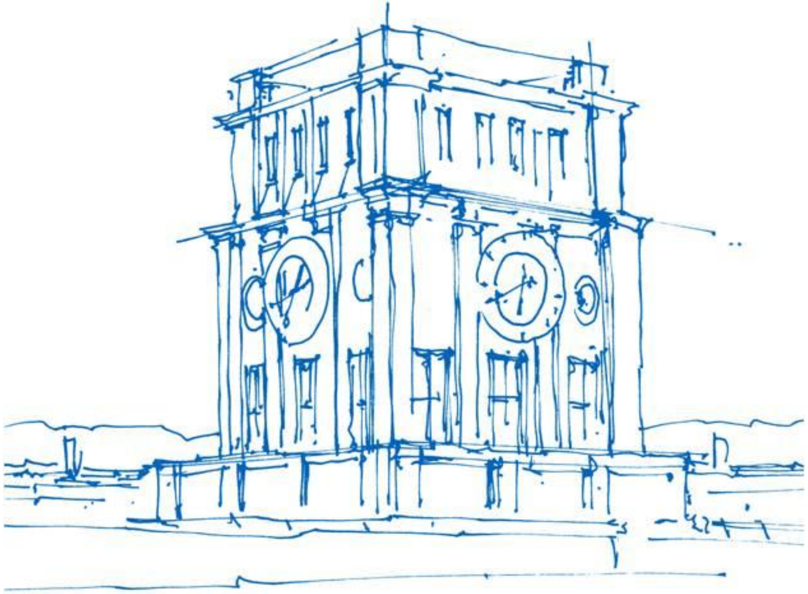
TUM Data Innovation Lab

Energy Peak Load Prediction in a BMW Plant

M. Herrmann, P. Labintceva, D. Simancas Mateus,

F. Vogeltanz, B. Yazganoglu



Uhrenturm der TUM

Technical University Munich

Department of Mathematics

Munich, February 17th 2018

Agenda

1. Introduction to the project

2. Dingolfing Production Plant

3. Energy Data

4. Production Data

5. Logistics Related Data

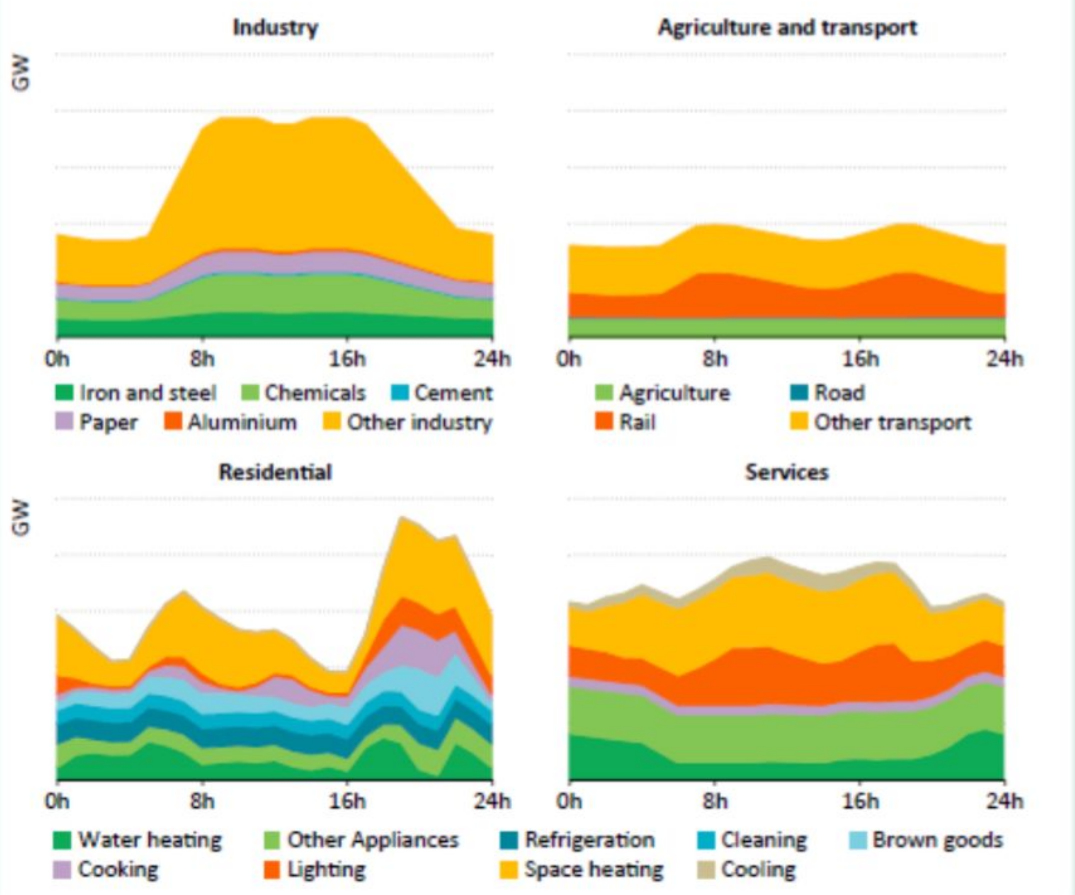
6. Conclusion

Agenda

Agenda

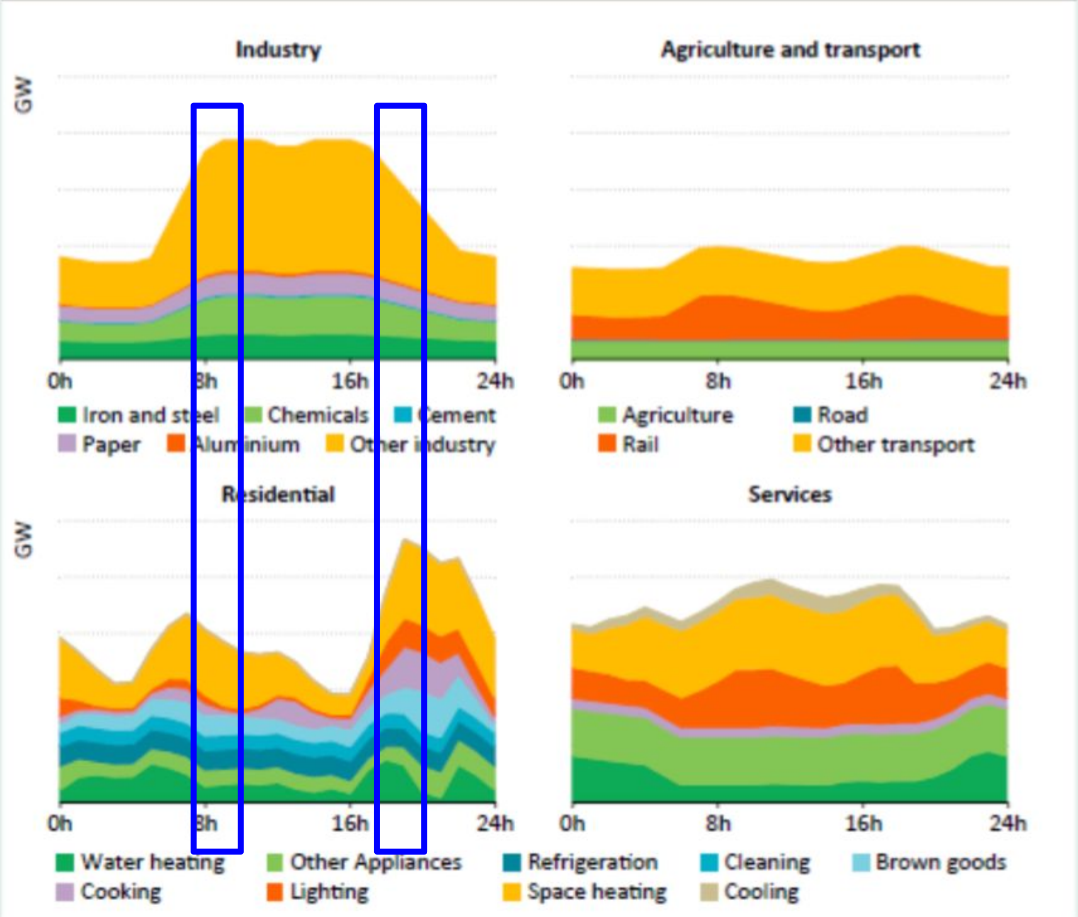
1. Introduction to the project
2. Dingolfing Production Plant
3. Energy Data
4. Production Data
5. Logistics Related Data
6. Conclusion

Figure 12.10 ▷ Illustrative load curves by sector for a weekday in February in the European Union compared with the observed load curve by ENTSO-E in 2014



Source: International Energy Agency: World Energy Outlook 2016

Figure 12.10 ▷ **Illustrative load curves by sector for a weekday in February in the European Union compared with the observed load curve by ENTSO-E in 2014**



HPLW for BMW;

January 2017[1)]:

07:30 a.m. - 09:00 a.m. and

05:00 p.m. - 19:30 p.m.

