Forecast of High Frequency Energy Data Based on Machine Learning Applications

1st IAEE Online Conference

Concurrent Session 75: Market Modeling

Erik Heilmann
University of Kassel
Chair for Energy Economics
Erik.Heilmann@uni-kassel.de
+49 561- 804 7175

UNI KASSEL VERSITÄT

Co authors:
Janosch Henze
Heike Wetzel

Structure

UNI KASSEL VERSITÄT

Part 1: Machine Learning (ML) in energy economics

- Motivation
- What is ‚Machine Learning'?
- Modeling process of ML
- Application on energy demand data

Part 2: Exemplary application

- Case study setting
- Selected forecast approaches
- Results and Discussion
- Conclusion

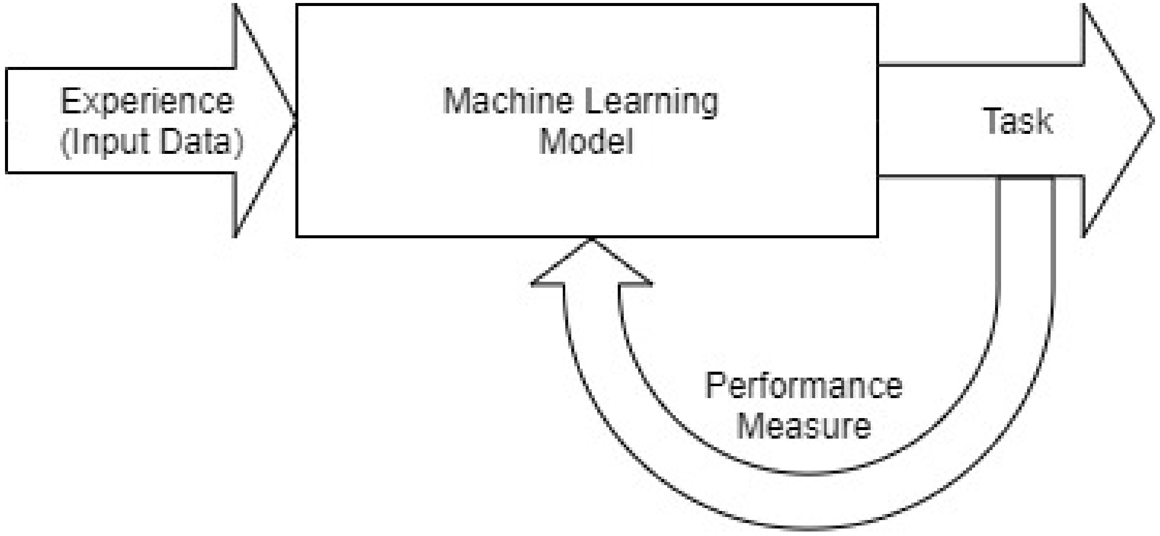
Motivation – Forecasts in Energy Economics

