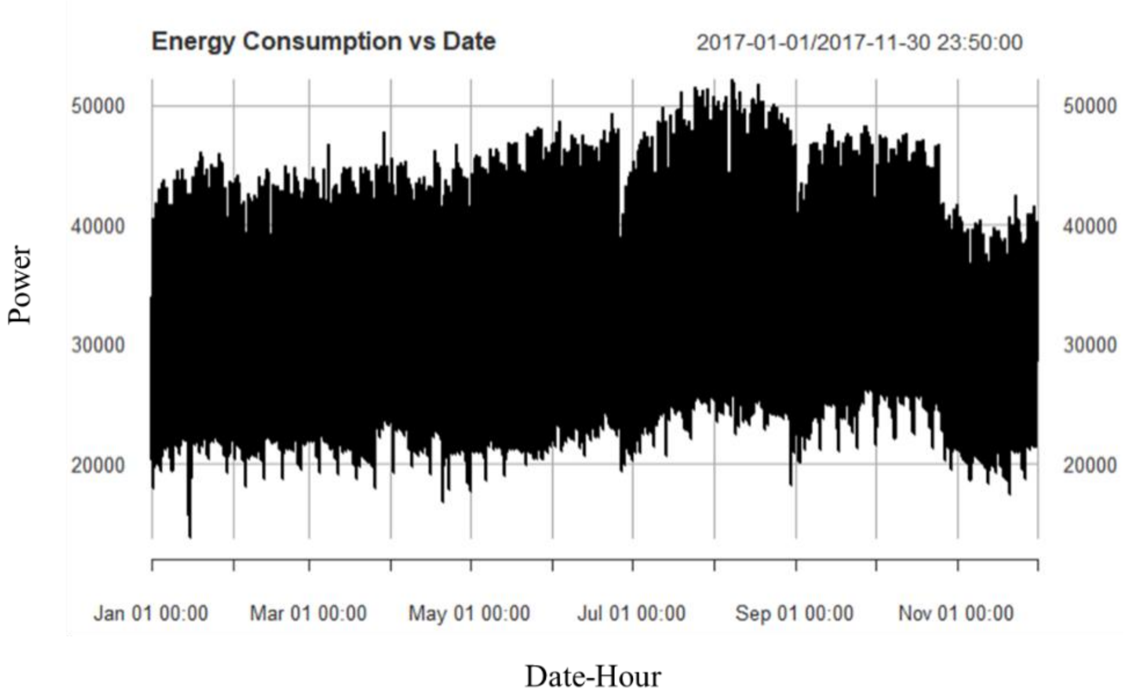
University of Milano-Bicocca
Master's Degree in Data Science
Streaming Data Management and Time Series Analysis
Academic Year 2022-2023

