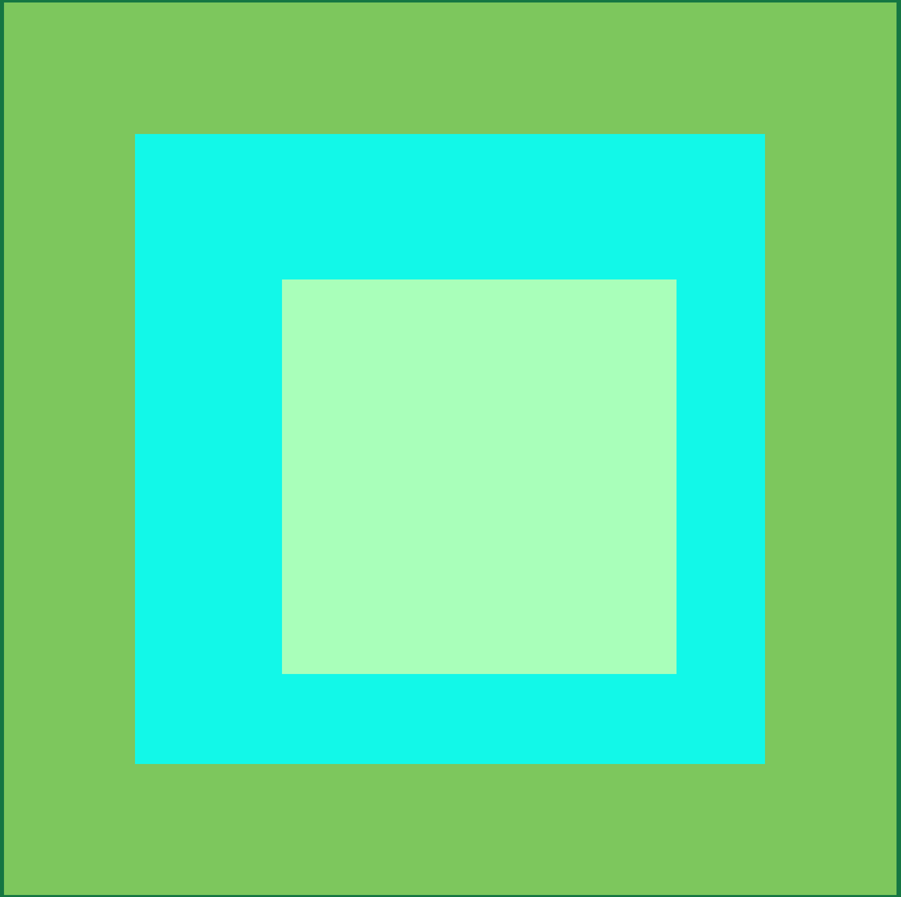
Single and multiple step forecasting of solar power production:

applying and evaluating potential models.

Final Presentation 27 May 2019



Process data

Machine learning

Start-up

Sun Labs

SUN LABS

Data visualization and real-time logging for solar cell facilities

Primary research question

Create a solar power forecasting model based on Sun Labs collected data.

Can exogenous variables increase accuracy of the models?