UNIKASSEL VERSITÄT

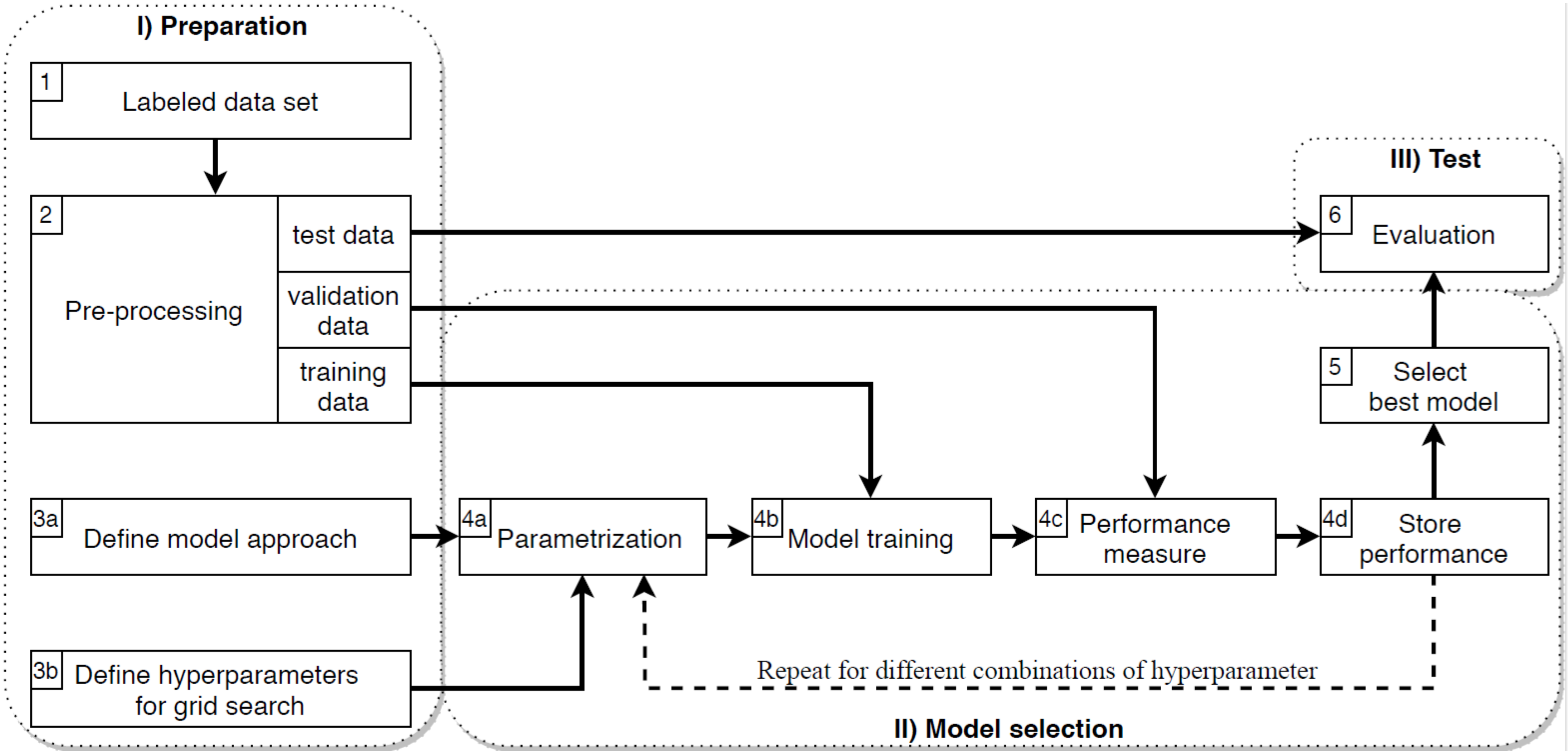
- Classical short-term forecasts‘ tasks:
  - Optimal scheduling of energy plants.
  - Network operation tasks.
  - Price indications for optimal marketing.
  - …
- Classical long-term forecasts‘ tasks:
  - Analysis of different energy sectors and peak demands to design future energy supply and networks.
  - Identification of strategies for energy efficiency enhancements and decreasing greenhouse gas emissions.
  - Giving the basis for policy decision-making.
  - …
- New use cases (e.g. local balancing of electricity) require forecasts with a high temporal and local resolution.
- Machine Learning (ML) techniques can provide forecasts based on big amount of data → state-of-the-art in technical oriented literature, but not in economic oriented literature.

What is ‚Machine Learning‘?

- No general definition in the literature.
- Main characteristics of a ML model:
  - It is based on some experience and therefore data-driven.
  - it gets trained on a specific task that can be measured by some performance measure (for example an error term). The performance, in general, improves with the quantity of appropriate experience.
- Most common model tasks are **prediction**, classification and clustering.
- Data can be **labeled** or unlabeled.