[1)]Hochlastzeitfenster für 2017 nach § 19 Abs. 2 Satz 1 StromNEV

Source: International Energy Agency: World Energy Outlook 2016

Motivation: High Peak Load Windows

To be able to profit from cost reduction, several conditions have to be fulfilled.
For BMW[1)] one is:

highest electricity demand of the year **within a HPLW** < 0.9 x (highest electricity demand of the year **outside the HPLW**.)

[1)] Hochlastzeitfenster für 2016 nach § 19 Abs. 2 Satz 1 StromNEV

Milestones

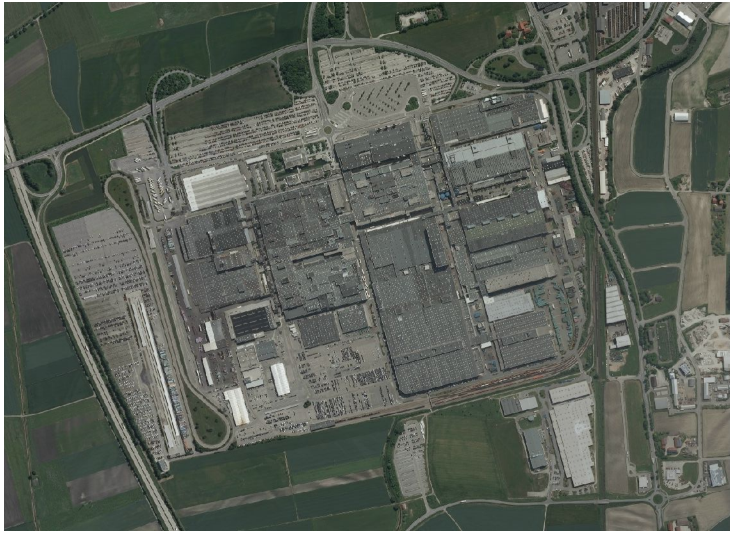
1. Statistical exploration of energy data and outlier correction
2. Detection of relevant features to explain and forecast the energy profile
3. Using machine learning methods for data exploration and developing prospective improvements
4. Documentation of project

PLANT DINGOLFING

press.bmwgroup.com

Dingolfing Production Plant

- BMW Group’s largest vehicle production site in Europe
- 1,600 BMW vehicles every day.
- models of the 3 to 7 Series
- components for BMW’s electric vehicles
- car bodies for Rolls-Royce Motor Cars
- the operation is spread over **different production halls**



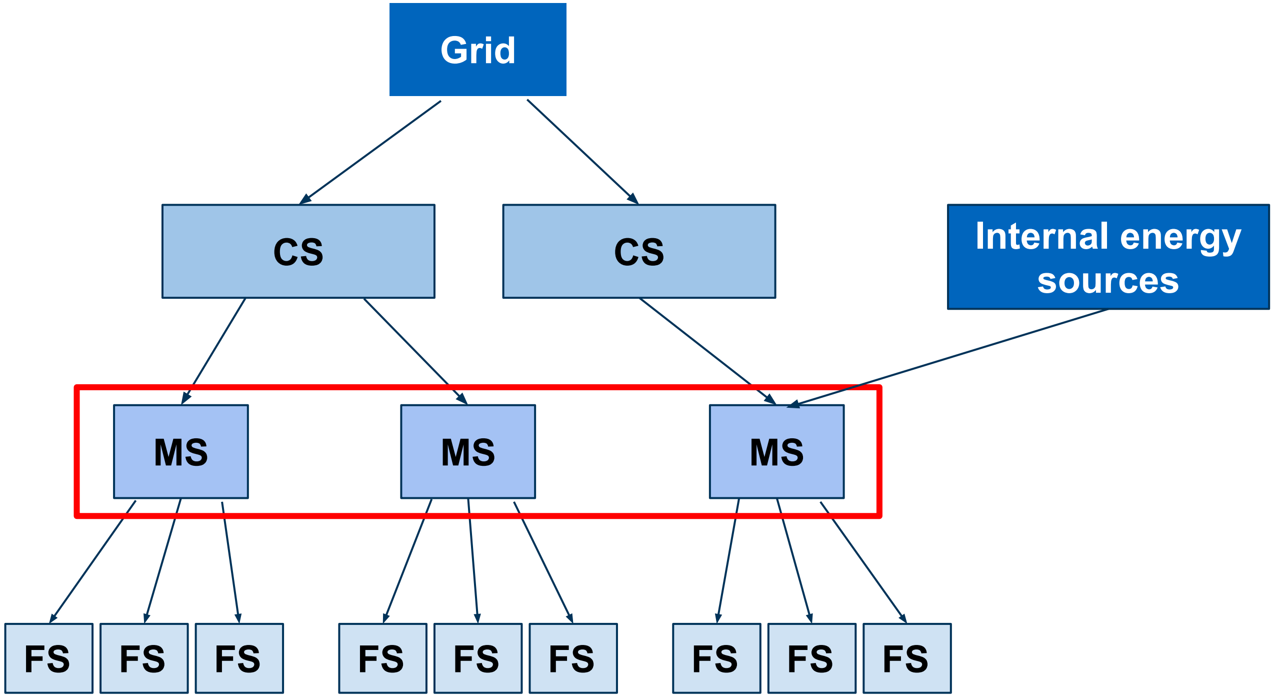
https://www.bmwgroup-werke.com/dingolfing/en.html

press.bmwgroup.com

press.bmwgroup.com

ENERGY DATA

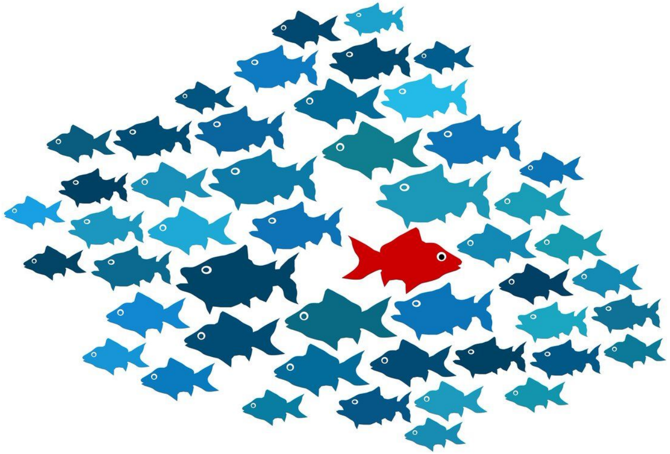
Energy Flow


Grid
CS
CS
Internal energy sources
MS
MS
MS
FS
FS
FS
FS
FS
FS
FS
FS
FS

Energy Data: Exploratory Data Analysis

The energy data is a large time series data set consisting of:

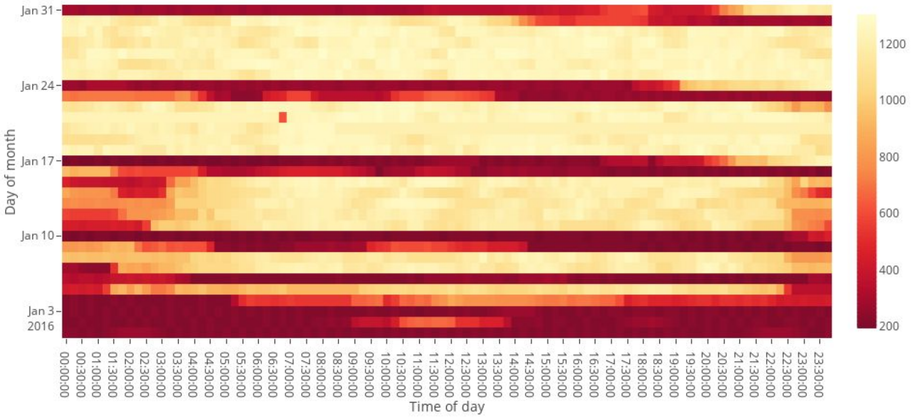
- energy data from the whole plant
- incomplete: sum of elements is not equal to the total
- aggregated by 15 minute consumption values
- no production data
- mainly coming from measuring devices which can be subject to recording errors
- there are common identifiable measurement errors