ELECTRICITY CONSUMPTION FORECASTING USING ARIMA, UCM AND MACHINE LEARNING MODELS

Author:
Giorgio CARBONE matr. n° 811974

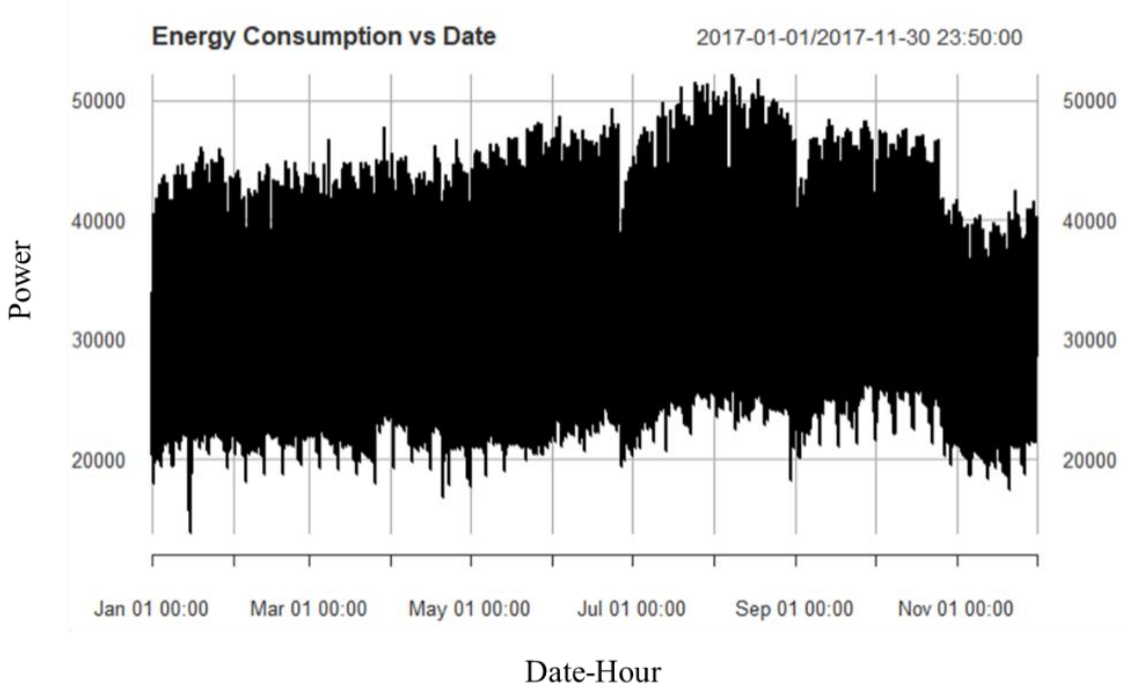
Introduction

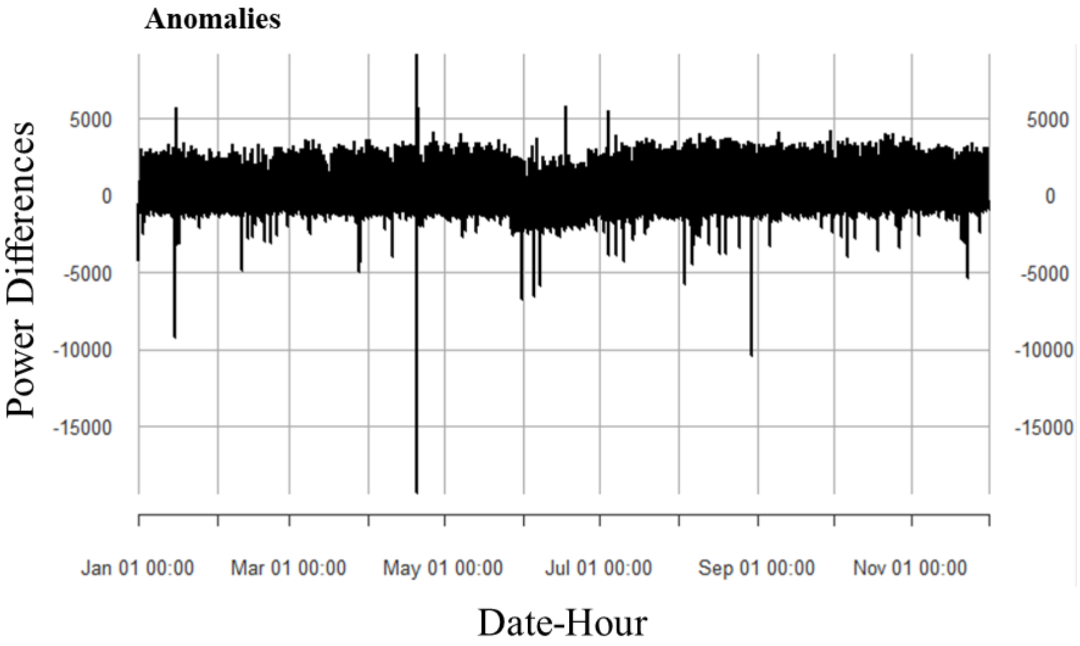
- **Univariate**, **regular**, **high-frequency** time series
    - **48096** electricity consumption measured every **10 minutes**
    - from 01/01/2017 00:00:00 to 30/11/2017 23:50:00
- Comparing **different approaches to modelling**
    - Statistical approaches → **ARIMA, UCM**
    - **Machine Learning** → Random Forest, k-NN
    - **Deep Learning** → GRU Recurrent Neural Networks
- **Goal** → predict the 4320 observations of December 2017
    - from 01/12/2017 00:00:00 to 30/12/2017 23:50:00



Data Exploration

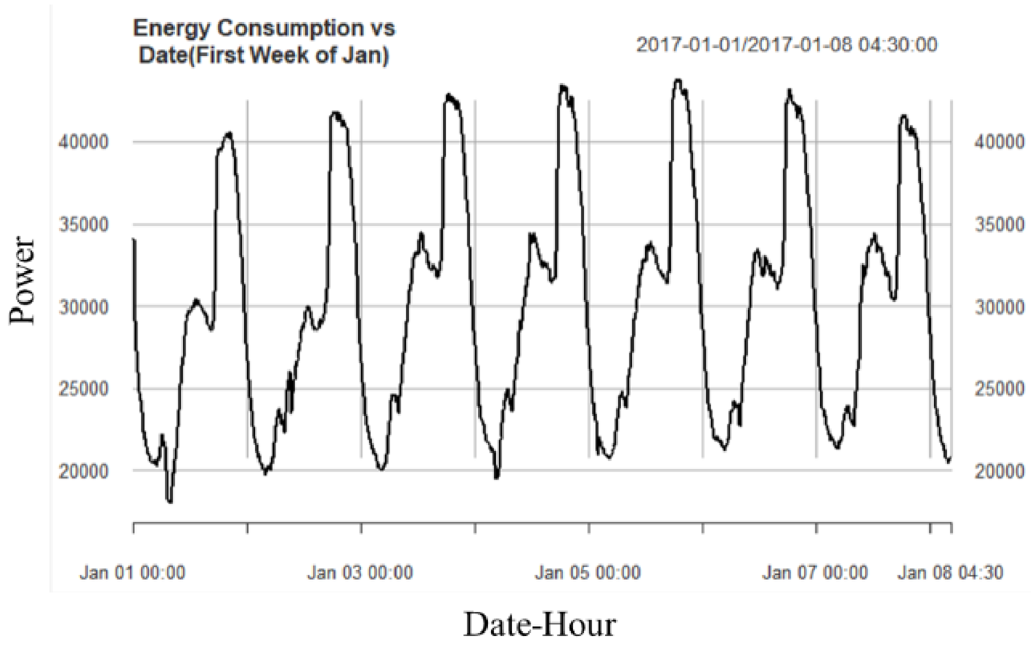
- ❑ **Two variables** → **Power**, Date (date-hour labels)
    - ❑ **no duplicates** or **missing** values
- ❑ energy consumption **increases** in **summer**
- ❑ **Yearly seasonality**
    - ❑ caused by the cycle of the seasons
    - ❑ **not observable** (only one year of data)
- ❑ Rare **anomalies**