Modeling process of ML models

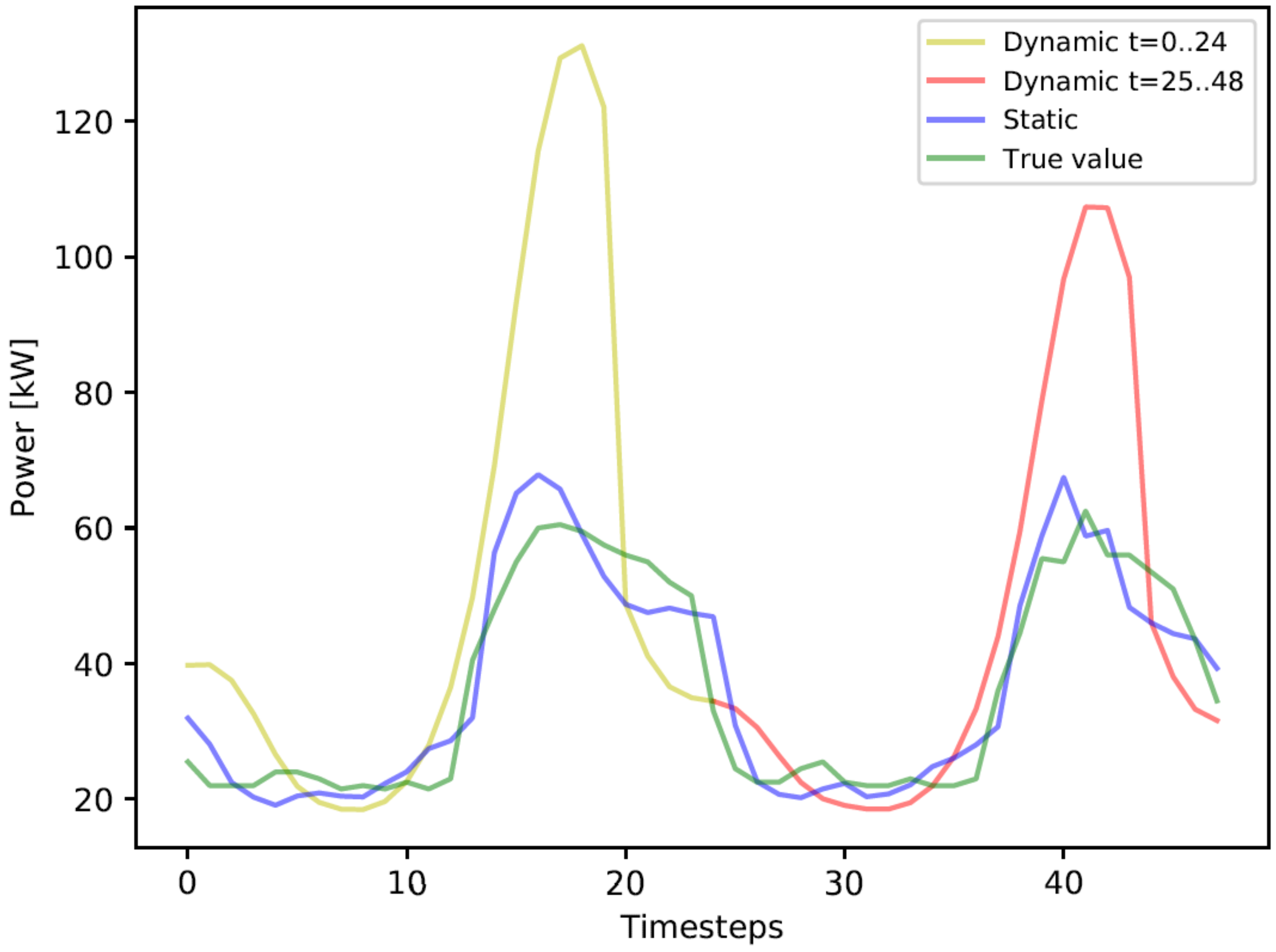


Application on Energy Demand Data

UNIKASSEL VERSITÄT

- Model task = prediction
- Target data = energy demand
- Input data:
  - (past) Energy demand
  - Time stamp
  - Weather (→ separate forecasts needed)
  - … any explanatory data
- Forecast horizon?
  - Static forecast for the next time step
  - Dynamic forecast for the next n time steps (using own prediction as inputs)

Exemplary static vs. dynamic prediction



Structure

Part 1: Machine Learning (ML) in energy economics

- Motivation
- What is ‚Machine Learning‘?
- Modeling process of ML
- Application on energy demand data

Part 2: Exemplary application

- Case study setting
- Selected forecast approaches
- Results and Discussion
- Conclusion

Case Study Setting

- **Aim of the case study:**
  - **Comparison of three different forecast approaches with different complexity.**
  - **Comparison of different input data complexity within these approaches.**
- One-year dataset:
  - Hourly electricity demand of 42 business consumers and one residual demand on the transformer (= target data).
  - Numerical weather prediction (NWP) data with 27 features.
- Input data:
  - Timestamp.
  - NWP.
  - Up to four past values of target data.

→ Variation of full input data set (35 exogenous variables) and reduced input data (5 exogenous variables).