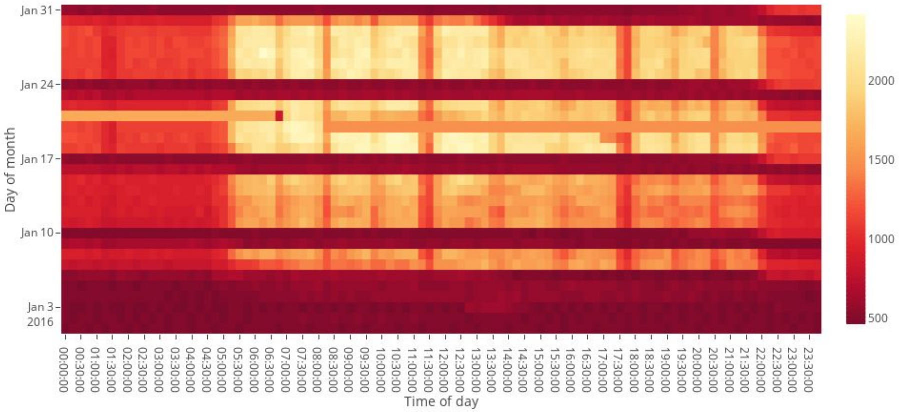
Energy Data: Visualization

Comparison between energy profiles at different stations of the plant, featuring **shifts**, **weekends** and occasional **energy peaks**.

MS-13



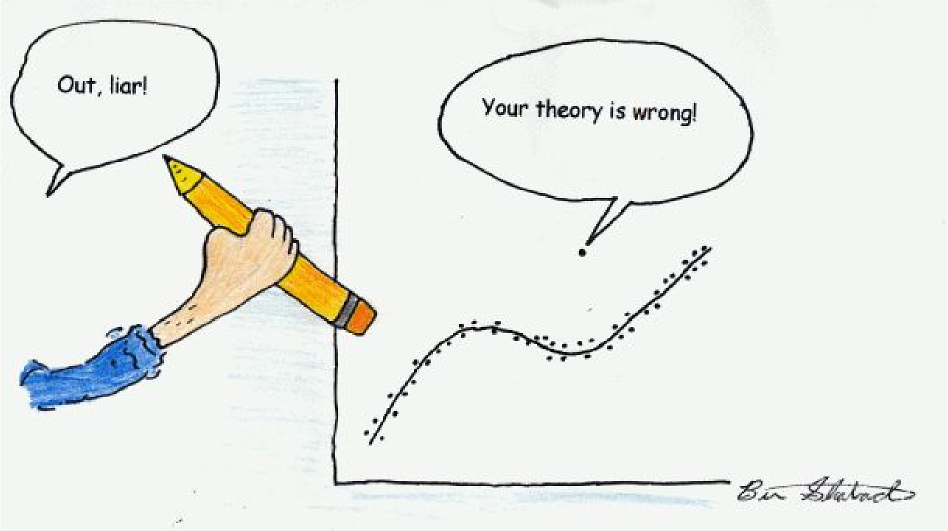
MS-4



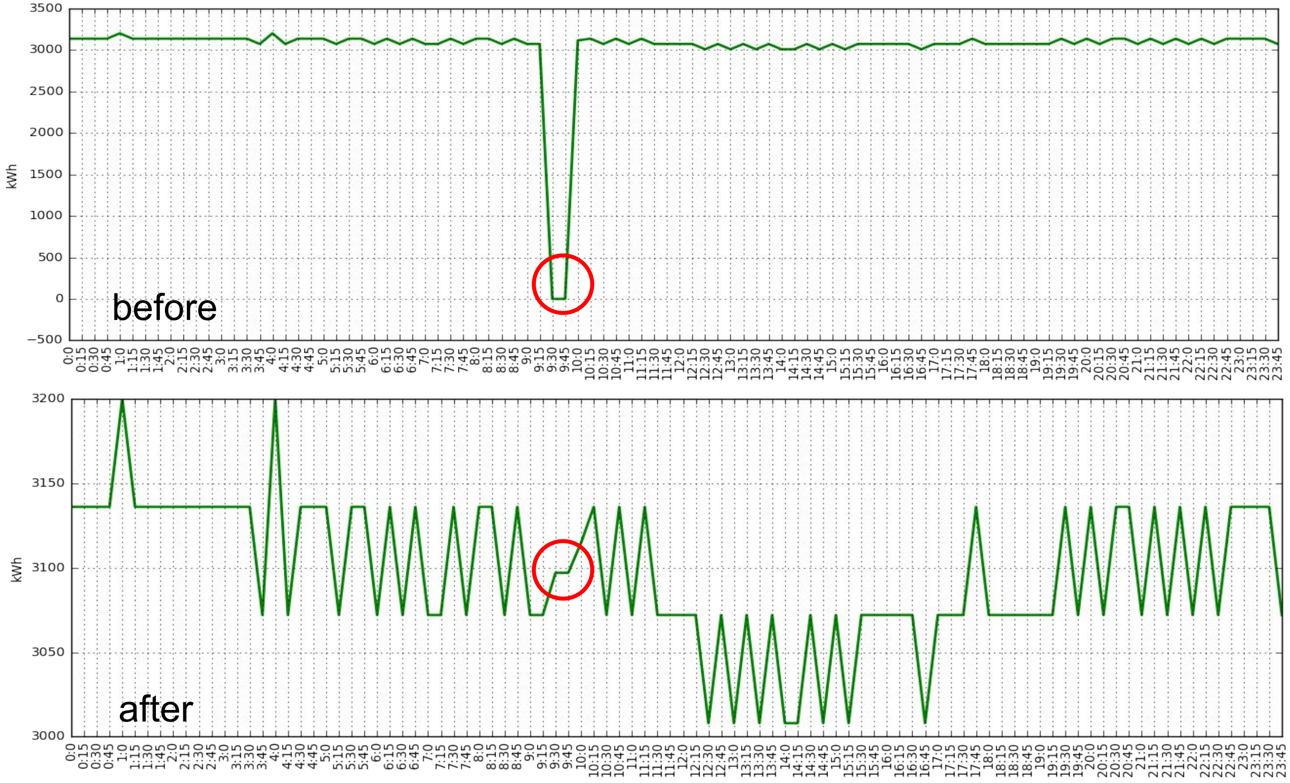
Energy Data: Outlier Detection and Missed Data

Types of wrong measurements in the energy data:

1. Missed data
   1.1. missed values
   1.2. “fake” zero consumption
2. Too long/too short measurements
3. “Wrong” outliers



