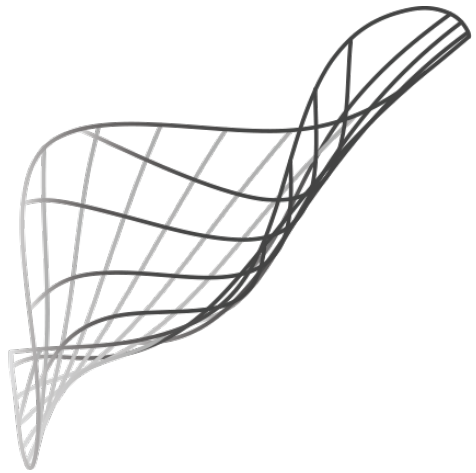
TANGENT WORKS

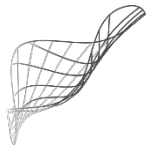
ADVANCED FORECASTING



Automatic Model Building for Time-series in Energy Industry

Ján Dolinský
jan.dolinsky@tangent.works

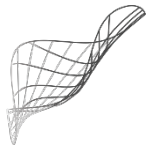
TANGENT WORKS
ADVANCED FORECASTING



TIM
Tangent Information Modeller

Ján Dolinský
jan.dolinsky@tangent.works

TANGENT WORKS
ADVANCED FORECASTING



Content

- Birthday Problem
- Time-series Problems in Energy Industry
- Data Science Process
- Live modeling demonstration
- Large-scale Forecasting systems & Why Automation Matters
- GefCom 2014, 2017 results & Summary

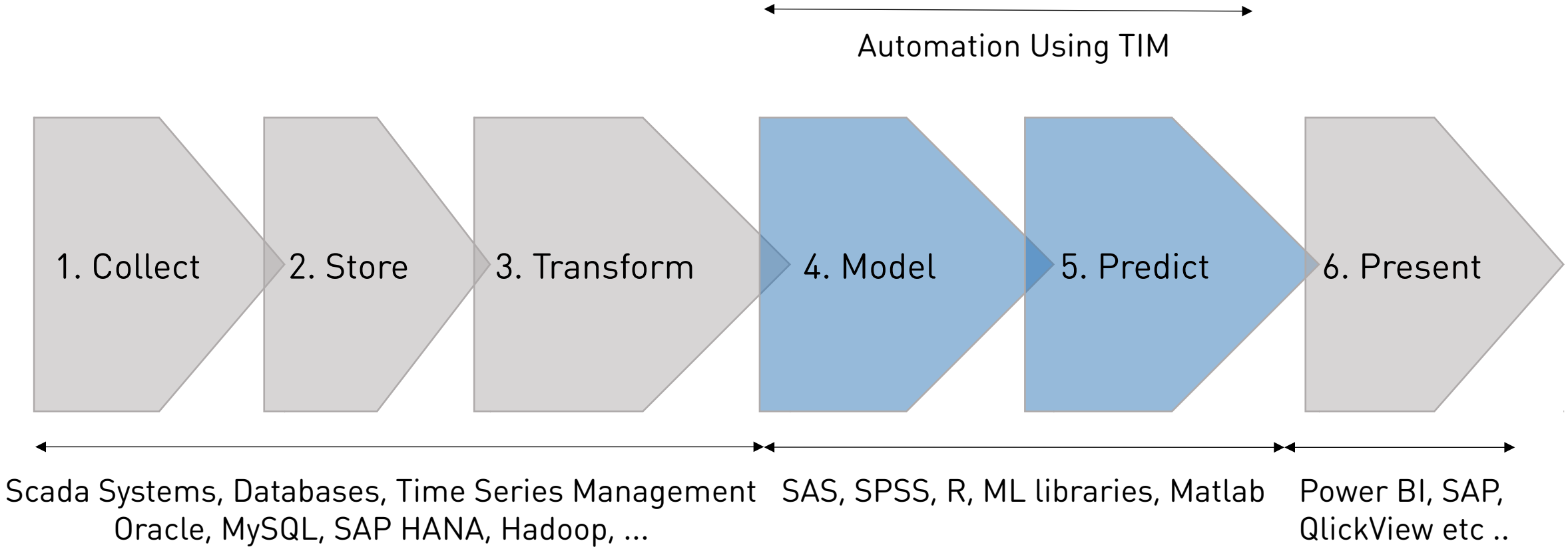
Birthday Problem

- 366 → 100%
- 120
- 75
- 23
- 6

Time-series Problems in Energy Industry

- Electricity Load (aggregated and individual)
- Technical Losses
- System Imbalance
- Gas Consumption (District Heating)
- District Cooling
- Wind Production
- Solar Production
- Electricity Price