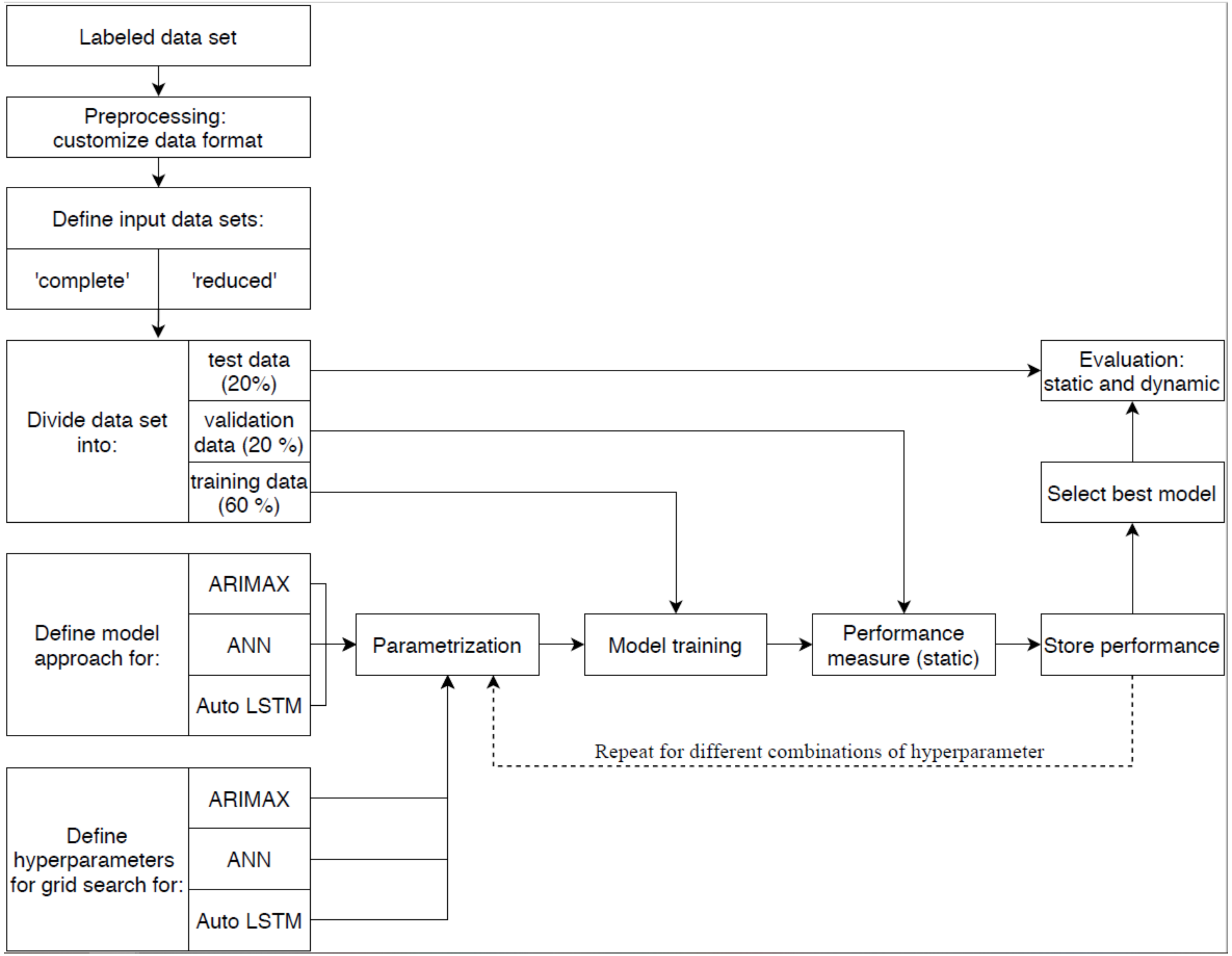
- Evaluation on normalized root mean squared error (RMSE) for **static** and **dynamic** forecasts.

