FORECASTING @ DPDHL

An introduction to forecasting projects at DPDHL Group

Data Science Meetup
Bonn, February 1st 2018

Agenda

1 Forecasting Use Cases @ DPDHL

2 Forecast Model

- ❑ Workflow for Developing a Forecasting Model
- ❑ Time-series vs Machine Learning approaches
- ❑ Particularities about Forecasting Projects

3 Use Case Deep Dives

4 New technologies for FC

Agenda

1 **Forecasting Use Cases @ DPDHL**

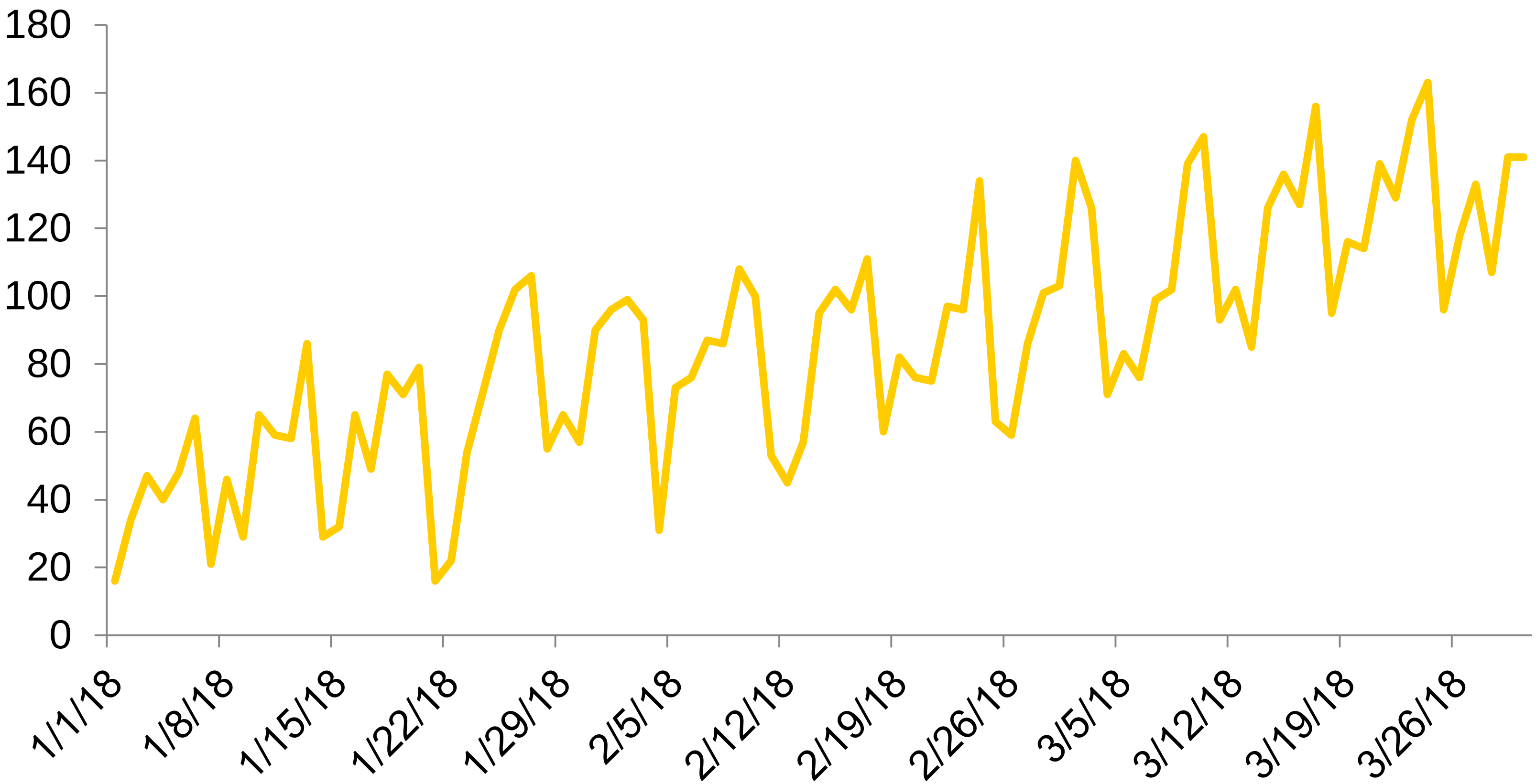
2 **Forecast Model**

- ❑ Workflow for Developing a Forecasting Model
- ❑ Time-series vs Machine Learning approaches
- ❑ Particularities about Forecasting Projects