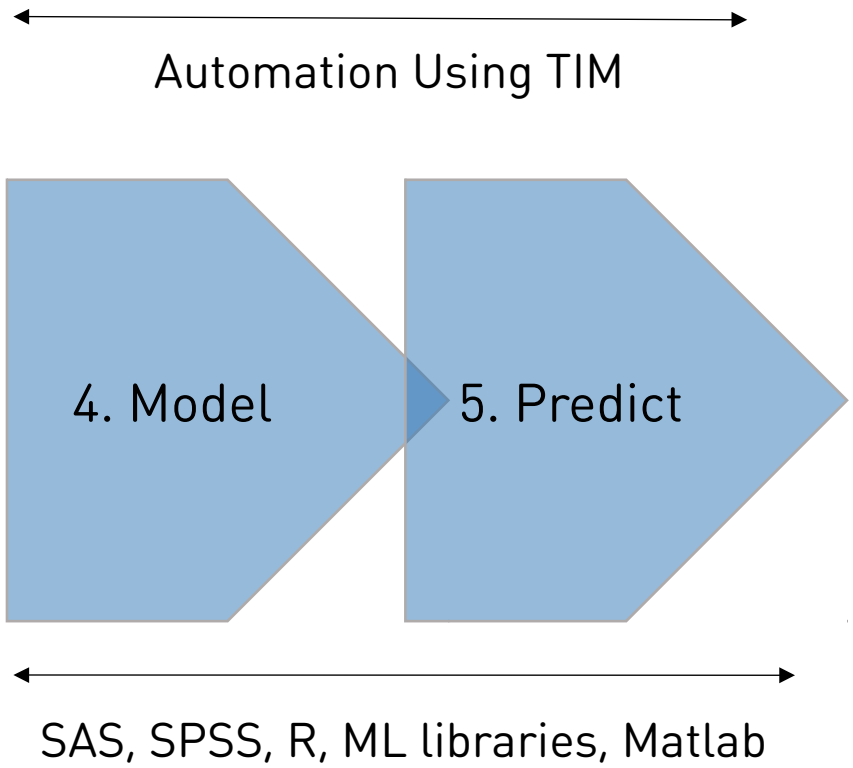
Data Science Process



Data Science Process

- feature engineering
- model selection

- tedious
- costs money



Data Science Process

Which features and how many ?

$$y(k) = f\left(Temp(k-3), Temp(k-22) * Wind(k-1), y(k-24)\right)$$

NN, SVM, MARS, LASSO, RF, ... ?

Data Science Process

$$y(k) = f\left(Temp(k-3), Temp(k-22) * Wind(k-1), y(k-24)\right)$$

Temp(k-1), Temp(k-2), ..., Temp(k-24)
Wind(k-1), Wind(k-2), ..., Wind(k-24)
y(k-1), y(k-2), ..., y(k-24)

24 + 24 + 24 + 24*24 = 24*27 = 648

Temp(k-1), Temp(k-2), ..., Temp(k-96)
Wind(k-1), Wind(k-2), ..., Wind(k-96)
y(k-1), y(k-2), ..., y(k-96)

96 + 96 + 96 + 96*96 = 96*99 = 9504

Information Geometry: Let Your Data Speak

Prof. Akaike – 1974 proposed AIC, one of the first information criteria

Prof. Amari – cca 1998 proposed to use differential geometry for non-parametric inference (e.g., training of neural-nets)

Prof. Konishi – 2004 proposed GBIC and general way of inferring new information criteria
Book: Information Criteria and Statistical Modeling

IG is taking advantage of using manifolds and geometrical structures to understand data and generate meaningful and interpretable models.

TIM: How It Works

seconds

Data intake

Synthesis of features based on a set of heuristic rules

Variable selection is carried out in Euclidean space using Bayesian inference

Model Complexity is estimated using a customized IC

Model is generated

Data Science Process – Industrial View on Data Science

$$y(k) = f\left(Temp(k-3), Temp(k-22) * Wind(k-1), y(k-24)\right)$$

Business/Industrial users value models that are

- Transparent
- Easy to deploy and use
- Easy to maintain
- Have accuracy which is good enough

Data Science Process – Industrial View on Data Science

$$y(k) = f\left(Temp(k-3), Temp(k-22) * Wind(k-1), y(k-24)\right)$$

Over-engineered models of e.g., Kaggle competitions are of limited use in industrial practice

- Such models are often (over) optimized for a single dataset
- Neural-net combined with genetic algorithm may have the best accuracy but ...