Energy Data: Missed Data

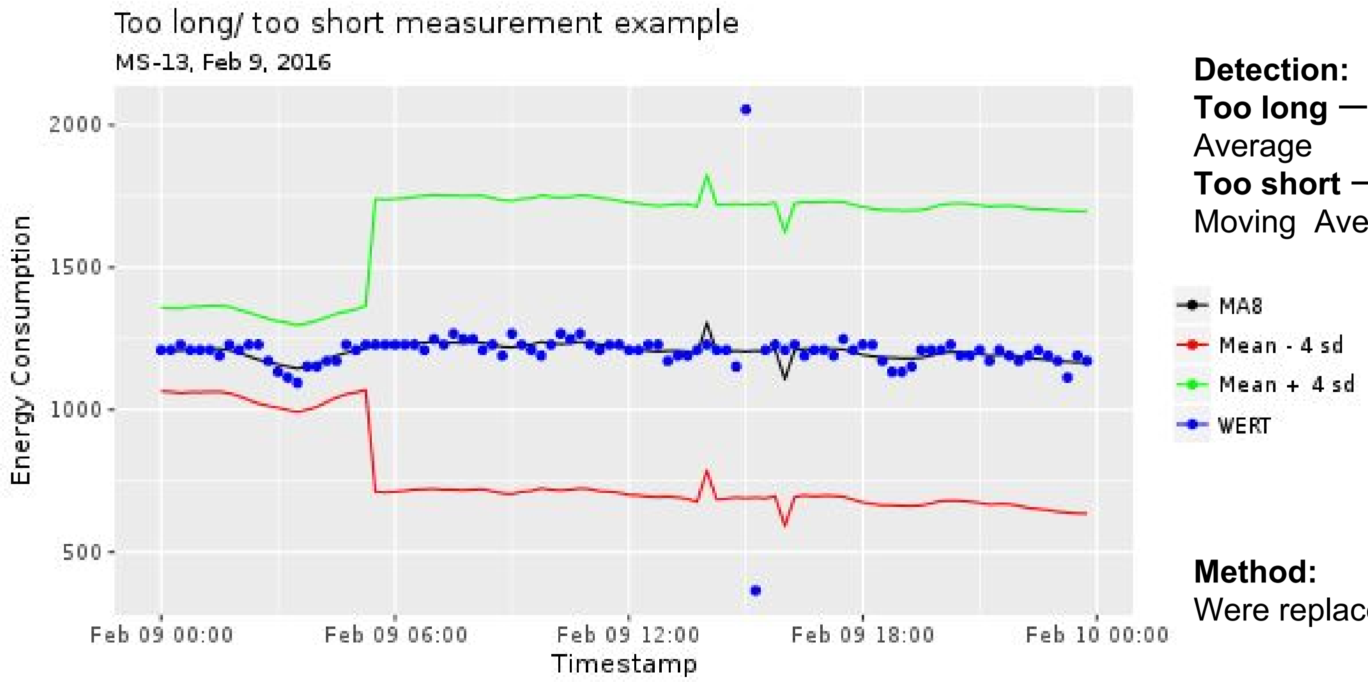


Method:

Several consecutive zeros or -1 during the production time were replaced with the **mean of the day**.

Single zeros or -1 during the production time were replaced with the **mean of neighbouring values**.

Energy Data: Too long/too short Measurements



Detection:
Too long — 3 std above the Moving Average
Too short — 3 std. below the Moving Average

Method:
Were replaced with their **average**

Energy Data: Wrong Outliers

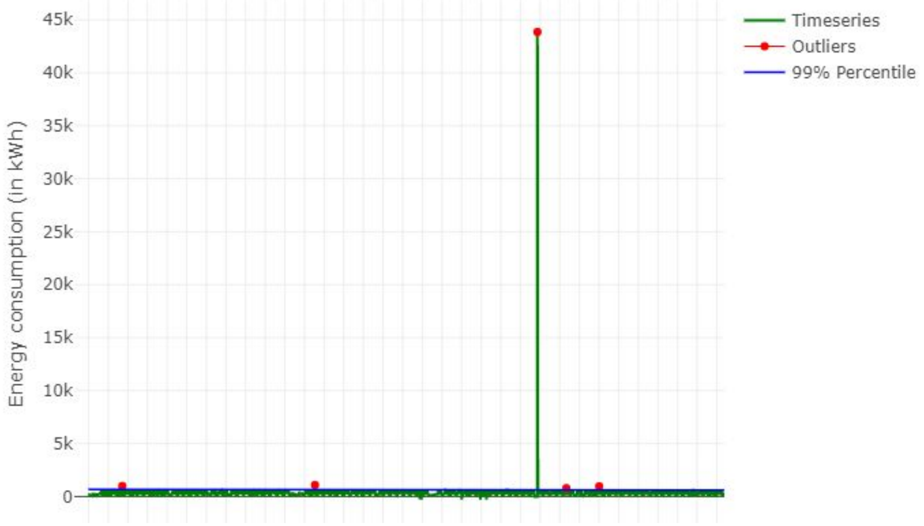
Detection:

Running Standard Deviation

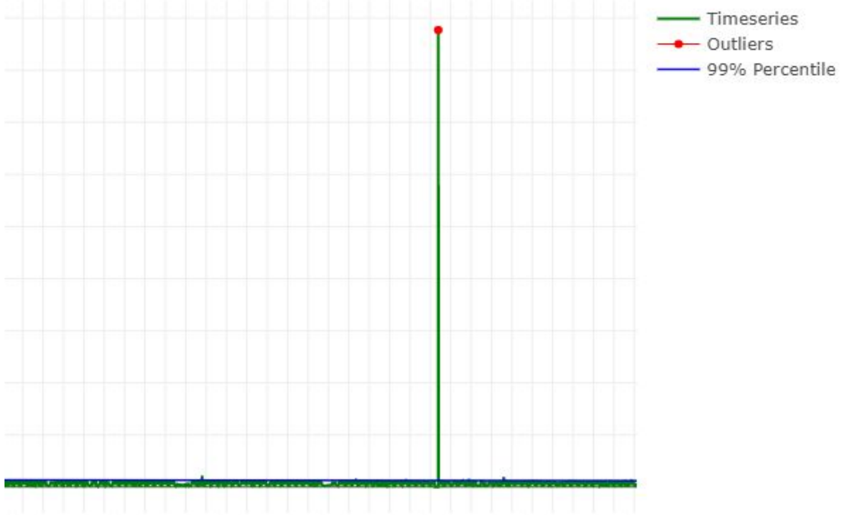
Lower bound: 99th Percentile

Method:

Were replaced with mean of neighbouring values



01.2016 10.2017



01.2016 10.2017

PRODUCTION DATA

press.bmwgroup.com

Production Data: Exploratory Data Analysis

The data set contains information about **production** together with **energy.** It mainly is:

- data from **car body assembly** only
- nodes are **processing units**
- cars come in and come out, energy is measured
- timestamps are *very* precise, to milliseconds
- **incompatible** with energy data

Production Data: Preprocessing

Preprocessing of production data consisted of:

- **Robust outlier detection**
  - Median Average Deviation (**MAD**)
  - Energy distributions are determined **by node**
- Accumulating in 15 minute time frames
  - Make it compatible with energy data
  - Shorter table, minimum information loss

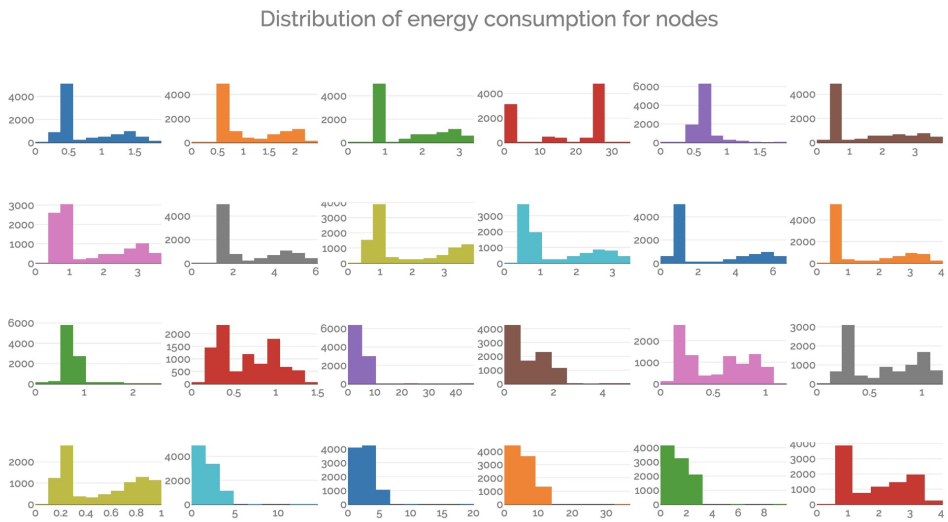
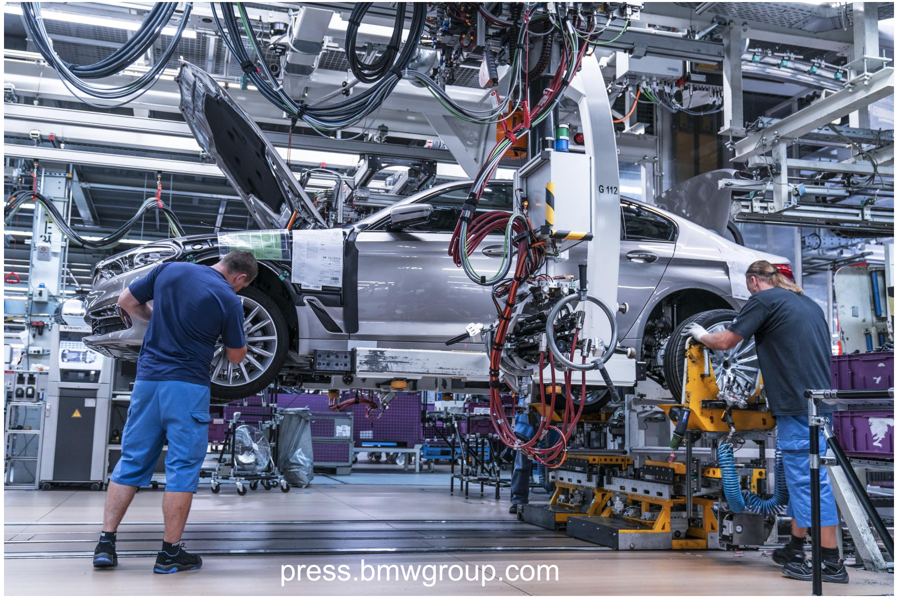