3 **Use Case Deep Dives**

4 **New technologies for FC**

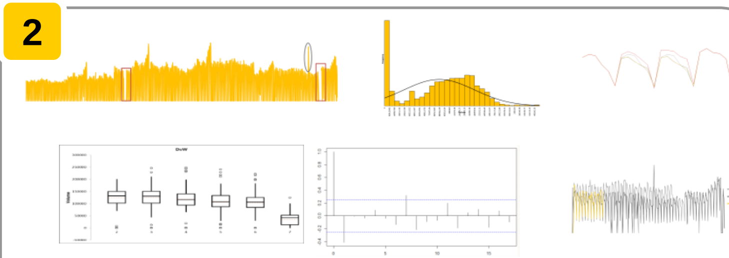
What is a time series?



Workflow for Developing a Forecasting Model

1 Data Sources
- Historical data
- Holidays
- Business Specific data

2 Plot!!!



3 Pre-process
- Divide data: train/ validation/test
- Clean for missing values
- Check for special events
- Understand outliers
- Analyze pattern changes
- Generate features (and more features)
- Select the best features

4 Train your model
- Test as many methods as possible
- Select best method (for your application)

5 Get your results
- Forecast
- Post-process
- Display results

Time-series vs Machine Learning approaches

	Characteristics	Advantages	Disadvantages
Time-series	• No feature generation necessary • Predictions dependent mainly on trend and seasonality	• Traditional method, usually known by the business • Can easily incorporate a prediction interval • Already implemented in out-of-the-box solutions	• Data set needs to be complete • More difficult to add external effects to model • At least the traditional, simple methods assume linearity in the data
Machine Learning	• Feature based models • Predictions dependent on information encoded in feature vector • Trend and seasonality have to be modeled into features using historical information	• Can handle incomplete data easier than TS approaches • Can model external effects by encoding them into features • Some methods can easily model non-linearity in the data	• Not all methods can easily incorporate prediction intervals • Not always easy to explain to the business • Requires a lot of business understanding and effort for feature engineering

Time series

Machine Learning

Method examples

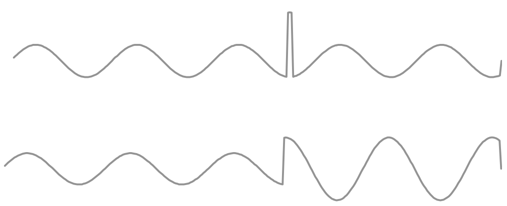
Method	Description	When to use
(Seasonal) Moving Average	Simplest forecasting technique. Forecast is the average of last n values	To get simple trends or to smooth irregularities in the data
Exponential Smoothing (and variations)	Forecast value is a weighted average of the past observations with weights reducing exponentially over the time	Commonly used to smooth the data by removing much of the noise
ARIMA (and variations)	Forecast value is based on autocorrelation and moving average patterns	Can be useful when the data has a long seasonal history with minimum outliers
Linear Regression (and variations)	Forecast value based on linear relationship with the target	When a simple and understandable forecast is required but data has important external effects
Random Forest	Forecast value is an average of the forecast produced by several decision trees	Can be useful for forecasting with non-linear relationships between features and target, easy to apply
Gradient Boosted Trees	Forecast is a result of sequential ensemble of decision trees with successive more weight to hard to predict observations	Similar to random forest but more careful application including tuning of parameters is required
Support Vector Regression	Forecast is defined based on the determined support vectors (i.e. “most decisive” data points of the training set)	Similar to gradient boosted trees, might also be a good option when you don’t have enough data for complex methods
Neural Networks	Forecast based on a network of simple processing units who jointly create a prediction	For big data sets with complex patterns, some variants detect temporal structure of data automatically