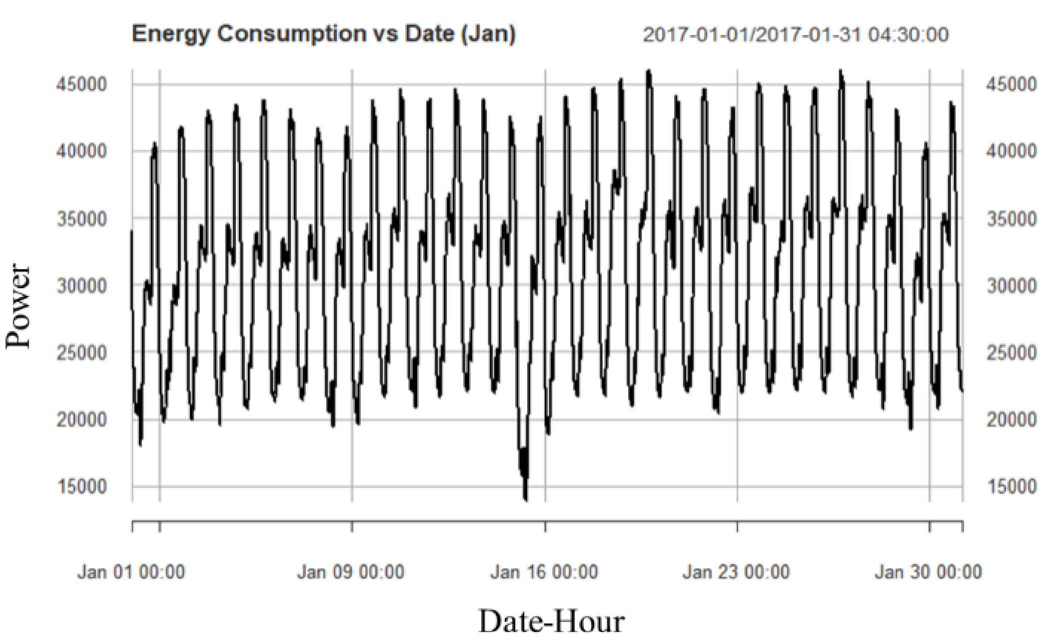


Data Exploration

- ❑ **Daily seasonality**
  - ❑ day-night cycle
  - ❑ 24-hour period → **144** observations

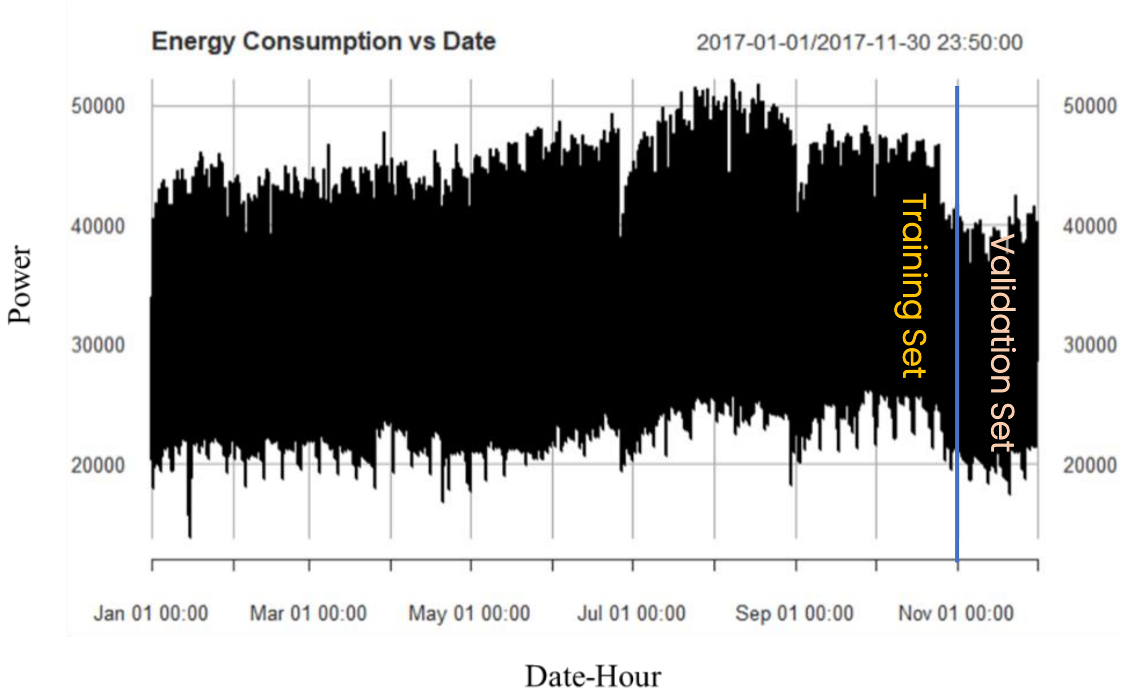


- ❑ **Weekly seasonality**
  - ❑ weekend consumption reduction
  - ❑ 7-days period → **1008** observations



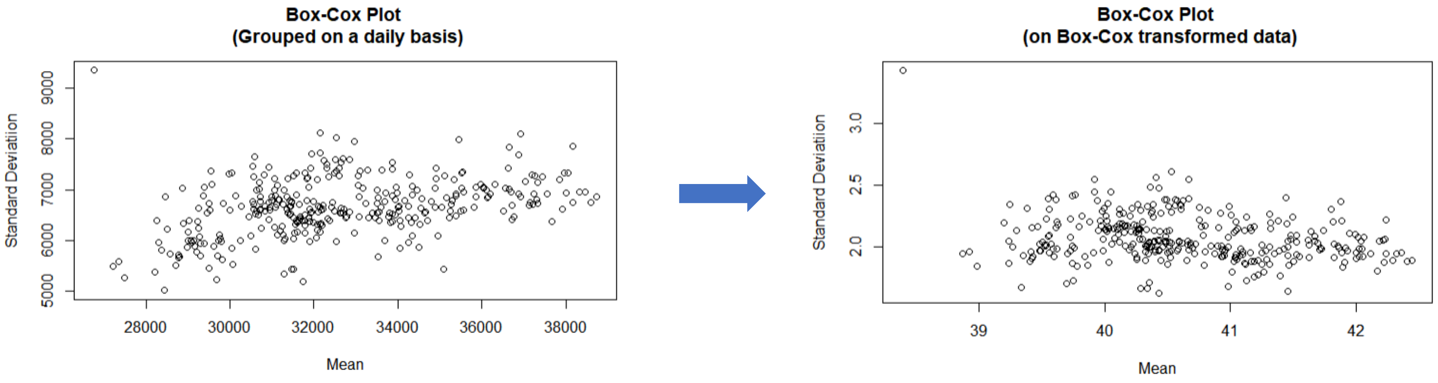
Hold-out models evaluation method

- **Training Set**
  - From 01/01/2017 00:00:00 to 31/10/2017 23:50:00
  - 43776 observations
- **Validation Set**
  - From 01/11/2017 00:00:00 to 30/11/2017 23:50:00
  - **4320** observations → consistent with Dec **forecast horizon**
- Choosing the **models** with the **lowest MAE** on **validation** set → for each type
  - highest generalization capabilities
  - **re-estimated** on **training+validation**
  - used to **predict** from **01/12**/2017 00:00:00 to **30/12**/2017 23:50:00