Selected Forecast Approaches

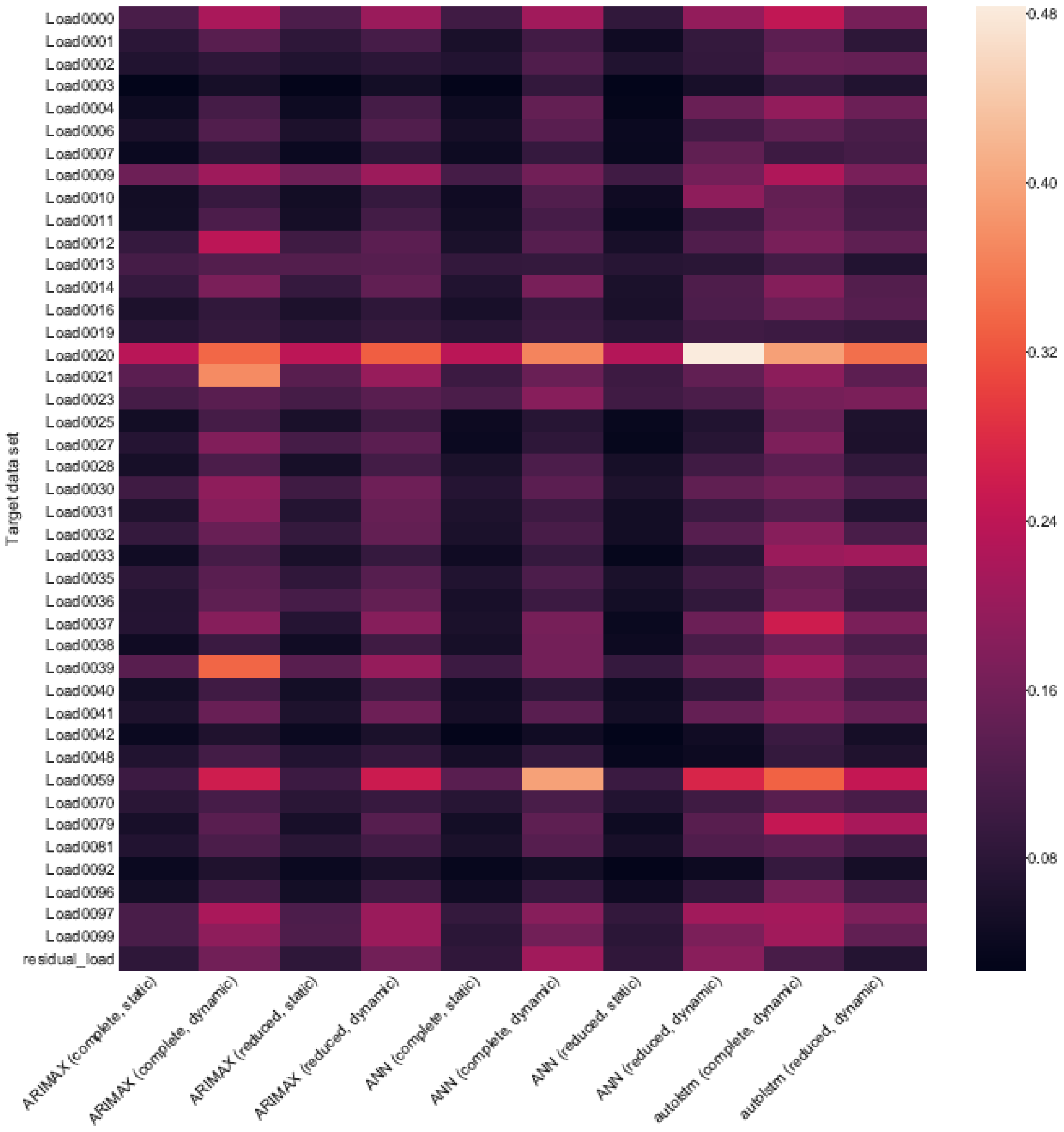
1. ARIMAX:
   - Autoregressive (AR) integrated (I) moving average (MA) models with exogenous input variables (X).
   - „Classical“ time series analysis.
   - Linear relationship between target variable and past values of target variable as well as exogenous inputs.
2. Artificial Neural Networks (ANN):
   - One of the most popular ML techniques. Part of the academic literature for more than 20 years.
→ Gain on importance in the last 5-10 years.
   - Network of ‚perceptrons‘ that process their input signal via a defined activation function. The perceptrons are connected via weights.
→ Non-linear relationship between target and input variables.
3. Auto-LSTM:
   - Combination of AutoEncoder (AE) neural network and a Long Short Term Memory (LSTM) network
→ Advancement of ANN.
→ Approach of the emerging field of deep learning.
   - Designed for dynamic forecast.

Case Study Overview

- Case study leads to 258 prediction models for the 43 target datasets (3 methods * 2 input variations = 6 models per dataset)
- Grid search for each model:
  - 75 ARIMAX configurations
  - 648 ANN configurations
  - 1152 Auto-LSTM variations
- Evaluation of Auto-LSTM only for dynamic forecast