Fig #. Data from each node can be interpreted as a sample from a different populations.

Production Data: Linear Regression Model

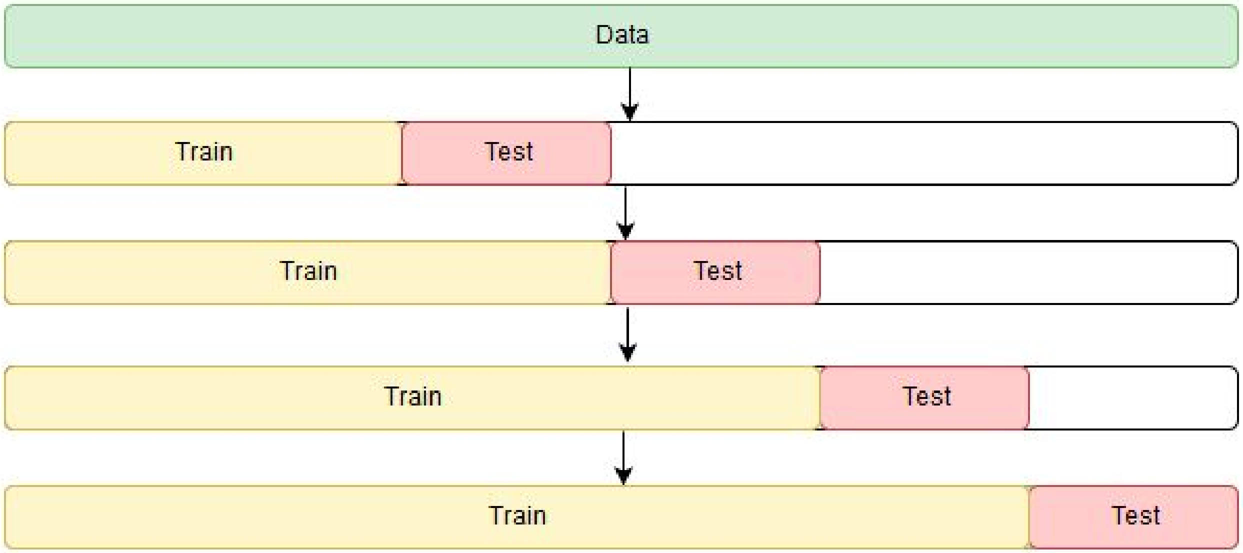
Used features:

- number of car passed through the node
- lags of the target variable (96)
- day of week coded by means of the day/hour
- weekend or not
- information about shifts and ramp-ups
- car model



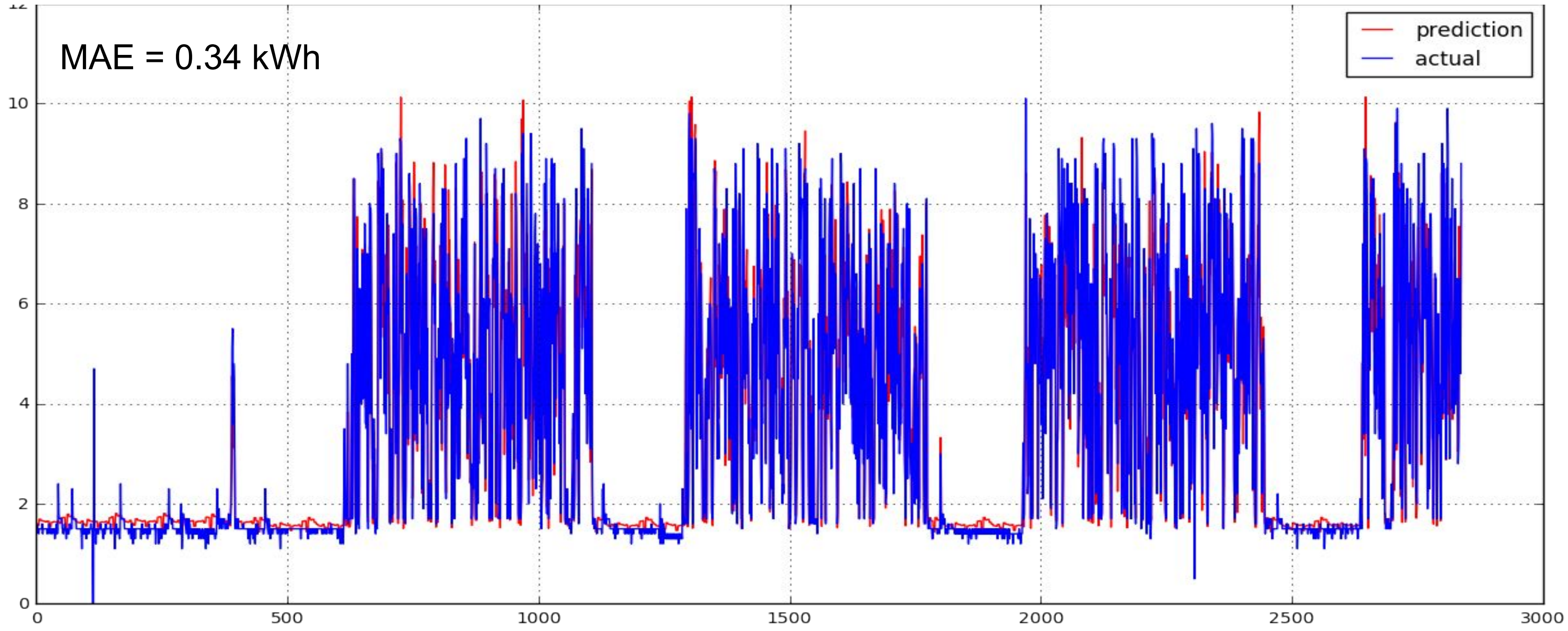
Production Data: Linear Regression Model

Time series Cross-Validation



Production Data: Linear Regression Model

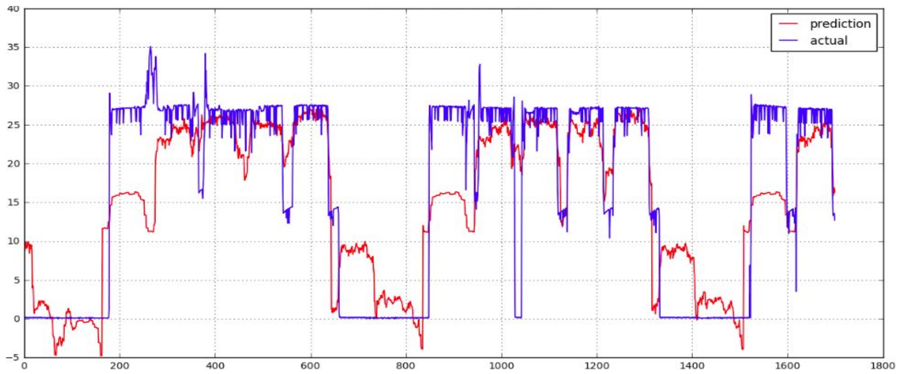
83 nodes have MAE < 1 kWh!