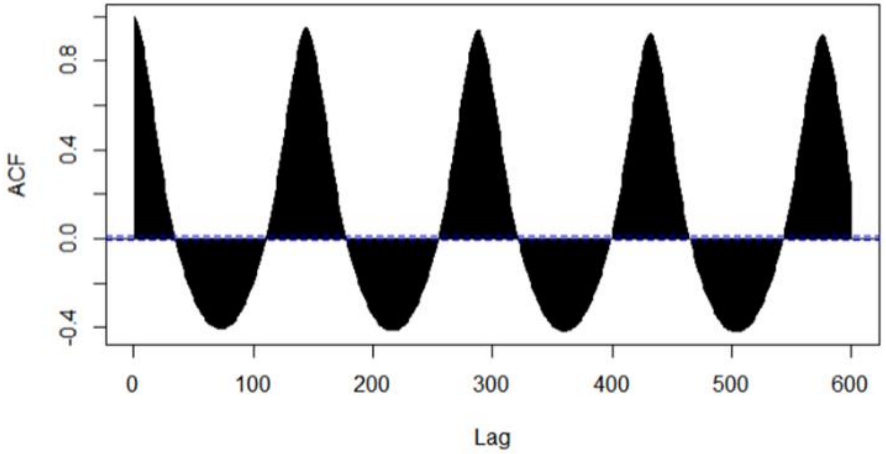


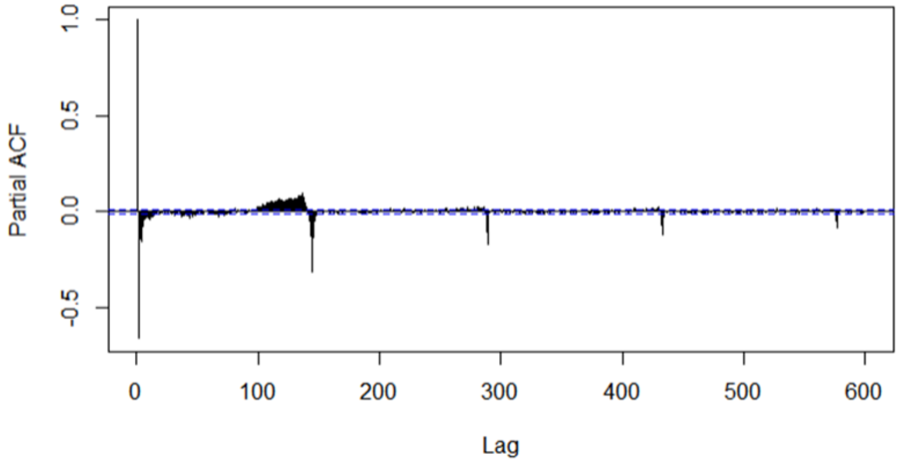
ARIMA MODELS

ARIMA: stationarity



- ❑ Non-stationarity in **variance**
  - ❑ Box-Cox plot → Slight linear correlation
  - ❑ Models estimated on Box-Cox transformed time series → performed worse → **no transformation applied**
- ❑ Non-stationarity in **mean**
  - ❑ **trend** and **multiple seasonalities**
  - ❑ **Augmented Dickey-Fuller** and **KPSS** (5%)
    - ❑ **rejected** hypothesis of need for **simple difference**
    - ❑ need for a **seasonal difference** with **lag=144**