Why TIM

- TIM does it all automatic
  - ✓ Zero degrees of freedom
  - ✓ No feature engineering
  - ✓ High speed computing on standard hardware
- Accuracy often matches or outperforms manually build models
- TIM models are transparent, bring extra knowledge and confidence

Live Demonstration of TIM

Large-scale Forecasting Systems & Why Automation Matters

Example 1 (totaling in 8*2 + 5*2 + 3*1 = 29 models)

- 5 electricity load asset and 3 solar farms = 8 assets in total
- Day-ahead and week-ahead scenarios for each asset + nowcasting (intra-day)

for the solar assest

- Dynamic data availability for a single asset e.g., historical data of yesterday may not come
- Models are re-built on-demand

Large-scale Forecasting Systems & Why Automation Matters

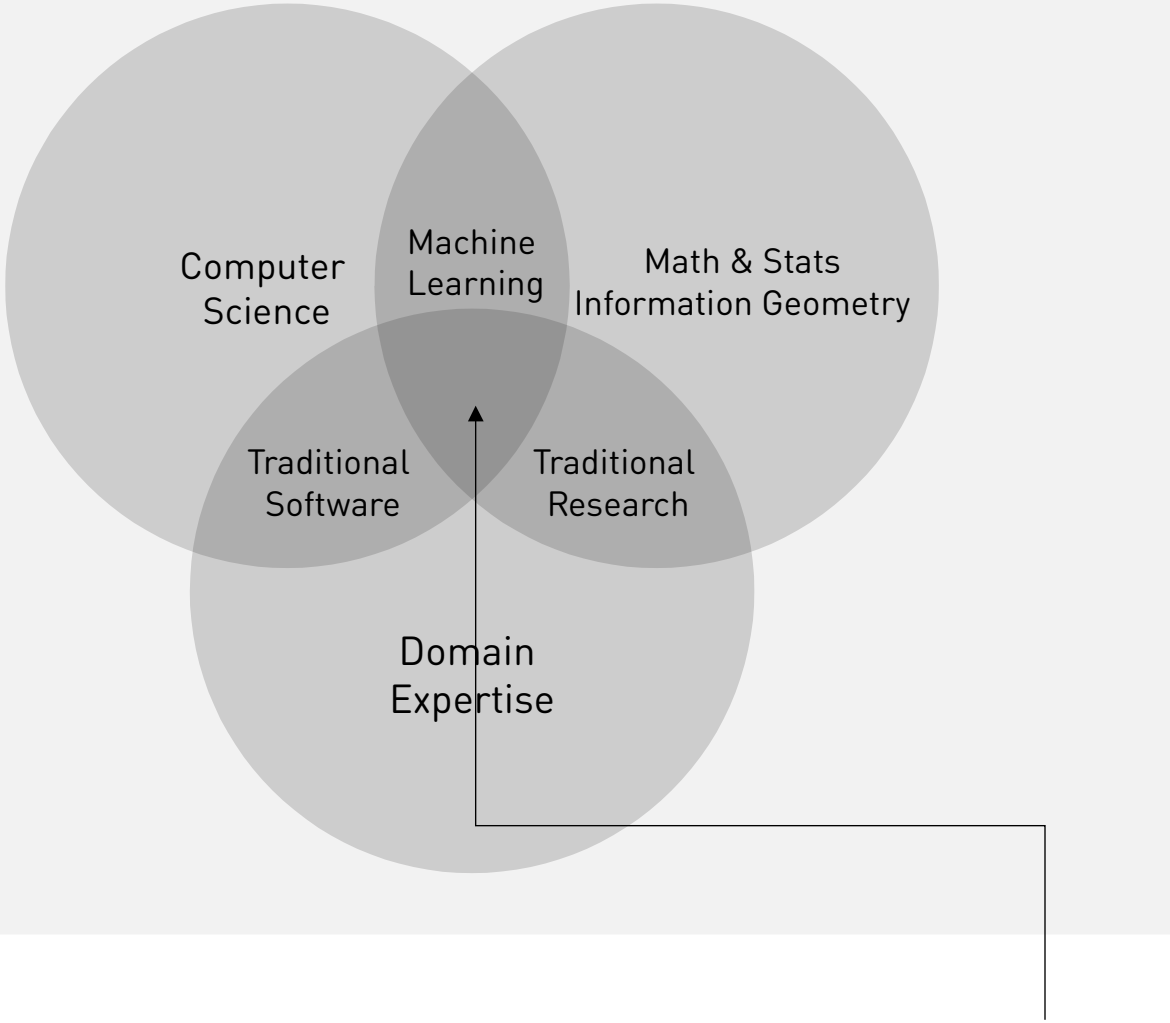
Example 2 (totaling in 1800*2 + 1800*2 = 7200 models)

- 1800 electrical assets
- Day-ahead and week-ahead model for each asset
- Dynamic data availability for any asset
- Automatic model re-building to account for structural changes