Production Data: Linear Regression Model

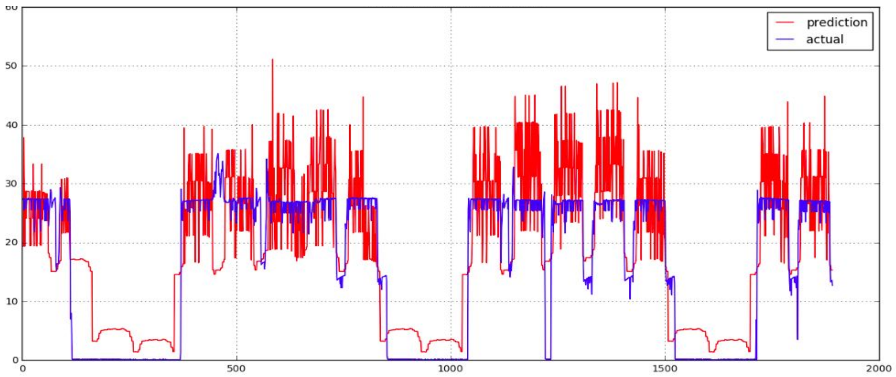
MAE ≈ 5.0 kWh

without number of processed cars



MAE ≈ 6.0 kWh

with number of processed cars



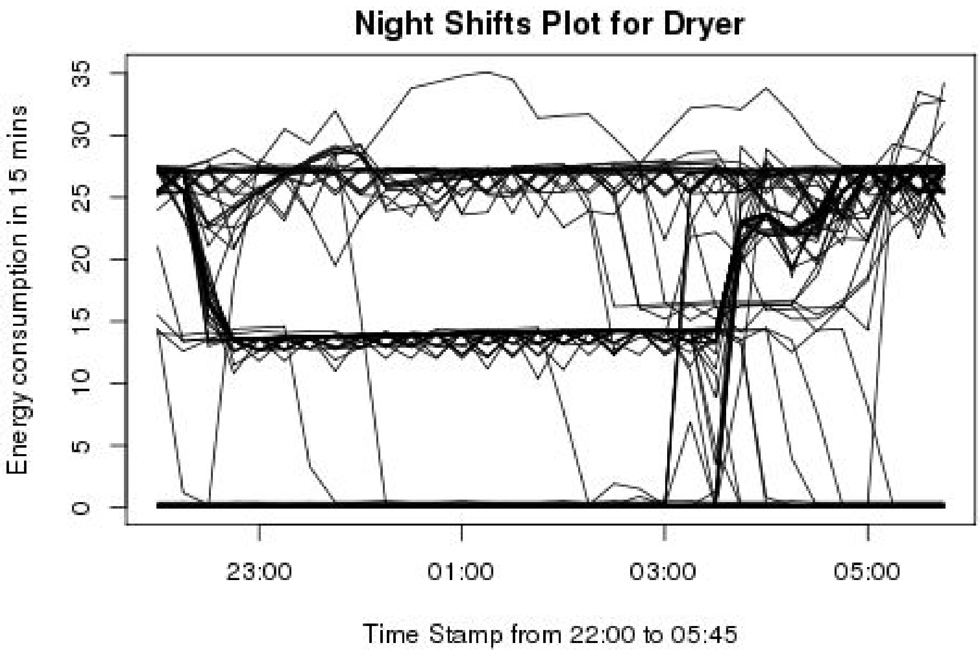
The Other 6 Nodes

- Understand what is happening in these nodes
- Energy consumption independent of units
- Nights

https://www.flickr.com/photos/pasma/485201047

The Dryer in the Body Shop

- Highest energy consumption
- 98 days, 32 timestamps
- To detect patterns
- To gain information within a cluster

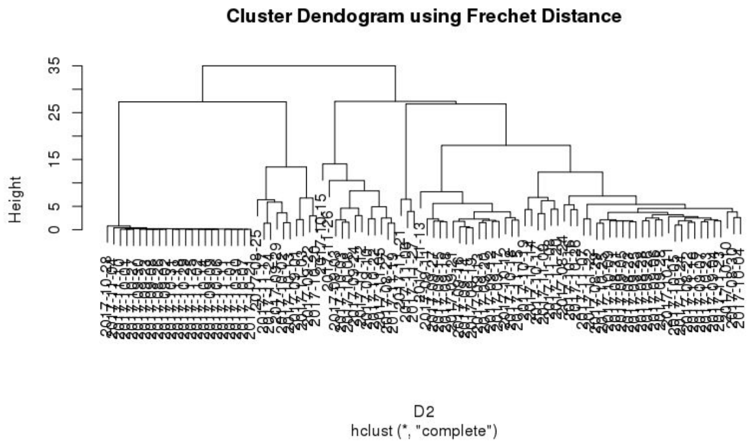


Time Series Clustering

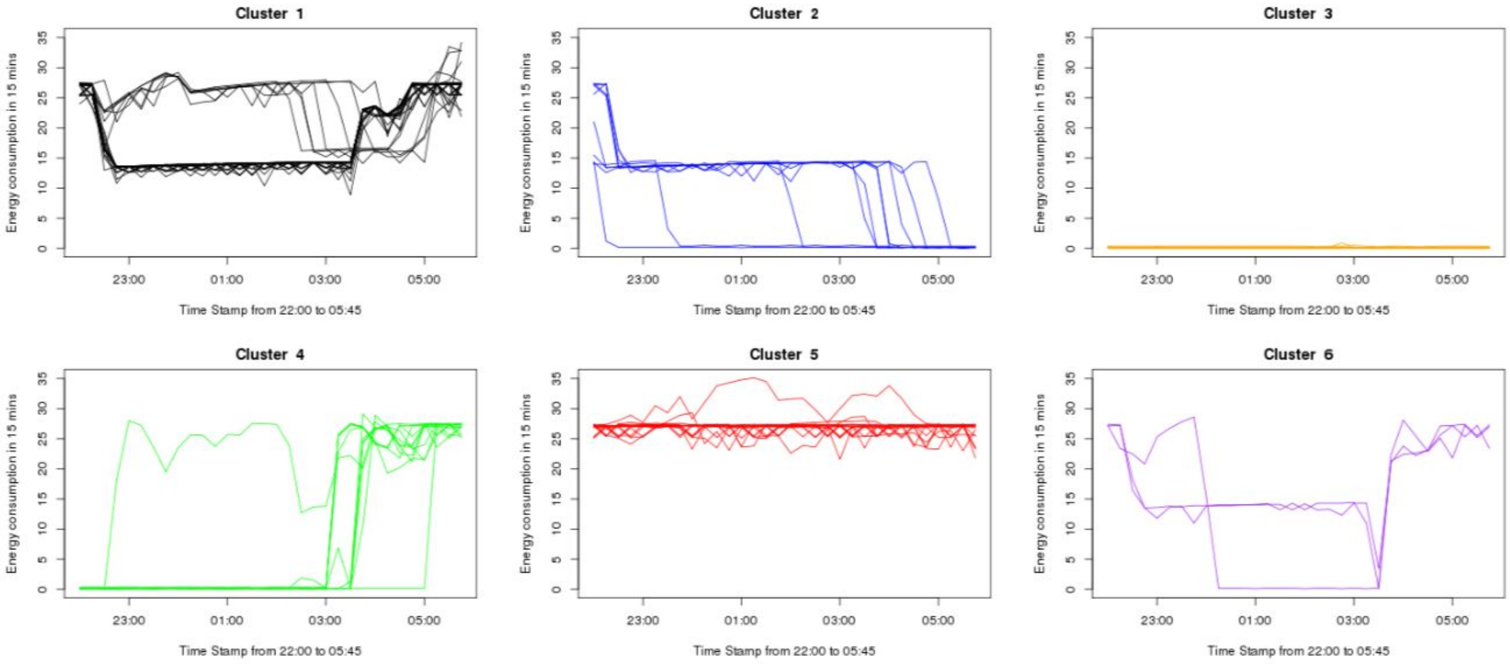
- 32 dimensional points in space vs. a curve with 32 observations
- How does Time Series Clustering work?
  1. Compute a dissimilarity matrix
  2. Sort out the observations
  3. Hierarchical Clustering

Dissimilarity Methods

1. Correlation: Based on the estimated Pearson's correlation of two given time series
2. Frechet Distance: The infimum of maximum distances between two curves
3. Dynamic Time Warping: Optimal match between time series regardless of their acceleration



Clusters



Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Energy consumption in 15 mins
0
5
10
15
20
25
30
35
23:00
01:00
03:00
05:00
Time Stamp from 22:00 to 05:45

LOGISTICS RELATED DATA

press.bmwgroup.com

Logistics Related Data: Basic Statistics.