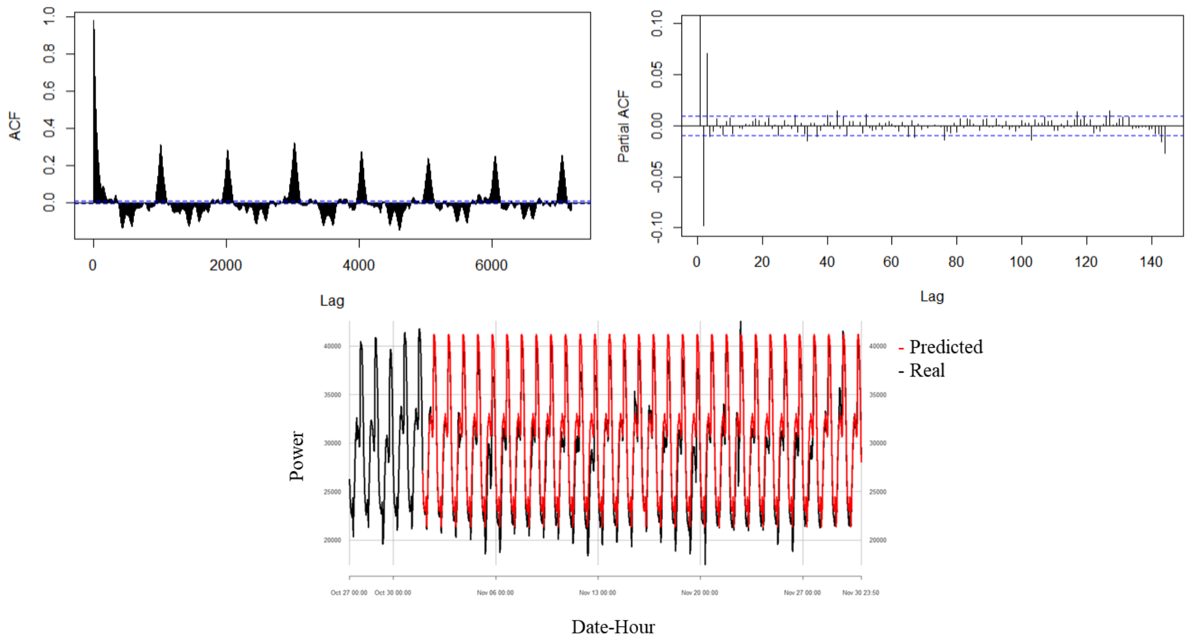




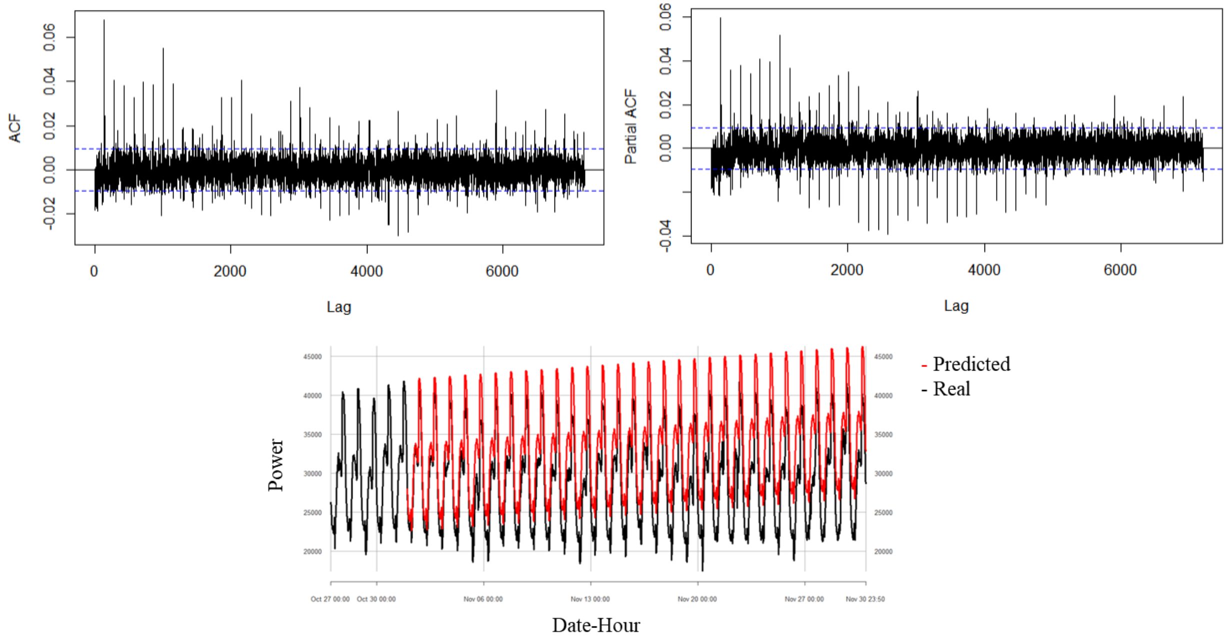
ARIMA: no regressors

- **First** model → $\text{ARIMA}(0,0,0)(0,1,1)_{144}$ (1.1)
  - **SMA(1)** and **seasonal difference (p=144)**
  - $\text{MAE}_{\text{validation}} = 1415.85$
  - Residuals ACF → linear memory in early lags and multiples of 1008 (weekly seasonality)
  - Residuals PACF → linear memory in first 3 lags
- Tried to **capture** more **linear memory**
  - $\text{ARIMA}(3,1,0)(0,1,1)_{144}$ (1.7)
  - Lowest residuals ACF and PACF values → good adaptation to training data
  - **High** $\text{MAE}_{\text{validation}}$ → poor predictive capabilities

[1.1] Residuals ACF and PACF and validation set predictions



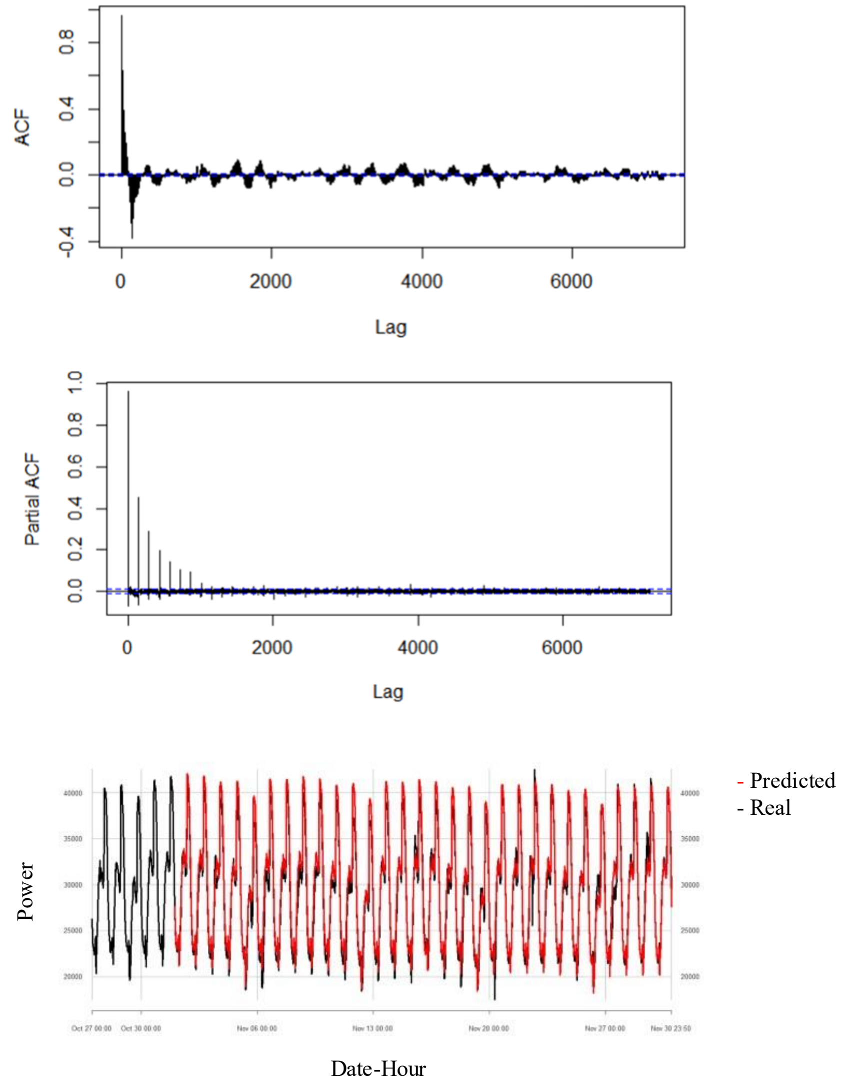
[1.7] Residuals ACF and PACF and validation set predictions