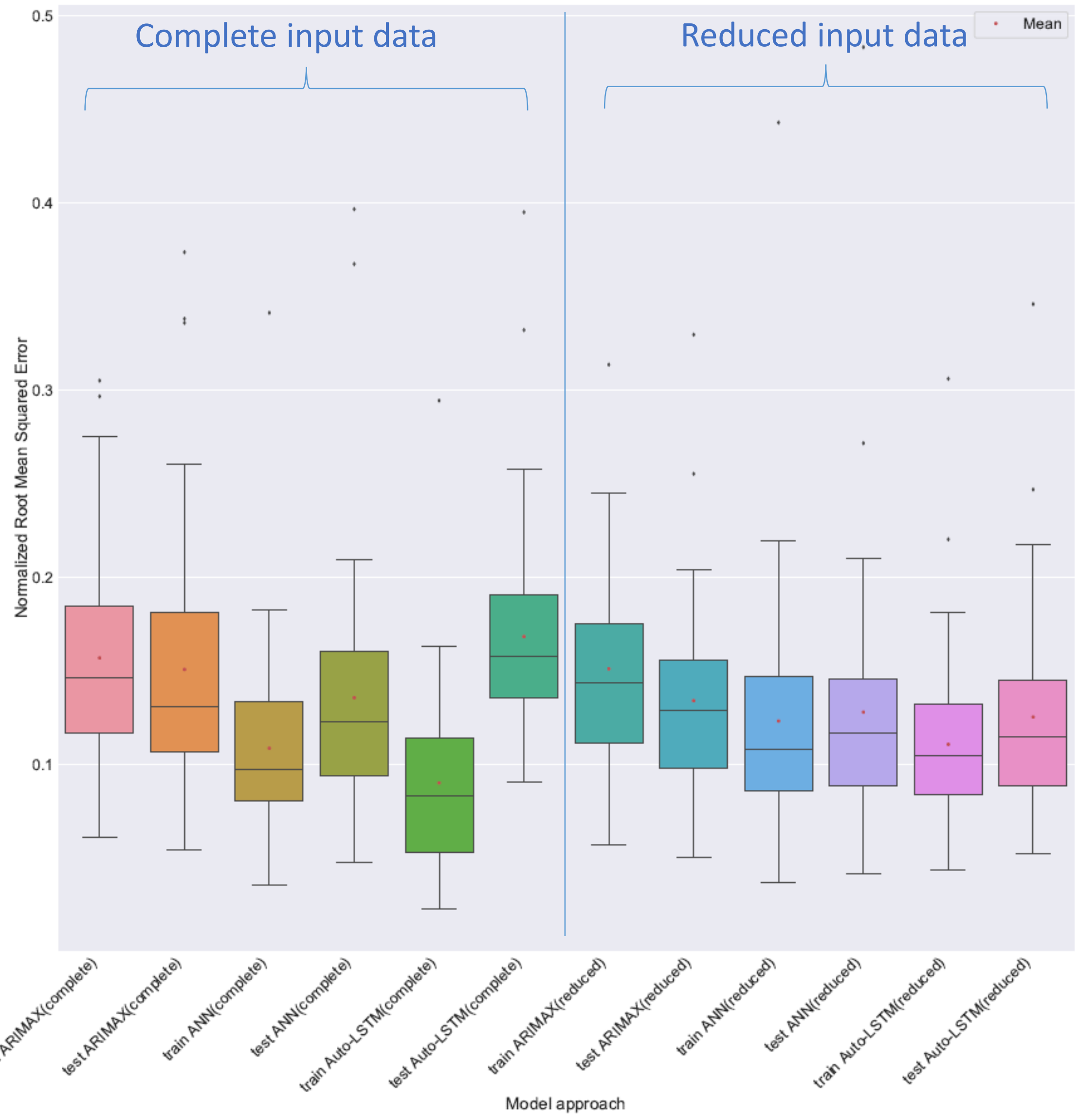


Results (normalized RMSE) - Overview



Results (normalized RMSE) - Dynamic Forecast

UNIKASSEL VERSITÄT



Conclusion

- ML as structured modeling process:
  - Data-based
    → generalize from training data onto unknown test data.
  - Aiming on specific target (e.g. prediction of a electricity demand).
    → Results give comparable information about the performance of each model.
- Results of the Case Study:
  - All the used approaches, in principle, able to provide forecasts with comparable quality.
  - No general relationship between high model complexity and good model performance.
  - No general relationship between the amount of different input data and good model performance.
  - However, it can be worth comparing models of different types and selecting the one that performs best for the pursued objective.
- The applied models provide black-box forecasts that can be used for further investigations, but may be difficult to interpret.

Thanks for your interest!
Questions?