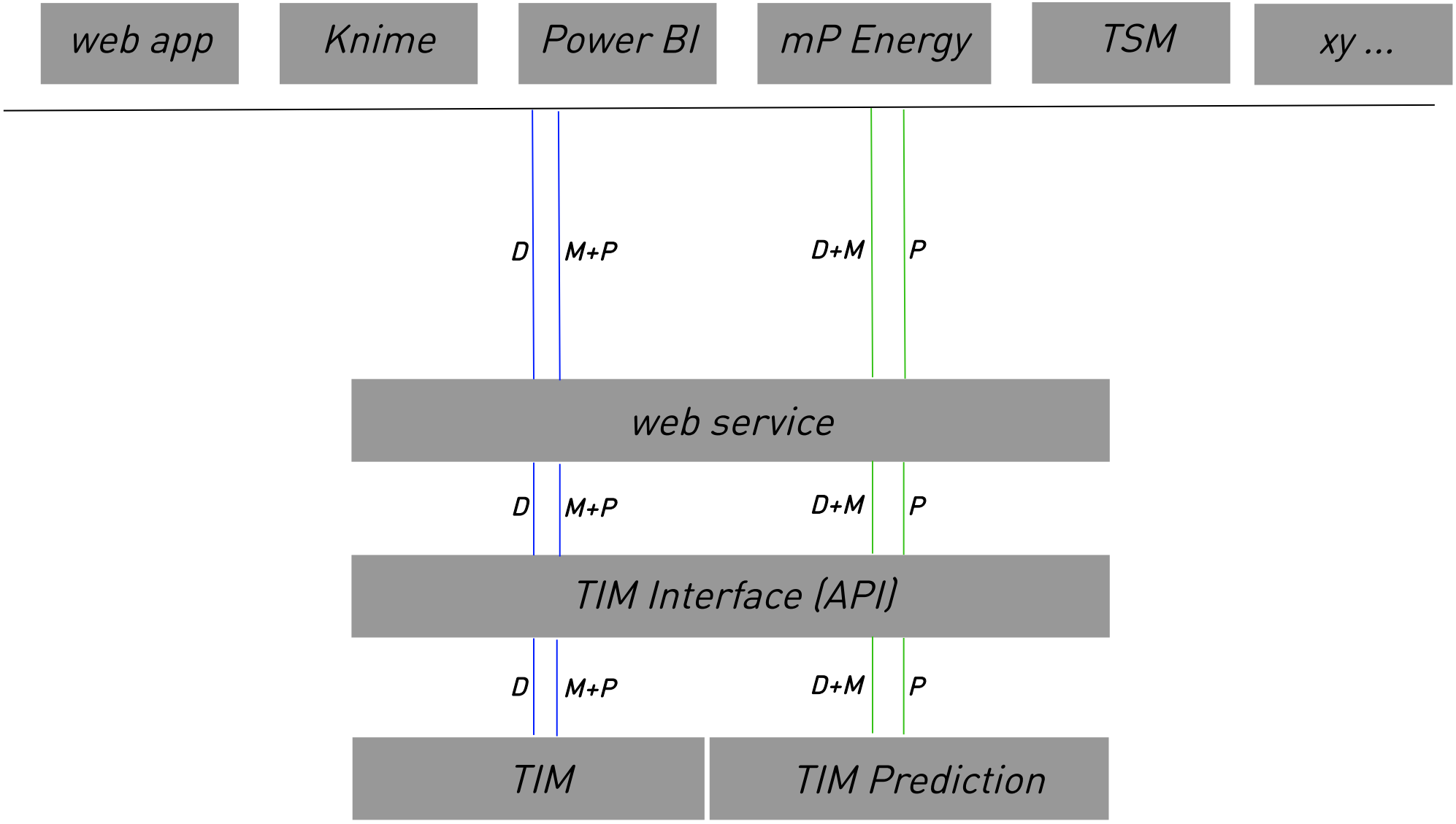
Julia Language

Walks like Python runs like C

- A single platform for prototyping and production → significant gains in efficiency of product development
- Vectorized code runs equally fast as de-vectorized
- Vectorization on a single thread level
  - SIMD out of the box
  - Direct calls to BLAS
- Distributed parallel computation (multiple threads)



TANGENT WORKS
ADVANCED FORECASTING



TIM Architecture

legend: D - data, M - model, P - prediction

Training flow

Testing flow

Clients

Prototype Clients

Technology Partners

GefCom 2014 ex-post results

- Electricity Load track: 3rd place out of about 300 teams
- Electricity Price trak: 2nd place out of about 250 teams
- Wind Generation track: 7th place out of about 250 teams
- Solar Generation track: 9th place out of about 250 teams

Andritz Hackathon 2017

1st place out of 7 ML companies

GefCom 2017 current ex-ante results

- Electricity Load track: 2nd place out of about 73 teams

Summary

- Fully automated model building
- Accuracy often outperforms manually build models
- Quick insights into a time-series of interest
- Tedious model building is an option not necessity

Thank you.

jan.dolinsky@tangent.works

Tangent Works
Na Slavíne 1
81106 Bratislava
www.tangent.works

TANGENT WORKS
ADVANCED FORECASTING