Sala Facility

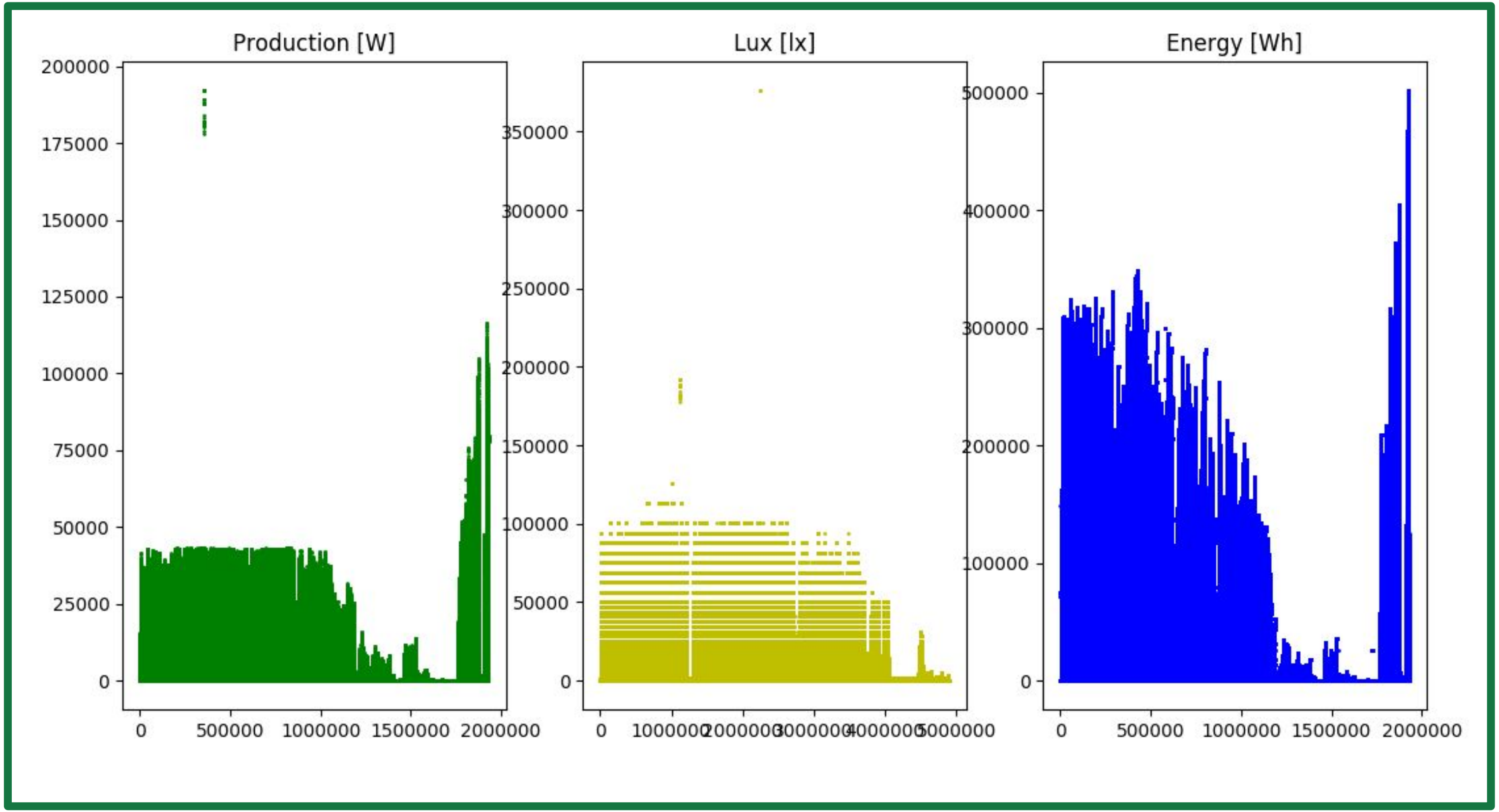
Photo from Solel i Sala och Heby

Installed max power: 47 kW

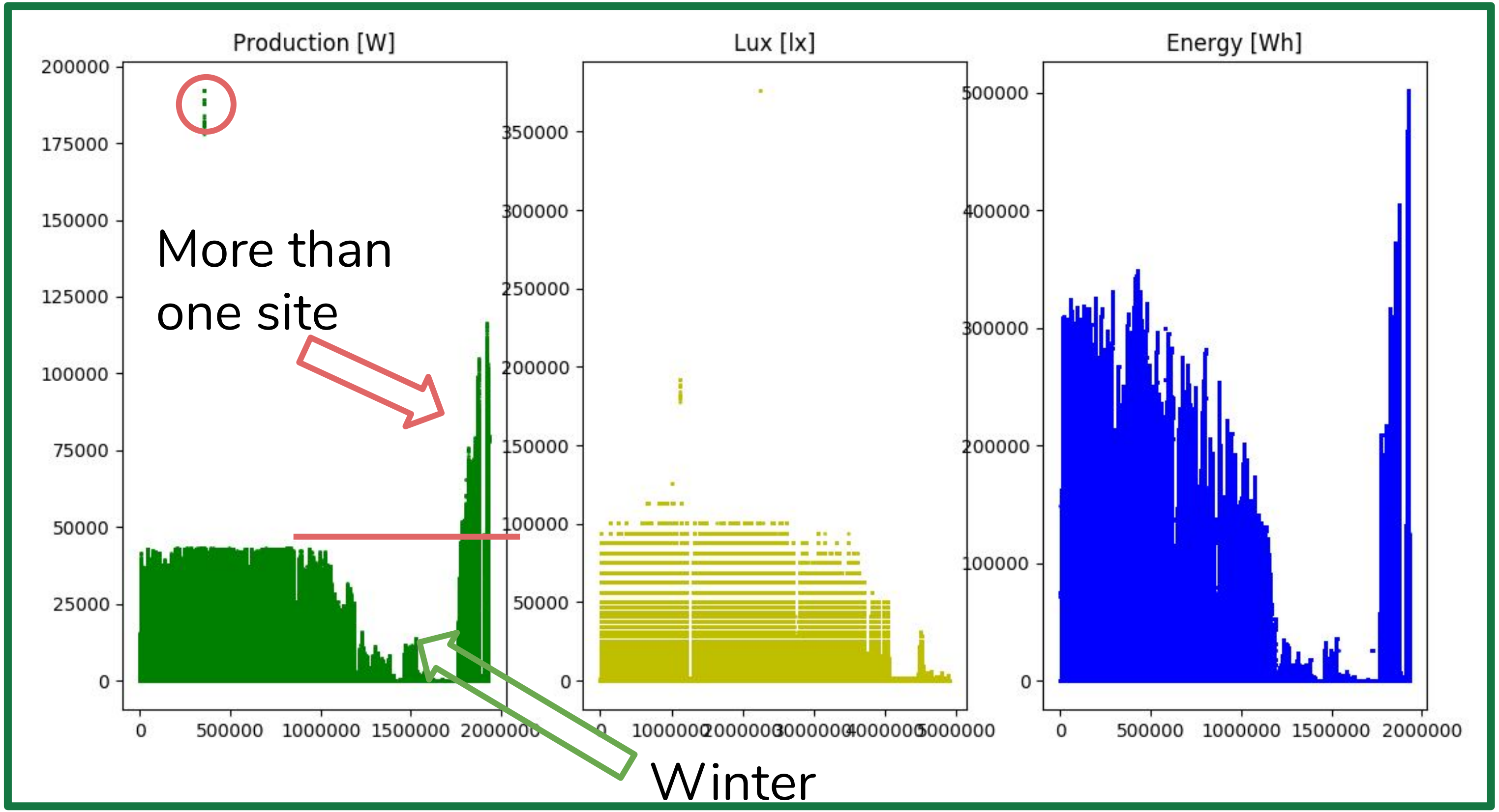
Yearly energy production: 40 MWh

Data Limitation

Ex: Data cleaning



Ex: Data cleaning



Data Limitation: exogenous predictors

Meteorological conditions

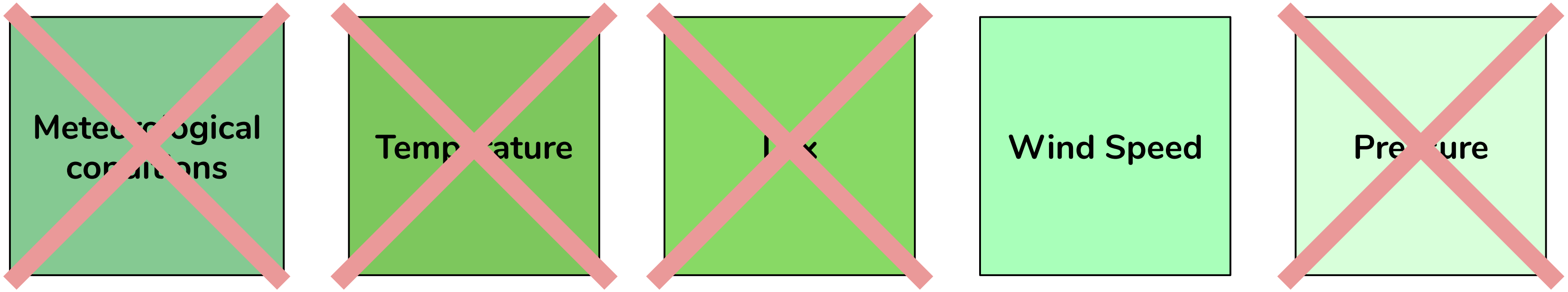
Temperature

Lux

Wind Speed

Pressure

Data Limitation: exogenous predictors



Model choice

1

Linear model: ARIMA and ARIMAX

Model choice

1

Linear model: ARIMA and ARIMAX ->
single step forecast

Model choice

1

Linear model: ARIMA and ARIMAX ->
single step forecast

2

Nonlinear model: Gaussian process ->
multiple step forecast

Research question

- Which of the various compared models has the highest single-step forecasting accuracy?
- How does the usage of an exogenous predictor change the performance of a forecasting model and ultimately the forecast accuracy?
- Which of the two compared models has the highest multiple step forecasting accuracy?

Research question

- Which of the various compared models has the highest single-step forecasting accuracy?

Root Mean Square Error

ARIMA, ARIMAX, Persistence

Research question

Wind Speed

How does the usage of an exogenous predictor change the performance of a forecasting model and ultimately the forecast accuracy?

Root Mean Square Error

ARIMA, ARIMAX

Research question

Gaussian process, persistence

Root Mean Square Error

Which of the two compared models has the highest multiple step forecasting accuracy?