Particularities about Forecasting Projects

Main things to consider in a forecasting project for people with regression background:

Data Cleansing

Outlier vs Changing point



Pre-process that leads to Post-processing

Removing special events effect	Adding special events effect
For training	**After Prediction**

Feature Generation

Temporal Features:

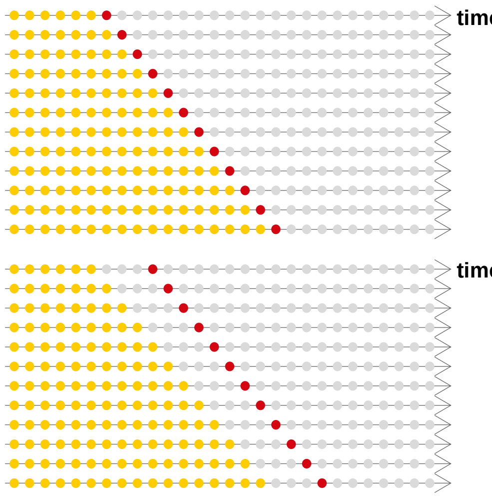
Year, Month, Week of Year, Day of Week, is holiday, distance to holidays and others

Features based on historical values:

value n lags ago, average over p lags ago, sums and differences between lags and others

Cross Validation

The cross validation also needs to take into consideration the temporal nature of the data.



Key point: Forecast projects demand attention to the temporal nature of the data! Are you using future information for data cleansing, feature generation or evaluation of your model?

Agenda

1 **Forecasting Use Cases @ DPDHL**

2 **Forecast Model**

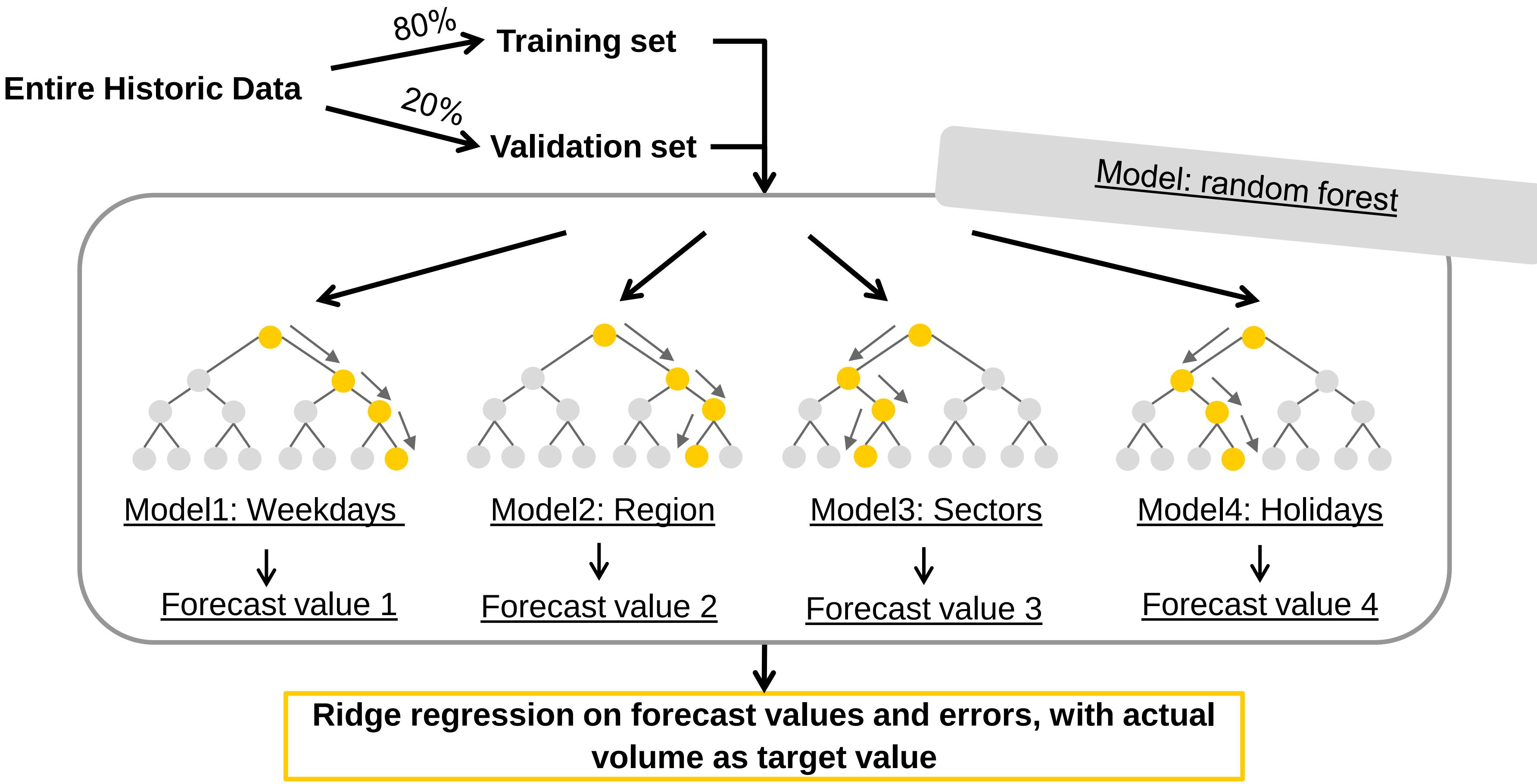
- ❑ Workflow for Developing a Forecasting Model
- ❑ Time-series vs Machine Learning approaches
- ❑ Particularities about Forecasting Projects

