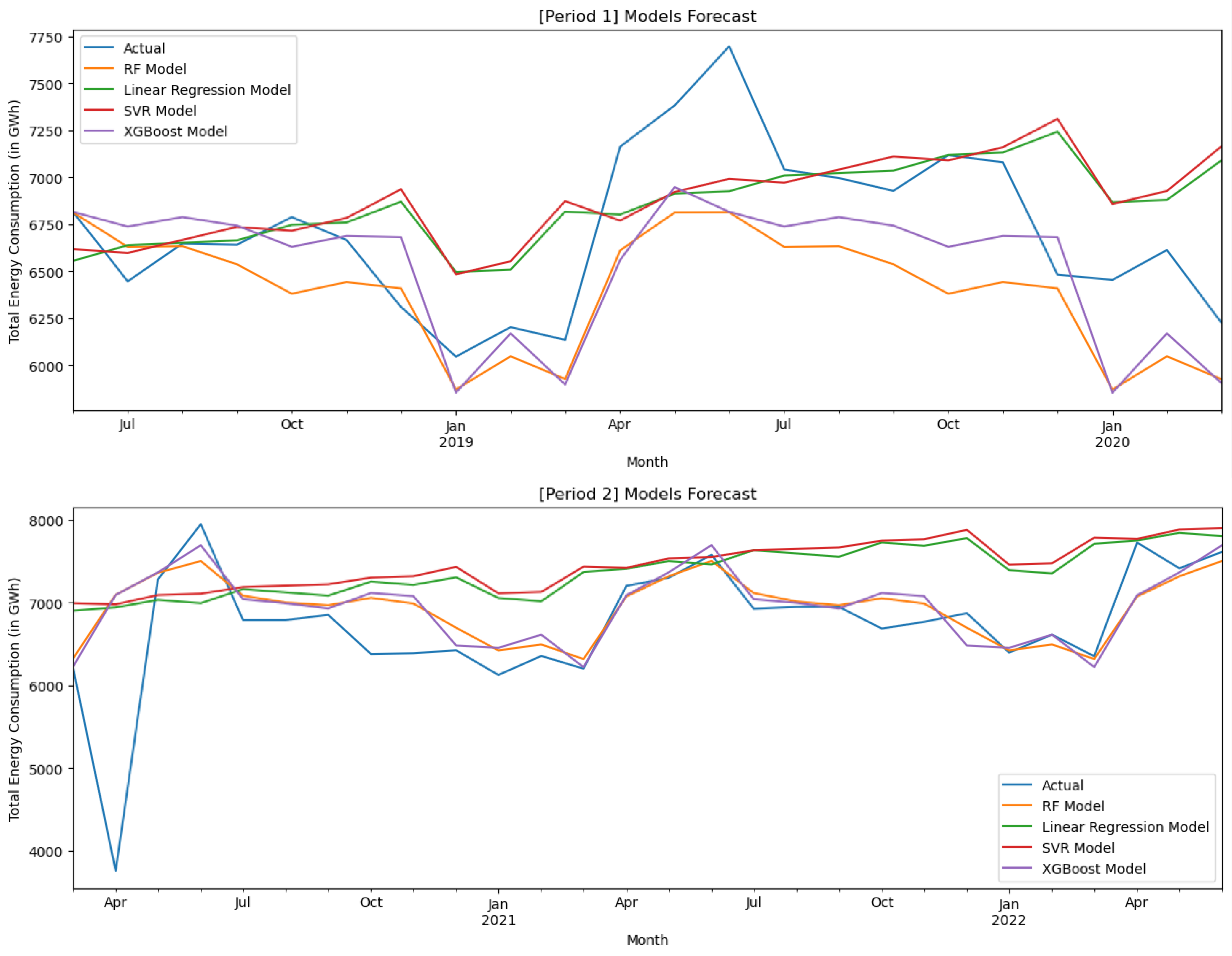


Fig. 8: SVR model comparison by period with prediction error.

RESULTS



[Period 1] Models Forecast
Total Energy Consumption (in GWh)
7750
7500
7250
7000
6750
6500
6250
6000
Actual
RF Model
Linear Regression Model
SVR Model
XGBoost Model
Jul
Oct
Jan
2019
Apr
Jul
Oct
Jan
2020
Month
[Period 2] Models Forecast
Total Energy Consumption (in GWh)
8000
7000
6000
5000
4000
Actual
RF Model
Linear Regression Model
SVR Model
XGBoost Model
Apr
Jul
Oct
Jan
2021
Apr
Jul
Oct
Jan
2022
Apr
Month

CONCLUSION

1. THE XGBOOST MODEL WAS THE MOST ACCURATE MODEL IN PREDICTING THE FIRST PERIOD, WHILE THE RANDOM FOREST MODEL HAD THE LOWEST RMSE IN THE SECOND PERIOD.

2. DEVELOPING PREDICTIVE MODELS USING XGBOOST AND RANDOM FOREST MODELS IS SIGNIFICANT TO FORECASTING ENERGY CONSUMPTION IN THE PHILIPPINES AND IS BENEFICIAL FOR FUTURE STUDIES THAT AIM TO ENGAGE IN THE SAME CONTEXT

RECOMMENDATIONS

1. STUDY HOW THE ACCURACY OF THE MODEL'S PREDICTION AND ANALYSIS CAN VARY DEPENDING ON THE DATA AND VARIABLE SETTINGS AND DETERMINE WHICH APPROACH IS SUPERIOR IN ALL CASES.

2. STUDY HOW FUTURE, POST-PANDEMIC DATA AFFECTS MODEL ACCURACY

3. INVESTIGATE ON TIME SERIES FORECASTING ALGORITHMS SUCH AS AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) AND SEASONAL TREND DECOMPOSITION TO ENHANCE THE ANALYSIS BY INTEGRATING ENERGY CONSUMPTION BY DAY OR SEASONS