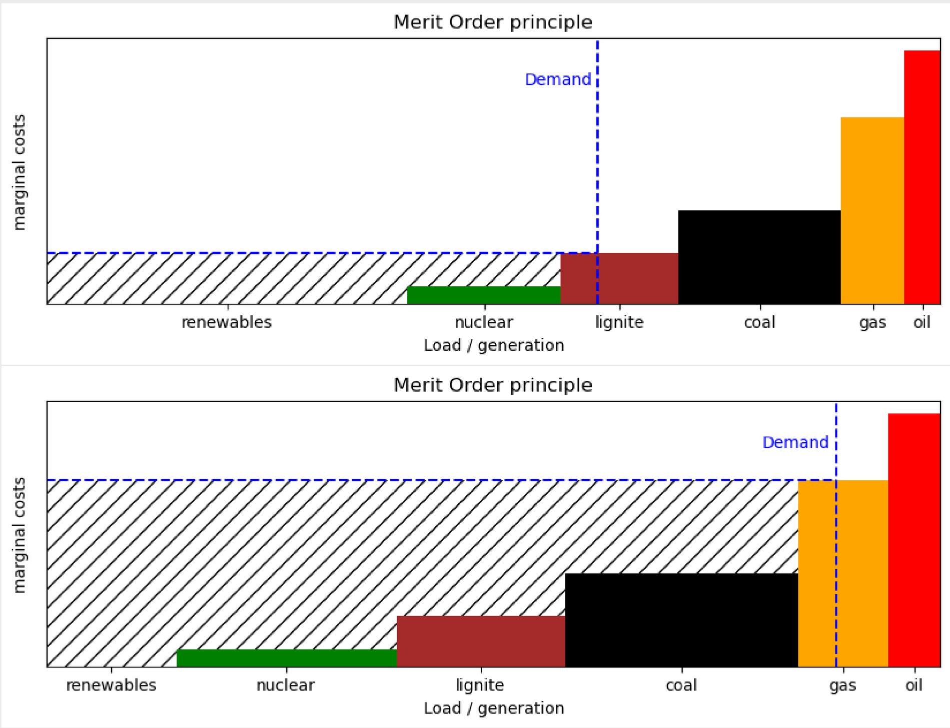
JÜLICH Forschungszentrum

# Classification and Prediction of extreme events in complex time series

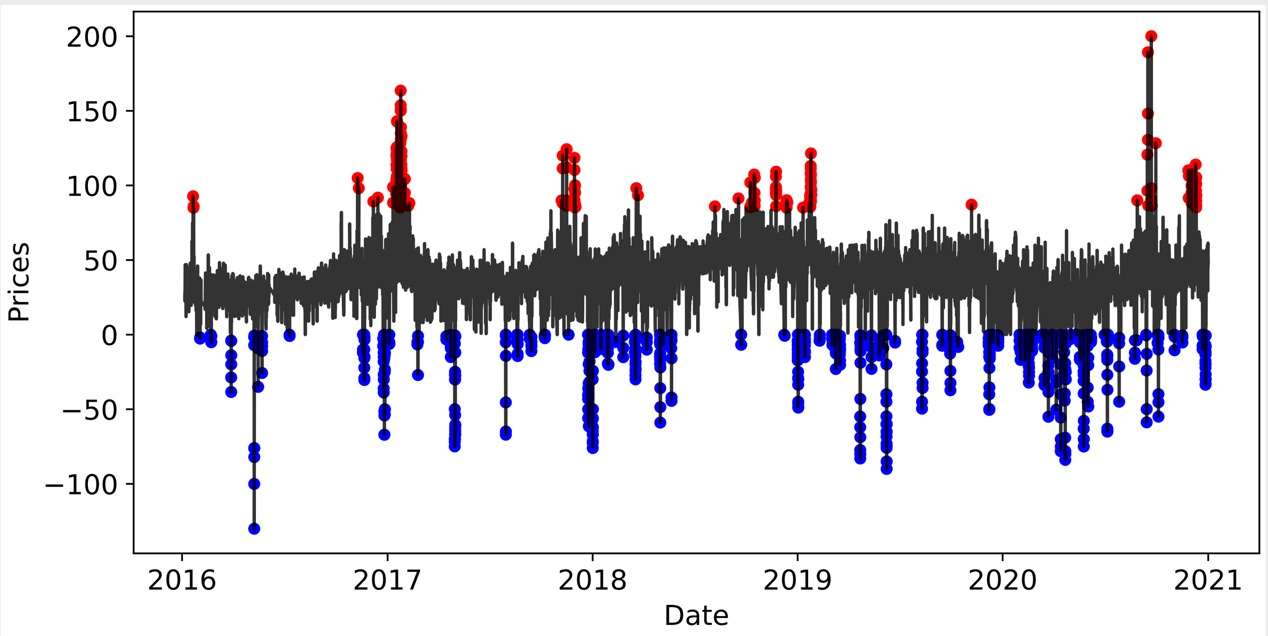
Understanding price spikes and negative prices in electricity markets

## INTRODUCTION

- A stable operation of the electric power system requires a balance of supply and demand at all times
- Markets on various timescales before delivery ensure that supply and demand are always met
- Cost efficiency leads roughly to Merit order effect




- Still abrupt price spikes occur
- Negative price spikes mean, that producers pay for power consumption