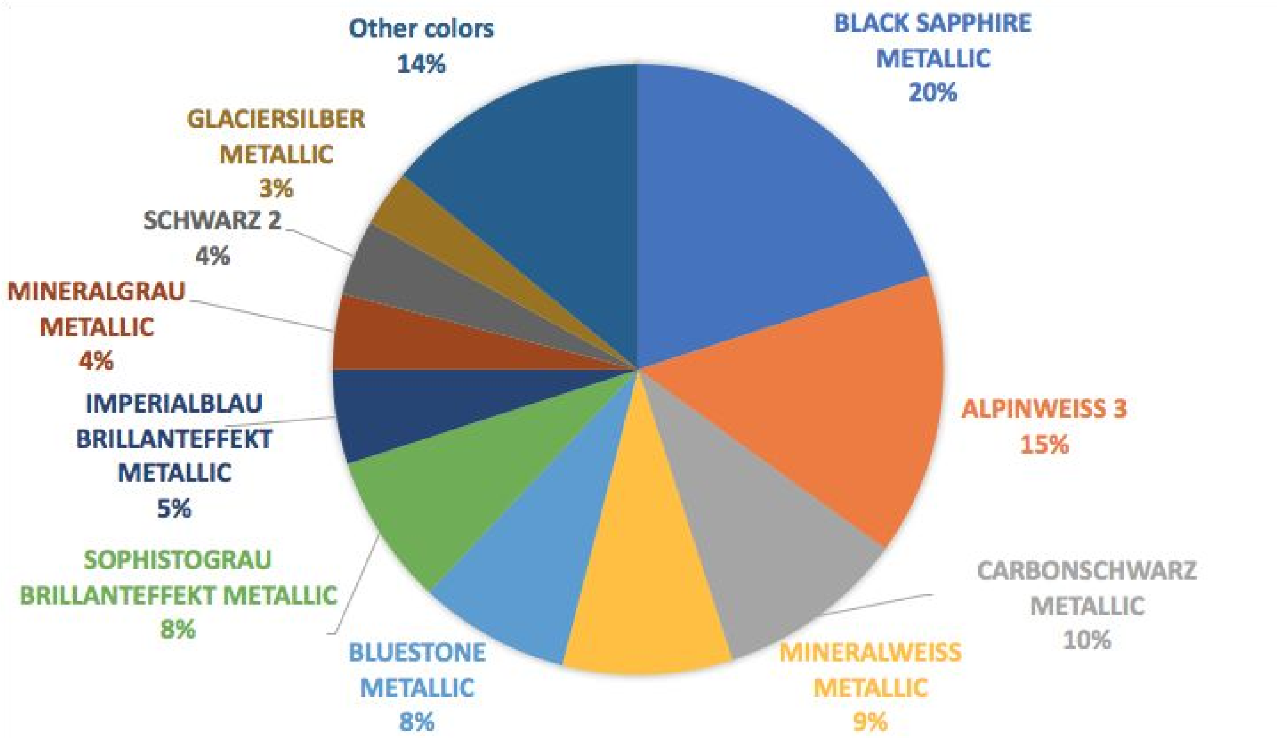
Logistic system data: Tracking unique cars at different production stations in the **paint shop.**

Measurements	Unique Cars	Colours
~ 370,000	~ 50,000	~ 100

Body specification (Variante)	Production stations (Zaehlpunkt)	Time Period
~ 1,000	36	~ 6 Month

Not complete production data; only extract

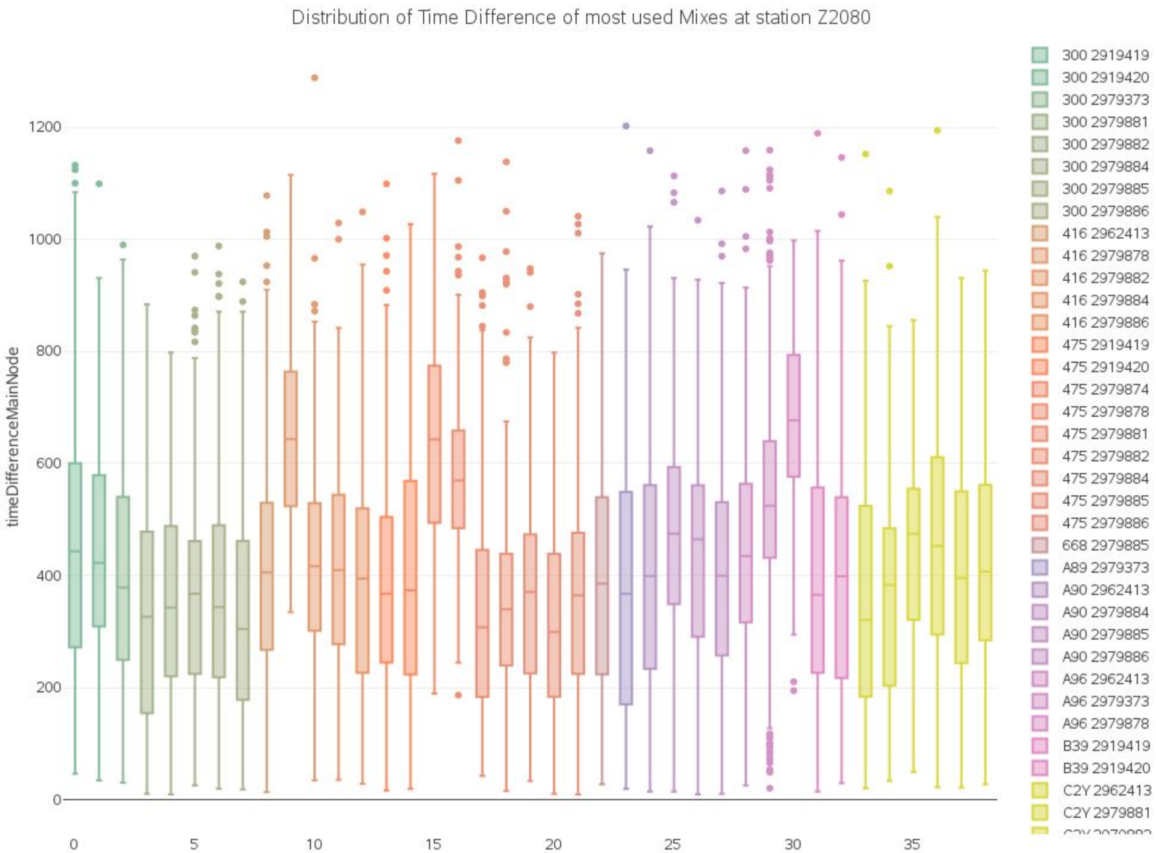
Logistics Related Data: 1. Basic Statistics.



For 85% of the cars just 10 different colours are used.

For 2% of the cars we don’t know which colour is used.

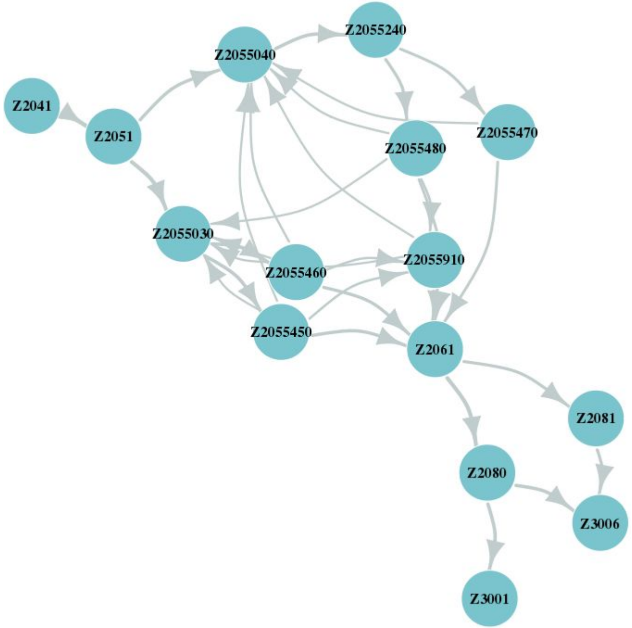
Logistics Related Data: 2. Boxplot.