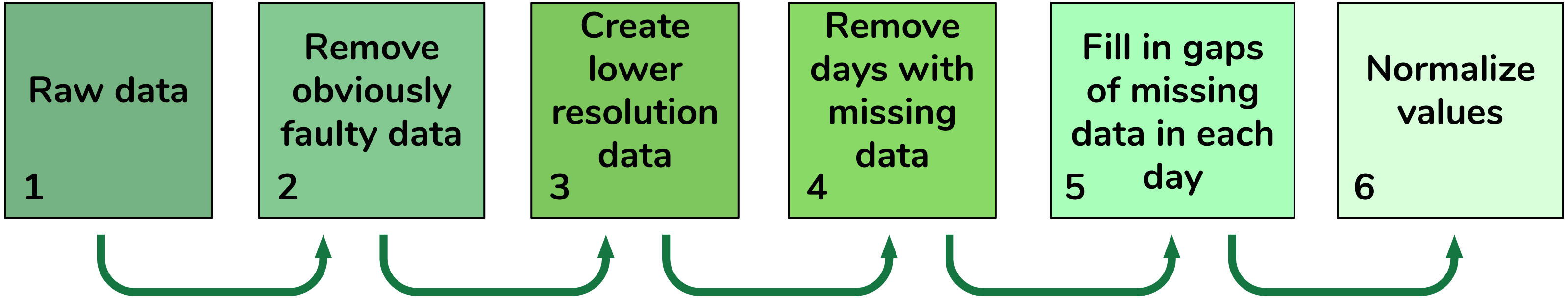
To answer research questions

Data preprocessing

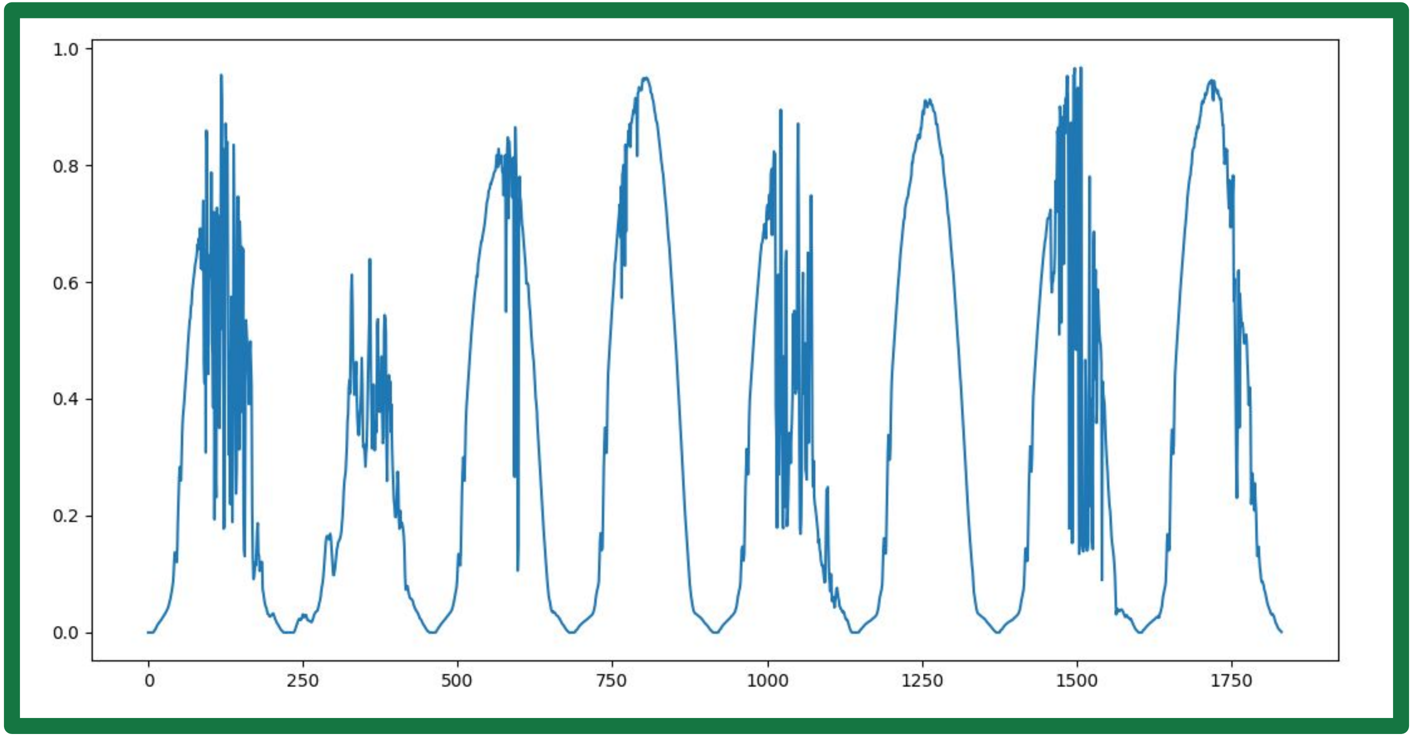
Single step implementation	Multiple step implementation
Persistence	Persistence
ARIMA	Gaussian process
ARIMAX	