3 **Use Case Deep Dives**

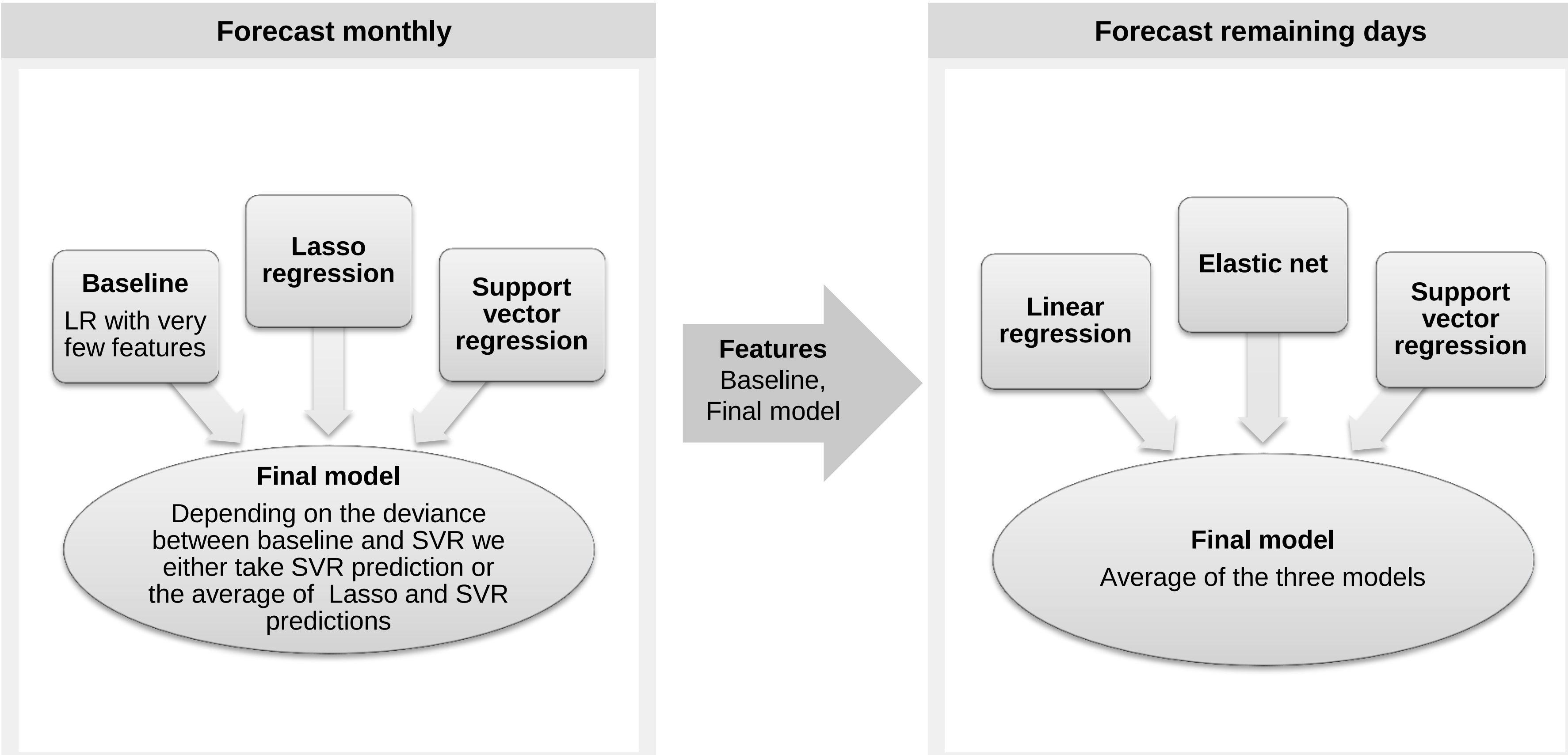
4 **New technologies for FC**



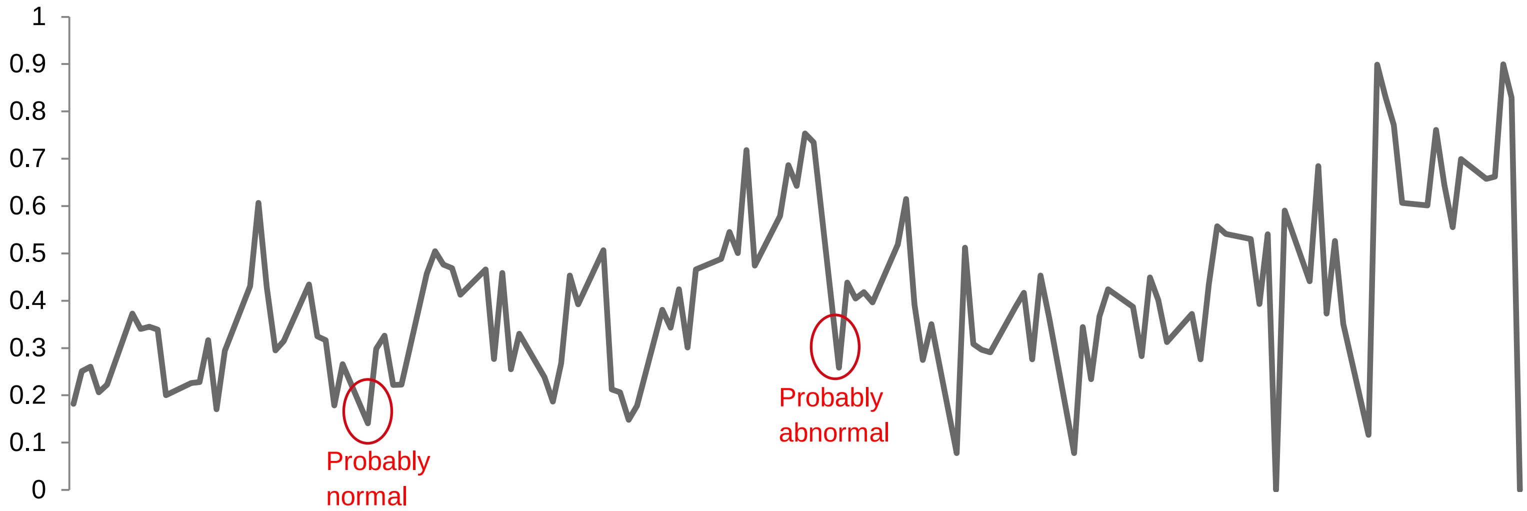

Case 1: Ensemble model focusing on different features of the data



Case 2: Two stage approach with first stage model generating features



Case 3: Data preprocessing based on wavelets



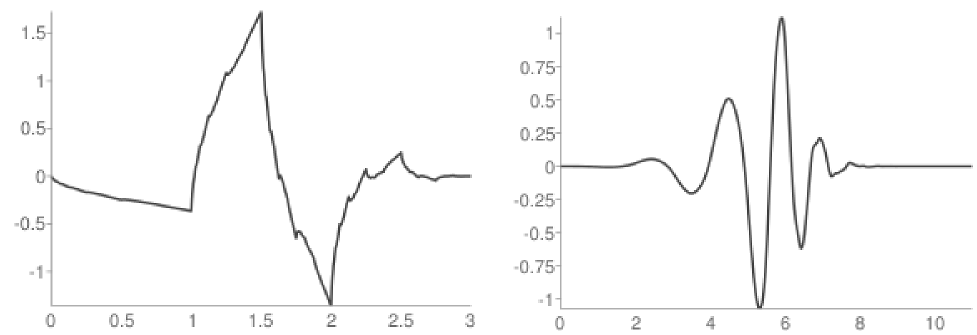
- Outliers in the data have a high impact on the forecast