ARIMAX

- model **weekly** seasonality using **6 dummy** regressors
  - $ARIMA(2,1,0)(0,1,1)_{144}\ +\ 6$ dummies [1.10]
    - captures a lot of linear memory
    - $\text{MAE}_{\text{validation}} = 4116.94$
  - $ARIMA(0,0,0)(1,0,0)_{144}\ +\ 6$ dummies [1.11]
    - $\text{MAE}_{\text{validation}} = 1065.45$
- model **weekly** seasonality using **sinusoids**
  - $ARIMA(0,0,0)(1,0,0)_{144}$ + 8 sinusoids (p = 1008) [1.12]
  - $\mathbf{MAE}_{\boldsymbol{validation}} = \mathbf{1010.08}$ → best ARIMA model
- model **daily** seasonality using **sinusoids** + **weekly** seasonality using **dummies** → no improvements in prediction accuracy
  - $ARIMA(0,0,0)(1,0,0)_{144}$ + 8 sinusoids (p = 144) [1.13]
  - $ARIMA(0,0,0)(1,0,0)_{144}$ + 8 sinusoids (p = 144) + 6 dummies [1.14]

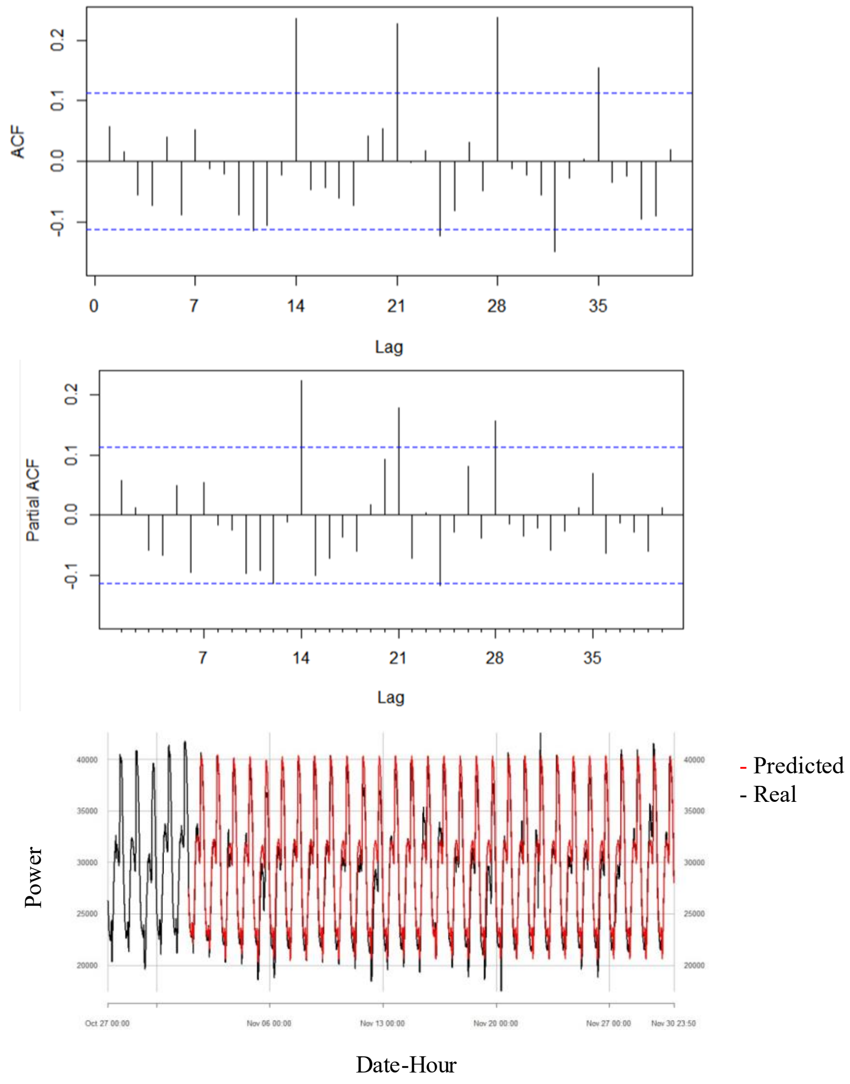
[1.12] Residuals ACF and PACF and validation set predictions



ARIMA: time series aggregated by hour

- ❑ Time series **aggregated by hour**
  - ❑ **24 time series** → 24 hours of the day
  - ❑ **averaging the 6 values** within each hour
- ❑ Evaluation
  1. **Fitted** the **same model independently** on each of the 24 time series
  2. **Predicted** the **hourly observations** for November 2017
  3. **Restored 10-minute** granularity by **cubic interpolation** using spline functions
  4. Computed $\mathrm{MAE}_{\mathrm{validation}}$
- ❑ $ARIMA(0,1,1)(0,0,1)_7$ [1.15]
  - ❑ captures a lot of linear memory
  - ❑ $\mathrm{MAE}_{\mathrm{validation}} = 1142.87$

[1.15] Residuals ACF and PACF and validation set predictions



UCM MODELS

UCM: time series aggregated by hour

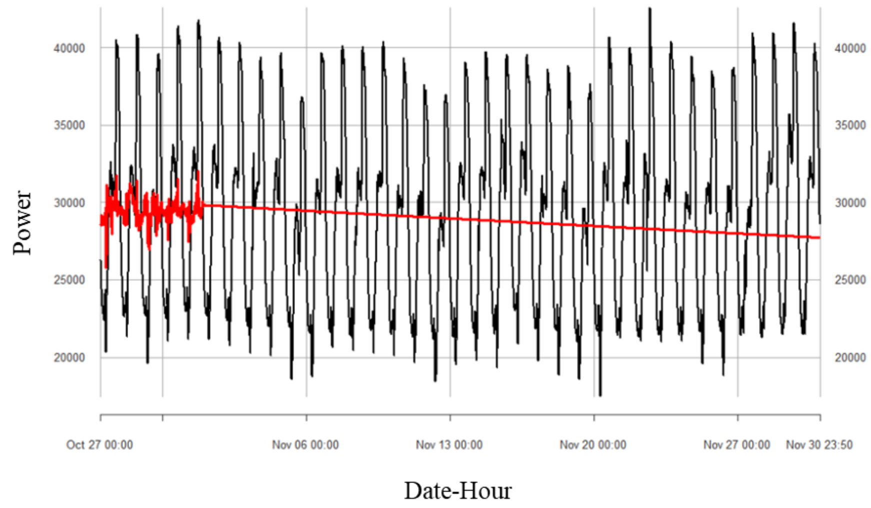
- Models are trained in time series **aggregated by hour** (**averaging** the 6 values within each hour)
  - **predicting** the **720 hourly observations** for **November** 2017
  - **restored 10-minute granularity** by cubic interpolation
- $\hat{y} = TREND + SEAS_{daily}(24h) + SEAS_{weekly}(168h)$
- Different models and combinations tested