RMSE

Data preprocessing



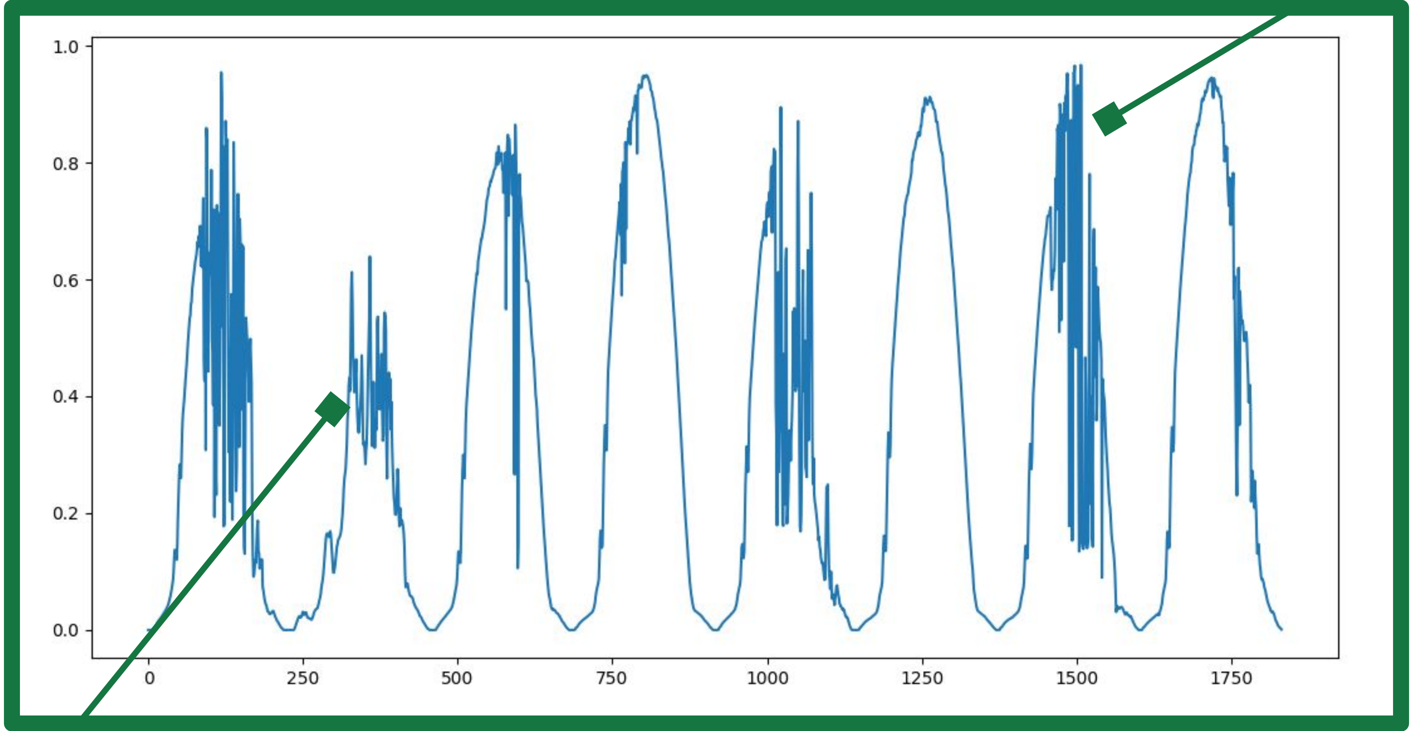
Data preprocessing



Example 8 days of July

Data preprocessing

Cloudiness



"Rainy day"

Example 8 days of July

Persistence

Forecast value the same as present value

Persistence

Forecast value the same as present value

$$\hat{y}_t = y_{t-1},$$

ARIMA

Integrated

AR | I | MA

Auto Regressive

Moving Average

ARIMA

ARIMA(p,d,q)

$$y'_t = c + \varphi_1 y'_{t-1} + \cdots + \varphi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \ ,$$

p: lags of earlier values	AR
d: order of differencing of y'	I
linear combination of q lags of forecasted errors.	MA

ARIMAX

Integrated

Exogenous Wind Speed

AR I MA X

Auto Regressive

Moving Average

ARIMAX

$$y'_t = bx_t + \varphi_1 y'_{t-1} + \cdots + \varphi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \,,$$

Exogenous Wind Speed

Gaussian process

A Gaussian process is a finite collection of random variables, which have a joint distribution.

Gaussian process

A Gaussian process is a finite collection of random variables, which have a joint distribution.