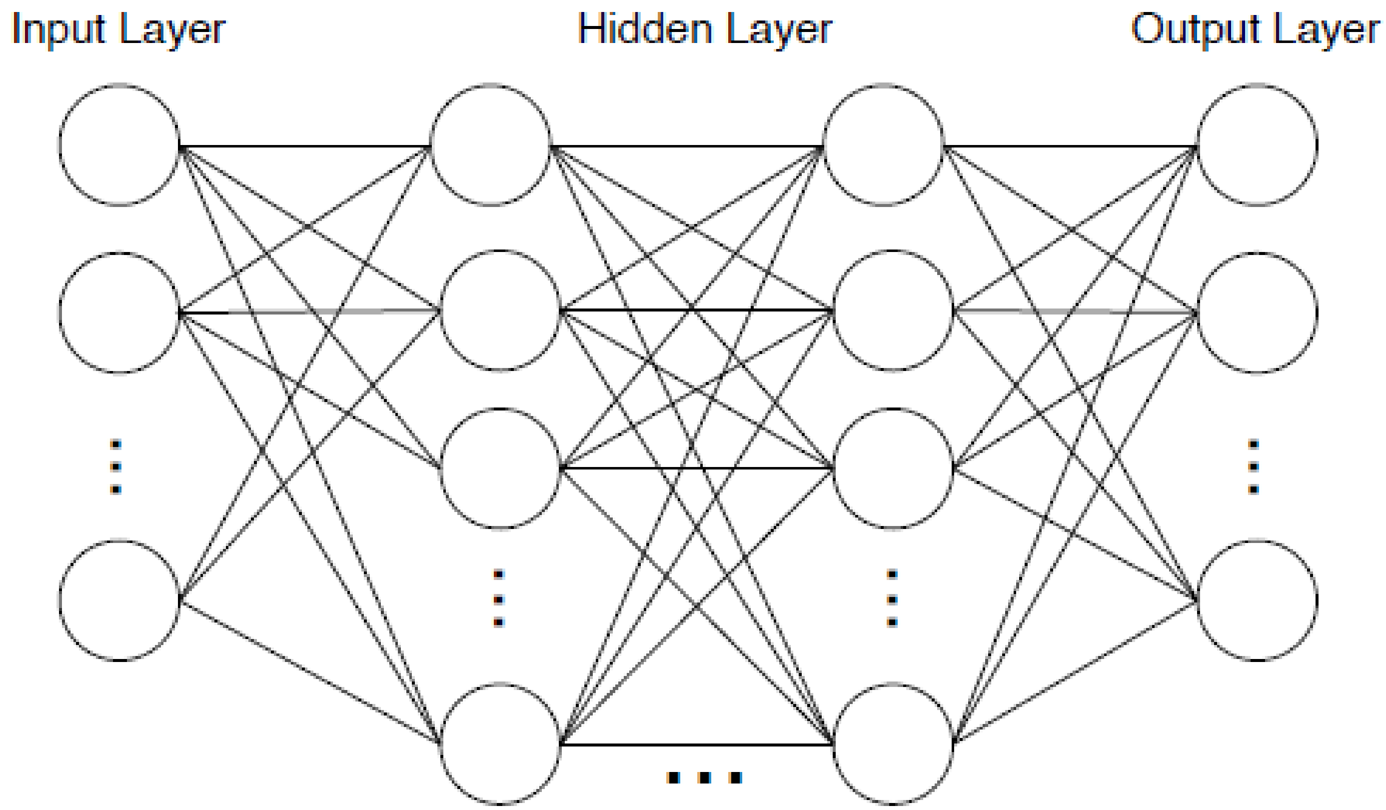
Erik Heilmann
University of Kassel
Chair for Energy Economics
Erik.Heilmann@uni-kassel.de
+49 561- 804 7175

UNI KASSEL VERSITÄT

Co authors:
Janosch Henze
Heike Wetzel

BACK UP

ANN structure



Grid search hyper-parameters

Table 1: ARIMAX grid search parameters.

Parameter	Range
p- and q-order	$[0, 1, 2, 3, 4]$
d-order	$[0, 1, 2]$

Table 3: AutoEncoder parameters split for full and reduced dataset..

	full	reduced
Learning Rate	$0.001 - 0.01$	$0.001 - 0.01$
Encoder Size	$[[34, 30, 20, 10]]$	$[[5, 10, 7, 4]]$

Table 2: ANN grid search parameters.