1. $LLT + 10\ stochastic\ sinusoids_{weekly}(p = 168) + (\mathbf{24}\ stochastic\ \boldsymbol{dummies}_{\boldsymbol{daily}}\ \boldsymbol{VS}\ stochastic\ \boldsymbol{sinusoids}_{\boldsymbol{daily}}(p = 24))$
   - $\mathbf{24}\ stochastic\ \boldsymbol{dummies}_{\boldsymbol{daily}} \rightarrow$ better accuracy
2. $LLT + \left(10\ \boldsymbol{stochastic}\ \boldsymbol{sinusoids}_{\boldsymbol{weekly}}(p = 168)\ VS\ \boldsymbol{stochastic}\ \boldsymbol{cycle}_{\boldsymbol{weekly}}(p = 168)\right) + 24\ stochastic\ dummies_{daily}$
   - $\boldsymbol{stochastic}\ \boldsymbol{cycle}_{\boldsymbol{weekly}}(p = 168) \rightarrow$ very high $\mathrm{MAE}_{\mathrm{validation}}$
3. $LLT + (1 - 16)\ \boldsymbol{stochastic}\ \boldsymbol{sinusoids}_{\boldsymbol{weekly}}(p = 168) + 24\ stochastic\ dummies_{daily}$ [2.1]
   - Testing the **optimal number** of weekly **sinusoids** for the **final model**
   - **2** sinusoids $\rightarrow MAE_{validation} = 1210.08$ and **6** sinusoids $\rightarrow MAE_{validation} = 1299.41$

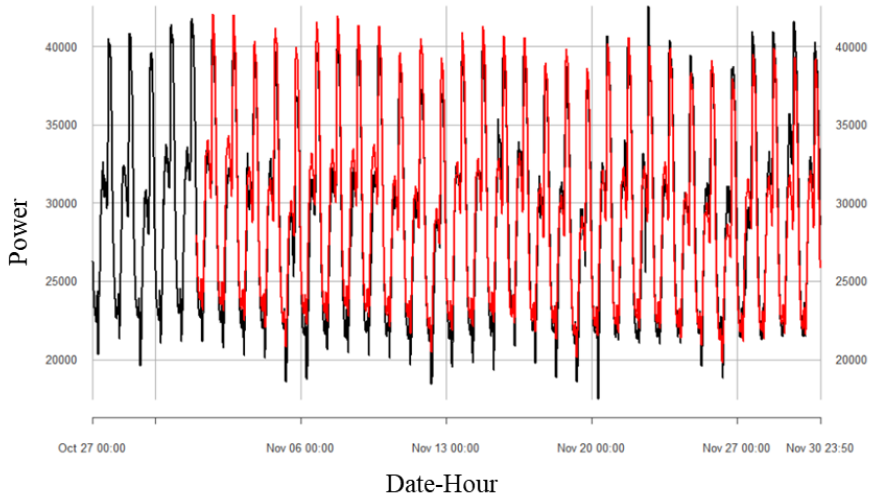
UCM: original time series

- Original 10-minutes frequency time series
  - **Training set: September and October** 2017 **observations**
  - Applied **logarithm** → improved performance
- Final and **best UCM model** [2.2] → $\boldsymbol{MAE_{validation} = 1183.40}$
  - $\boldsymbol{Local\ Linear\ Trend\ (LLT)} +$
  - $\boldsymbol{10\ stochastic\ sinusoids_{weekly}(p = 1008)} +$
  - $\boldsymbol{10\ stochastic\ sinusoids_{daily}(p = 144)}$
- Initial parameters (vy = time series variance)
  - $logVarEta: \frac{vy}{10}$ | $logVarZeta: \frac{vy}{10000}$
  - $logVarOmega144: \frac{vy}{10000}$ | $logVarOmega1008: \frac{vy}{10000}$
  - $logVarEpsilon: \frac{vy}{10}$
  - $a_1 = mean(time\ series)$ | $P_1 = vy * 10$

[2.2] Trend estimate and validation set predictions



- Predicted
- Real



MACHINE LEARNING MODELS

Deep Learning: GRU Recurrent Neural Networks

- tested different **hyperparameters**:
  - **number** of **GRU** layers and number of relative neuronal activities
  - presence or absence of **dropout layer** after GRU layers
  - batch size, epochs
  - learning rate and optimizer
  - number of **lag** and **lead**
  - activation function
  - **dense layers size**
- **Trained** on **original 10-minute** frequency time series
  - **Standardized** $(\mu = 0, \sigma = 1)$
  - **Differentiated** (**144** observations)
  - Using **lags as regressors**