Different kernel characteristics (assumptions) decide forecast function form

Gaussian process

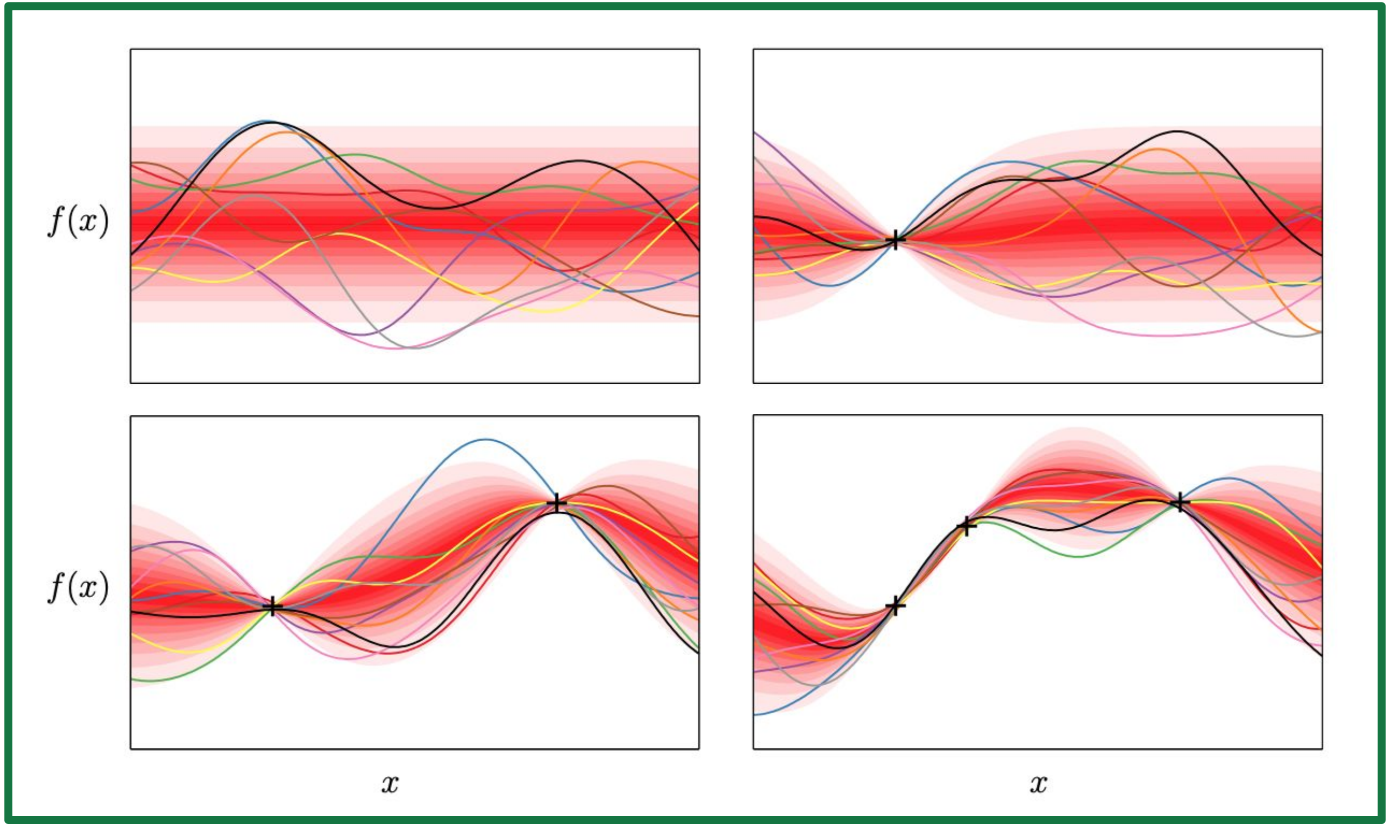


Figure from (Duvenaud, 2014)

RMSE

Forecast

Actual value

$$RMSE = \sqrt{\frac{\sum_{n=1}^{N}(\hat{y}_n - y_n)^2}{N}}$$

Final Presentation 27 May 2019

Single step model implementations

Compared models: Persistence, ARIMA, ARIMAX

Two different time steps 5 minutes and 10 minutes

ARIMA and ARIMAX orders (5,1,0), (5,1,1) and (4,1,2).

Random shuffle days

Trained using the first 24 days (≈86% of dataset) of July

Forecast next time step for 4 days (≈14% of dataset) of July

Result: single step models

RMSE	(5,1,0)		(5,1,1)		(4,1,2)	
	5 minutes	10 minutes	5 minutes	10 minutes	5 minutes	10 minutes
Persistence	7.65%	8.89%	7.65%	8.89%	7.65%	8.89%
ARIMA	7.49%	8.72%	7.52%	8.73%	7.54%	8.69%
ARIMAX	7.47%	8.68%	7.49%	8.70%	7.52%	8.66%