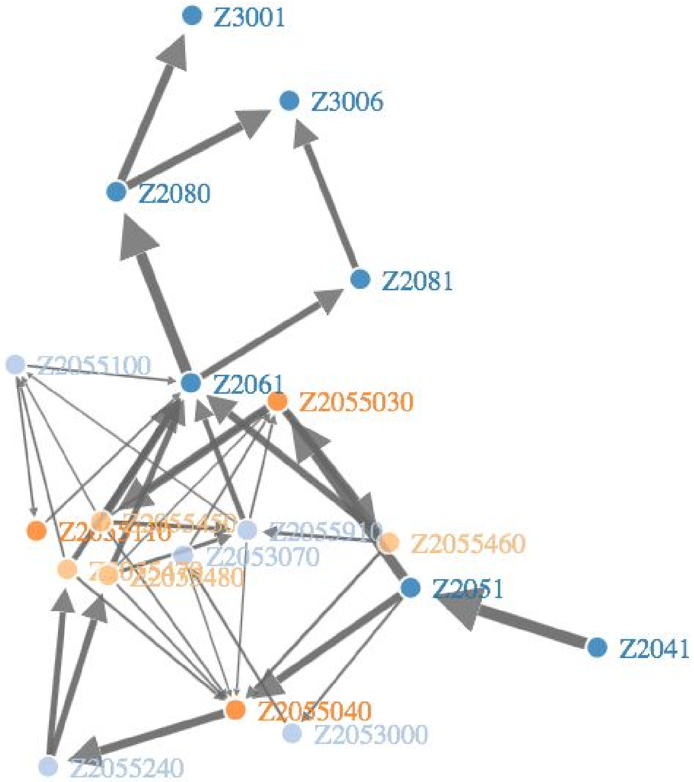


Mix =
Combination of color and variant

Logistics Related Data: Graphs.

Production flow of cars with color Atlantic Cedar Metallic.





Data without December 2017.

press.bmwgroup.com

Logistics Related Data: Regression Analysis

- Motivation:
  - ➢ Detection of correlations between features (color, car type) and energy consumption
  - ➢ Quick and flexible implementation

- Benchmark model: Linear Regression
  - ➢ No perfect linear correlation expected
  - ➢ But even other correlations (logarithmic, exponential, etc.) would be visible in slope

- Results:
  - ➢ Highly significant: Average Energy Consumption of previous day, Number of Working Steps
  - ➢ Significant: Color, Car Type, Number of Color/Car Type changes, Temperature, Wind Speed
  - ➢ Unclear: Radiation

Regression Analysis

- Prediction

Model	RMSD (in kWh)	Prediction Error
$EnergyUse = \beta_0 + \beta_1 Col_1 + \ldots + \beta_{14} Col_{14} + + \beta_{15} Car_1 + \ldots + \beta_{31} Car_{17} + \beta_{32} WS + \beta_{33} ColC + \beta_{34} CarC$	112.3	7.8%
$EnergyUse = \beta_0 + \beta_1 \; AverageEnergy_{previousDay} + \beta_2 WS$	92.5	6.5%

- WS: Number of working steps per interval
- ColC: Number of color changes per interval
- CarC: Number of car changes per interval

Color Clustering

- Motivation
  - ➢ Statistical significance of number of color changes
  - ➢ Goal: Find Clusters of frequent color combinations

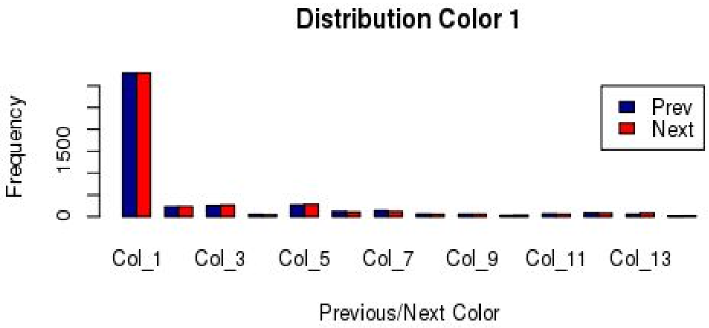
- Method
  - ➢ K-Means: Popular approach, only for numerical values
  - ➢ K-Mode: K-Means extended by categorical values

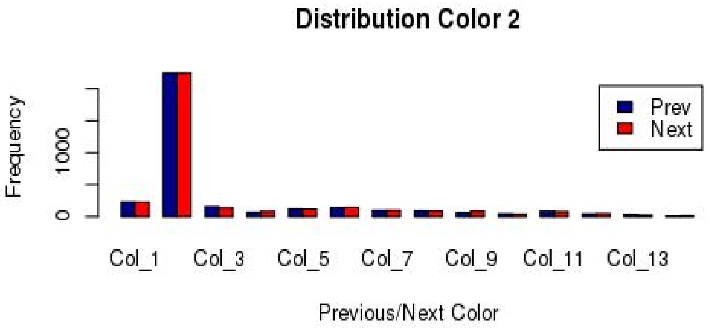
    → each color forms own cluster

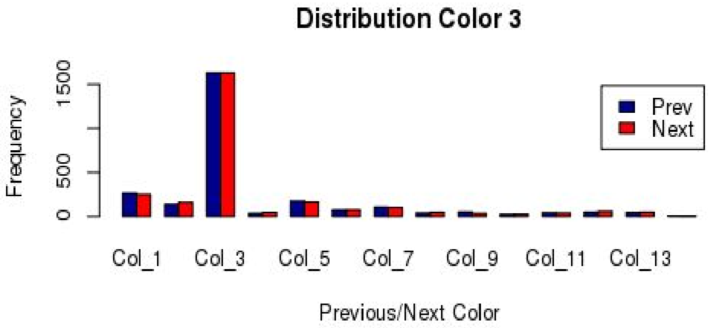
- Reason
  - ➢ Colors painted in sequence

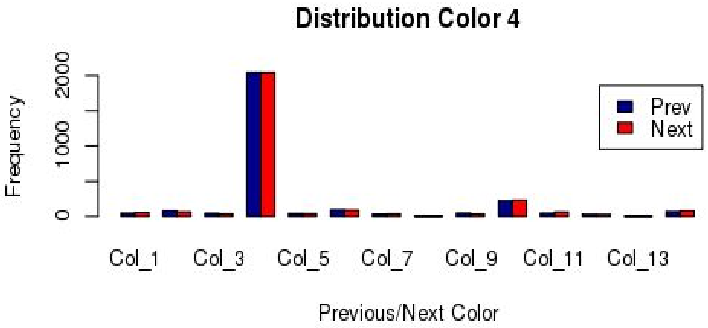
    → Painting process already optimized with regards to color

Color Clustering









Conclusion

Results

- Feature detection & electricity demand prediction
- Analyzing energy saving potential: ramp-ups
- Painting sequence: small room for improvement
- Visualization of production flow

Outlook

- Include data from the second MS connected to the paint shop into the analysis
- Experiment with neural network architectures
- Compare paint shop in Dingolfing with paint shops elsewhere (e.g. Regensburg, Leipzig)
- Consider other forms of energy consumption (water, heating, cooling)

Learning Outcomes

- Energy market and management at a company and in Germany
- Understand the data > Preprocess the data > Build models around the data > Interpret results
- Repurposing data can pose a challenge
- Balance coordination and manpower
- The value of diverse teams

https://www.flickr.com/photos/68751915@N05/6869764745

Thank you!

Sources:

Project related materials from 2016/2017

- Energy Data
- Logistics Related Data
- Production Data
- Energy Data Overview
- Hochlastzeitfenster für 2016/2017 nach
  § 19 Abs. 2 Satz 1 StromNEV
- Overview measurement points
- https://www.bmwgroup-werke.com/dingolfing/en.html

For this presentation:

- press.bmwgroup.com
- International Energy Agency: World Energy Outlook 2016
- www.ag-energiebilanzen.de 2016 preliminary figures
- flickr.com