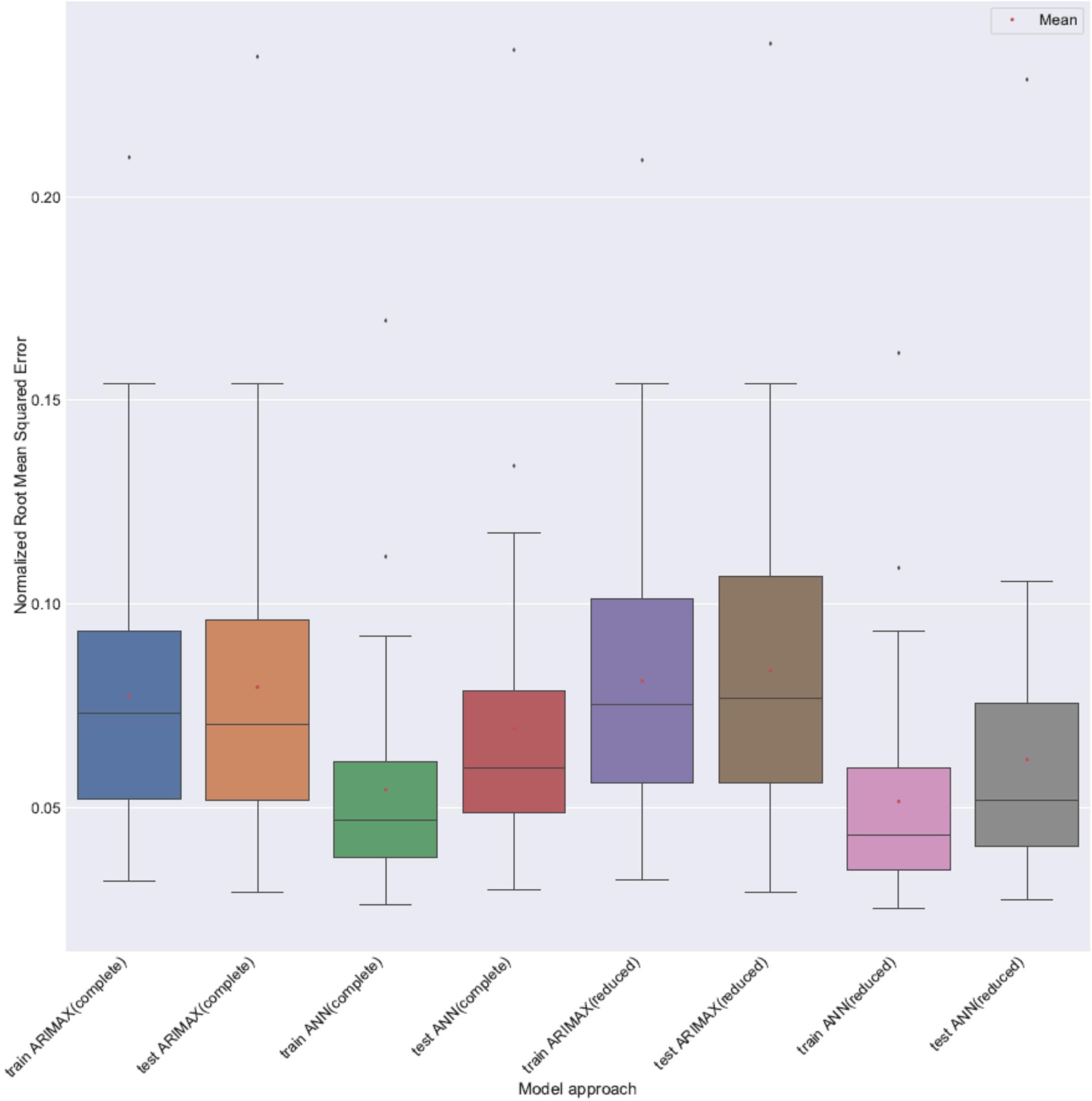
Parameter	Range
Number of layers	$[2, 3]$
Nodes in each layer	$[5, 10, 15]$
Regularization parameter	$[0.1, 0.01, 0.001]$
Activation functions	[`ReLU`, `tanh`]
Initial learning rate	$[0.1, 0.01, 0.001]$

Table 4: LSTM grid search parameters..

Parameter	Range
Learning Rate	$0.001 - 0.01$
Encoder Size (full or reduced)	$[34, 30, 20, 10]$ or $[5, 10, 7, 4]$
Number of LSTM cells	$[1, 2, 4]$
Hidden size of LSTM cells	$[1, 5, 10]$
Random initialization of LSTM weights	[`True`, `False`]

Results Static Forecast (normalized RMSE)

UNIKASSEL VERSITÄT



Discussion of Results

- Static errors are often smaller than dynamic errors, which is in line with the expectation that the error term accumulates with every additional forecast time step.
- Train errors of ML models in general lower than test errors.
- For static forecast ANN outperforms ARIMAX.
- For dynamic forecast no general advantage of any of the tested methods.
- No general difference in model performance between the full input dataset and the reduced dataset.