Result: single step models

ARIMA, ARIMAX order

RMSE	(5,1,0)		(5,1,1)		(4,1,2)	
	5 minutes	10 minutes	5 minutes	10 minutes	5 minutes	10 minutes
Persistence	7.65%	8.89%	7.65%	8.89%	7.65%	8.89%
ARIMA	7.49%	8.72%	7.52%	8.73%	7.54%	8.69%
ARIMAX	7.47%	8.68%	7.49%	8.70%	7.52%	8.66%

Result: single step models

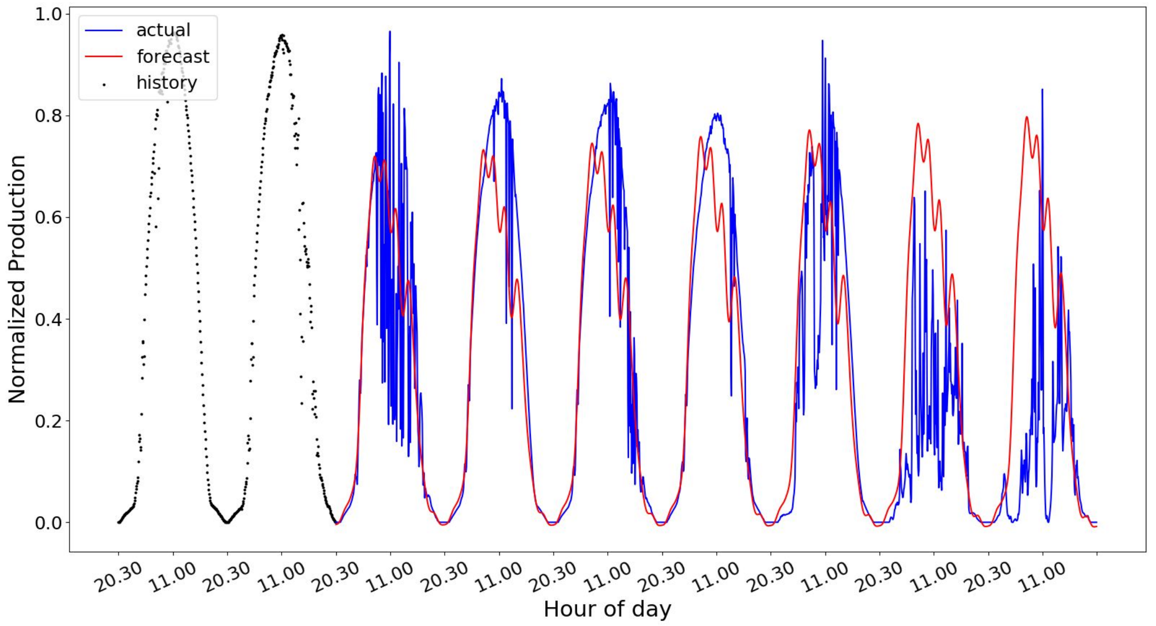
ARIMA, ARIMAX order

RMSE	(5,1,0)		(5,1,1)		(4,1,2)	
	5 minutes	10 minutes	5 minutes	10 minutes	5 minutes	10 minutes
Persistence	7.65%	8.89%	7.65%	8.89%	7.65%	8.89%
ARIMA	7.49%	8.72%	7.52%	8.73%	7.54%	8.69%
ARIMAX	7.47%	8.68%	7.49%	8.70%	7.52%	8.66%

Multiple step model implementations

- Compared models: One week persistence, Gaussian process
- Time step of 5 minutes
- Only observing days between production start and end 1.30 - 20.30
- According to (Dahl, 2017)
- Trained using the first 21 days (75% of dataset) of July
- Forecast all time steps for 7 days (25% of dataset) of July