**Input data:**

- Only data known at the time of trading was used (causal relationships)
- Hourly data from January 2016 - January 2021
- Renewable generation & load forecasts & repsective raps
- Fuel prices & nuclear availability
- Residual loads of neighboring bidding zones
- Day-ahead-prices for the Germany-Luxembourg biddingzone

## METHODS

**Over-/Under-sampling**

- Outliers are very sparse (0.5% negative, 1.9% high)
- Over-sampling helps balance minority class by duplicating rare instances
- Under-sampling helps reduce majority class
- Ensures the model doesn't get biased towards the majority class
- Helps improve model performance metrics for rare events
- Results of the this technique:
  $\rightarrow$ The best performance was obtained by an oversamplingrate of 1.6 (negative prices) / 2.0 (high prices) or undersampling rates of 0.6 (negative prices) / 0.45 (high prices)

**Feature reduction**

- Many features make interpretability hard
- Correlations between features decrease model performance
- Recursive feature elimination to reduce redundant features
- Residual loads of neighbours were aggregated


**Contact:** j.lange@fz-juelich.de


## MODELS: REGRESSION vs. CLASSIFICATION

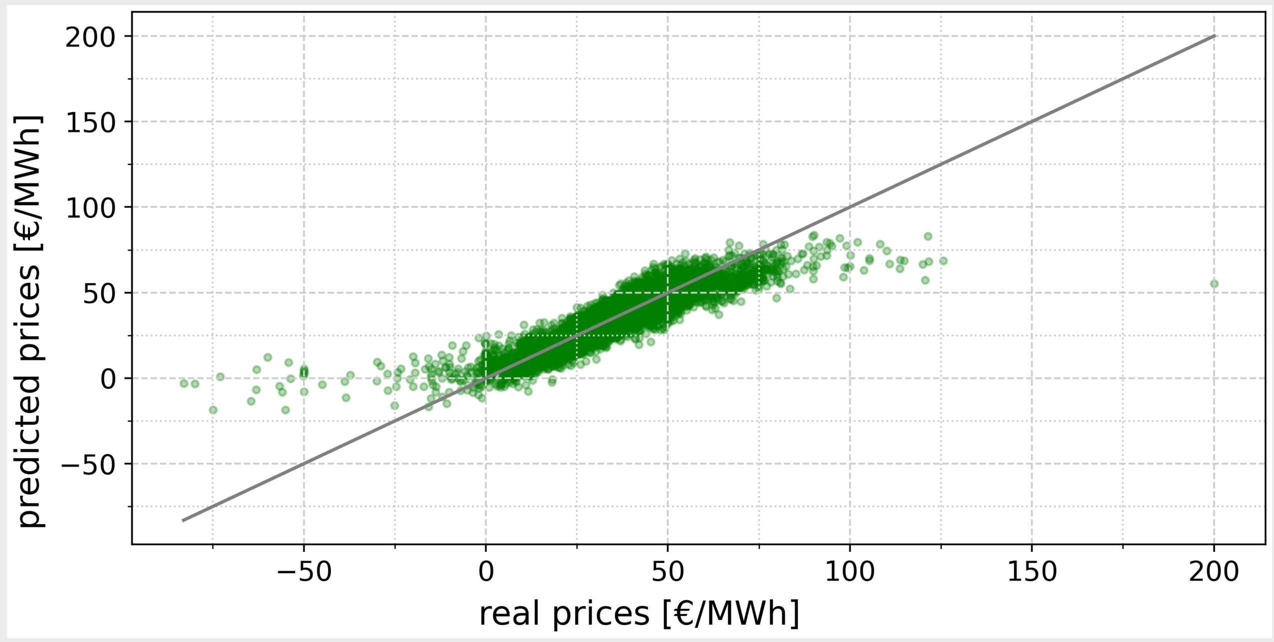
- General idea: predict day-ahead price $y$ for each hour from external features $\{x_1, x_2, \dots\}$ (load, renewable generation forecasts, oil prices,...)
- Simple linear models were used because of inherent interpretability

**Linear Regression**

- Fit

$$\hat{y} = \boldsymbol{\beta}^T \cdot \mathbf{x} = \beta_0 + \sum_{i=1}^{n-1} \beta_i \cdot x_i$$

  to minimize squared error
- Linear regression models fail to predict outliers




**Logistic Regression**

- Choose $\beta$ to maximize likelihood of an outlier event ($y = 1$)

$$p(y = 1 | \mathbf{x}, \boldsymbol{\beta}) = \frac{1}{1 + \exp(\beta_0 + \sum_{i=1}^{n-1} \beta_i \cdot x_i)}$$

- Two models: One for high ($> 85$ €/MWh) and one for negative prices

## RESULTS & INTERPRETATION

- Absolute value of coefficients $\beta_i$ is a metric for importance of the feature
- Oversamplig, Undersampling, and weighted model gave similar results
  $\rightarrow$ Consistent models
- Load & renewable generation show high importance for both high & low prices with inverse effects
  $\rightarrow$ Consistent with Merit order
- Residual load of neighbors $res_{agg}$ is very important for high prices, unimportant for negative prices
  $\rightarrow$ High prices in surrounding countries make export unlikely
- $CO_2$ price dependency high in both models
  $\rightarrow$ Proxy for increase in price over time?
- Negative price prediction led to much better results than high price prediction