Deep Learning: GRU Recurrent Neural Networks

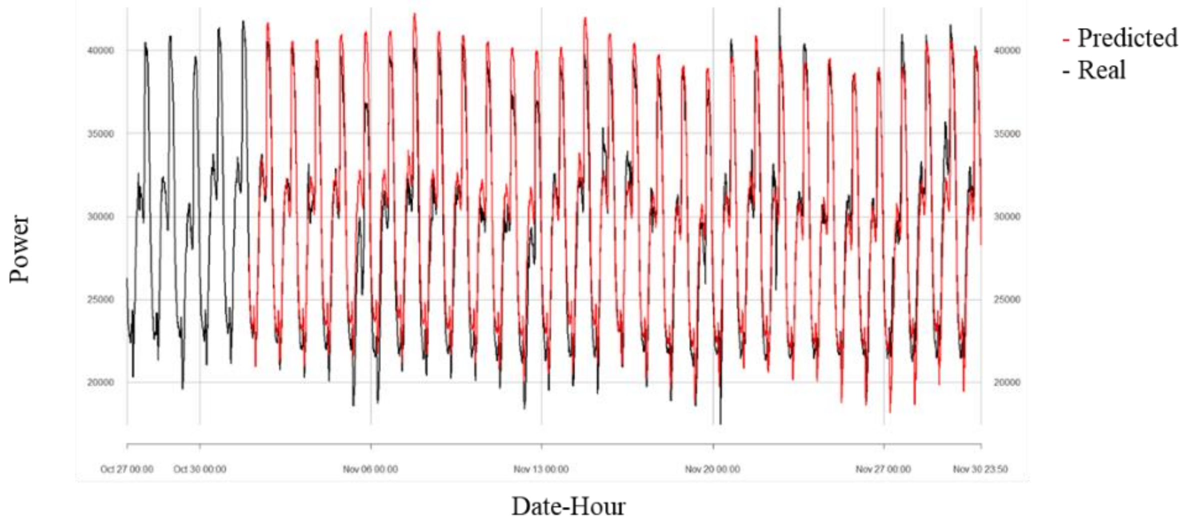
- Final Network **Hyperparameters**:
  - Batch Size: 144
  - Epochs = 40
  - **Lag** = 4320, **Lead** = 4320 (**multi-step prediction**)
  - LR = 0.001, Optimizer = Adam, Loss = MAE
- Network **Architecture:**

Input Layer

GRU Layer (16 units, no dropout)
Hyperbolic tangent Activation

GRU Layer (16 units, no dropout)
Hyperbolic tangent Activation

Dense Layer (FC, 30 neurons)
Linear Activation



- Good forecasting accuracy $\rightarrow \mathrm{MAE}_{\mathrm{validation}} = 1184.07$
  - best ML model

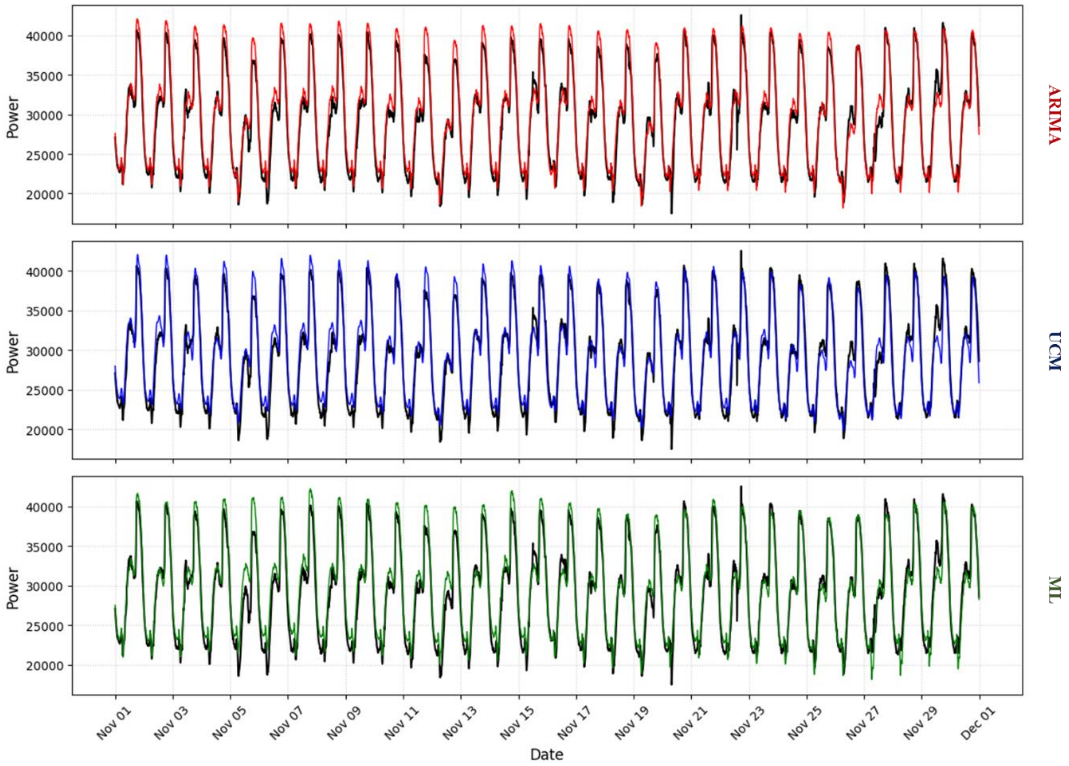
Machine Leaning: Random Forest and k-NN

- **Trained** on **original 10-minute** frequency time series
  - Using **lags as regressors**
- Random Forest
  - tested various **hyperparameters** configurations:
    - number of **lags**, number of **trees,** using external **regressors** (hour of the day and day of the week dummies)
    - **recursive** forecasting (using **estimates as new regressors**)
  - **Best Model: lags = 144*7, 250 trees, day of the week dummies** → $\text{MAE}_{\text{validation}} = 2350.14$
- **k-NN**, best model → $\text{MAE}_{\text{validation}} = 1694.75$
  - 144*7 **= 1008 lags**
  - Multi-Step Ahead Strategy **(MIMO) forecasting** → lead = 4320
  - **median** function to **aggregate** values

FINAL RESULTS

The most accurate models on the validation set

Model Class	Best Model for each class	$MAE_{validation}$
ARIMA	$ARIMA(0,0,0)\,(1,0,0)_{144}$ + 8 weekly sinusoids (p=1008)	1010.08
UCM	LLT $+$ 10 stochastic $\text{sinusoids}_{\text{weekly}}(p = 1008)$ $+$ 10 stochastic $\text{sinusoids}_{\text{daily}}(p = 144)$	1183.40
ML	Gated Recurrent Units (GRU) Network	1184.07



- **Best models** for each class
  - are **re-trained** on **training+validation**
  - used to **predict** the 4320 **December** 2017 observations