Result: multiple step models

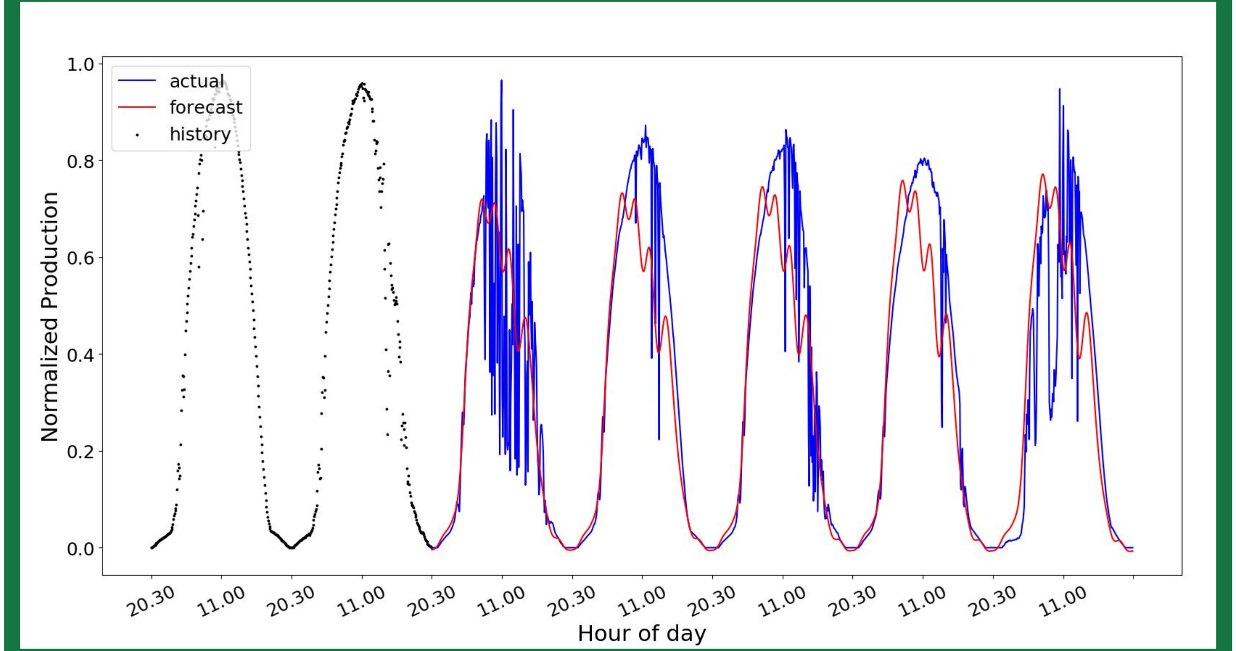


7 days

Persistence
RMSE: 22.61%

Gaussian process
RMSE: 19.64%

Result: multiple step models



5 days

Persistence
RMSE: 20.0%

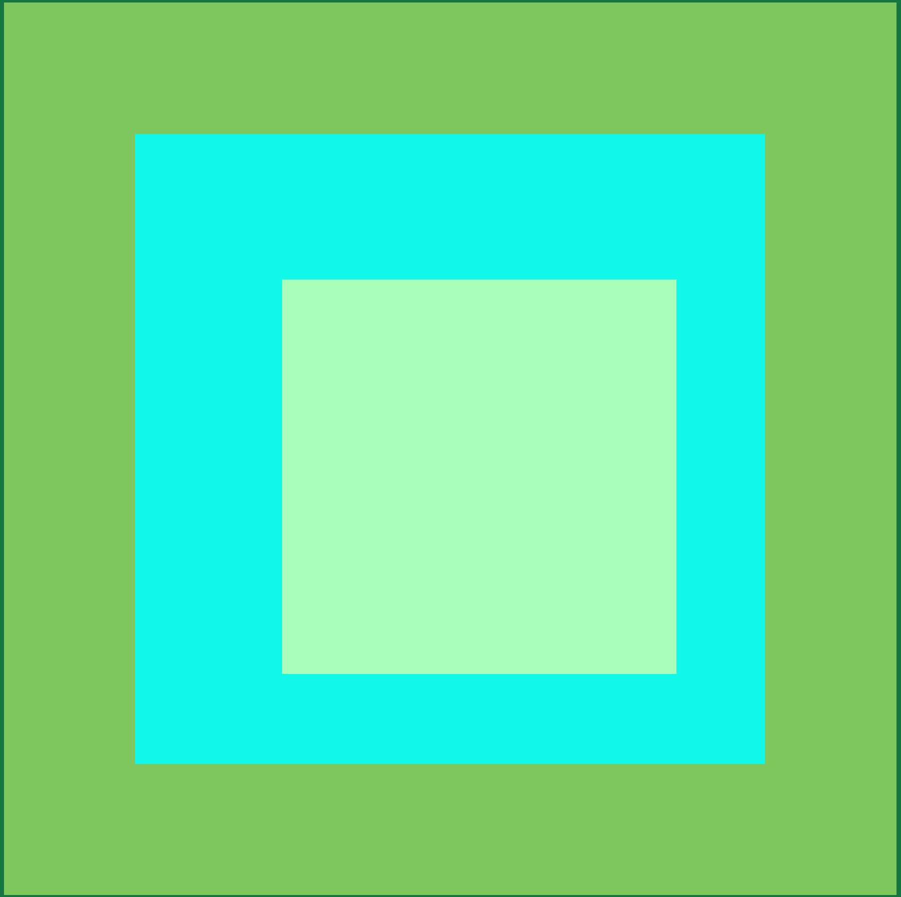
Gaussian process
RMSE: 13.7%

ARIMA is more accurate than persistence in applied model

Adding an exogenous variable did not help improve accuracy significantly

Gaussian process is more accurate than persistence in applied model

However, the applied model can be further developed by adding exog. variable or more history



Computational time

& time...

Final Presentation 27 May 2019

Thank you!