- Context sensitivity makes outliers harder to detect in time-series data

Case 3: Data preprocessing based on wavelets

Discrete wavelet transform (DWT)

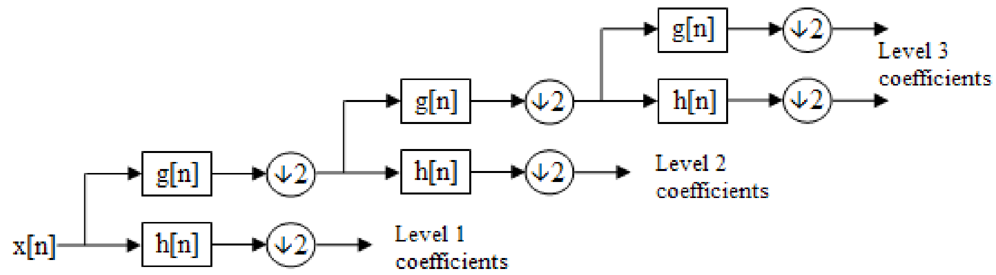
- Wavelets can be used to decompose signal by passing it through a series of filters
- In contrast to Fourier transforms, wavelet transforms capture both frequency and temporal information
- Based on the DWT coefficients the signal can be reconstructed
- By setting some coefficients to zero or only using coefficients of certain levels for the reconstruction a modified version of the signal is obtained in the original time domain (e.g. a smoothed time series)

Some examples of wavelets



Source: http://wavelets.pybytes.com/wavelet/db2/, http://wavelets.pybytes.com/wavelet/db6/

Schematic representation of a DWT



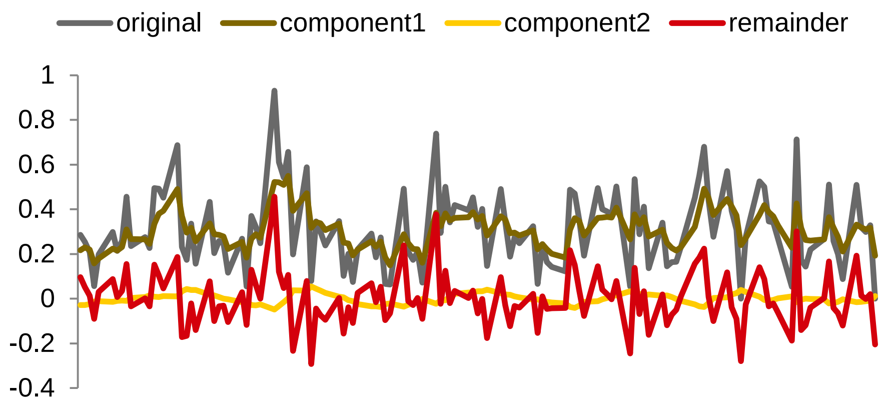
Source: https://upload.wikimedia.org/wikipedia/commons/2/22/Wavelets_-_Filter_Bank.png

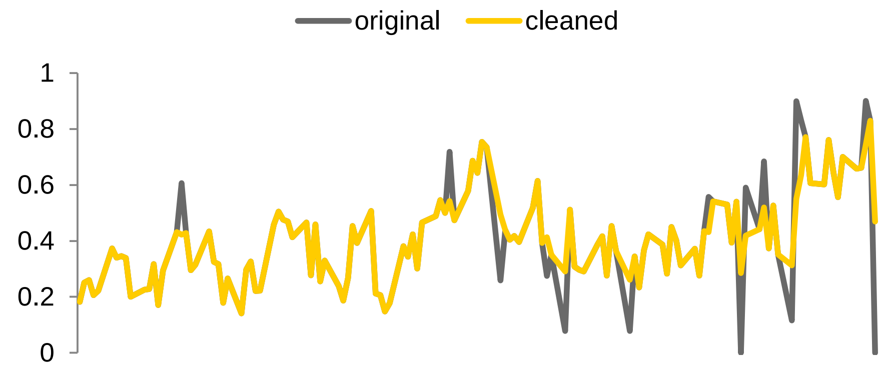
Case 3: Data preprocessing based on wavelets

Approach

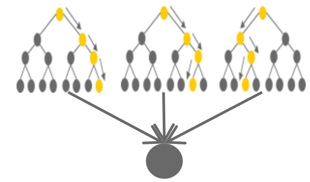
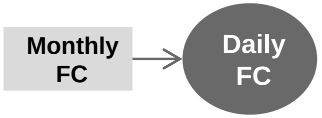
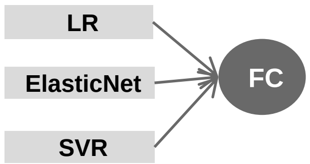
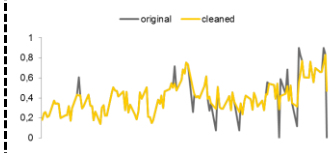
Two main approaches are used depending on the time series to forecast:

1. Decomposition into wavelet components
   - Time series is decomposed into wavelet components
   - Components are predicted separately
   - Predictions for components are added up again
2. Outlier correction based on wavelets:
   - Create a smoothed transformation of the time series by reconstructing it based on only the first few wavelet components
   - If the smoothed value deviates “too much” from the original value the data point is replaced by the smoothed value





Tips and tricks summary

1	Use ensemble of models focusing on different aspects of the data in each submodel	Case 1	
2	Include forecasted features to the final model	Case 2	
3	Use ensemble of models using a different method in each submodel	Case 2	
4	Invest time for pre-processing of your data → use “fancy” techniques if necessary	Case 3	

Agenda

1 Forecasting Use Cases @ DPDHL

2 Forecast Model

- ❑ Workflow for Developing a Forecasting Model
- ❑ Time-series vs Machine Learning approaches
- ❑ Particularities about Forecasting Projects

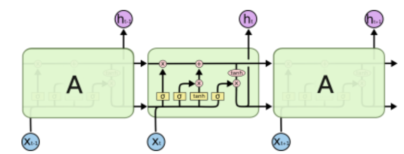
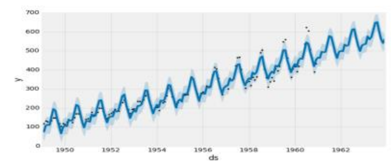
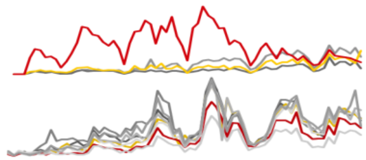
3 Use Case Deep Dives

4 New technologies for FC

New technologies for FC

We try to stay up to date with state-of-the-art technologies for Forecasting.

Some of the new models we’ve tested lately and might use in productive projects soon:

1	LSTM	
2	LightGBM	
3	Prophet Forecast	
4	Clustering of time-series before forecasting	

Questions?

If you have any question just drop us an email:

data_analytics@dpdhl.com

Katrin.Koenig@dpdhl.com