TALL SOM FORTELLER

Prophet

FORECASTING TIMESERIES FROM STATBANK NORWAY ☺ - HACK4SSB

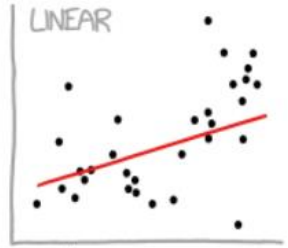
Statistisk sentralbyrå
Statistics Norway

Prophet – forecasting Statbank ☺

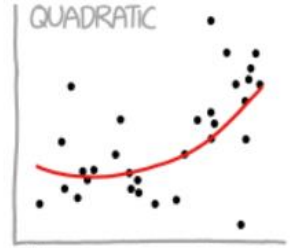
Team

- Marius Andersen, 240
- Jansen, Xiaoming Chen, 211
- Eivind Egge, 212
- Christine Kiran Kaushal, 214 / 721
- Jan Bruusgaard, 660

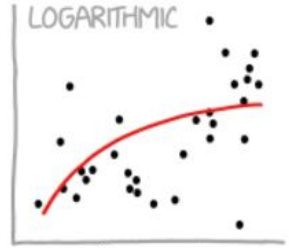
CURVE-FITTING METHODS
AND THE MESSAGES THEY SEND



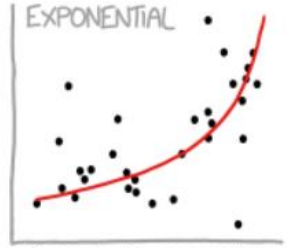
"HEY, I DID A REGRESSION."



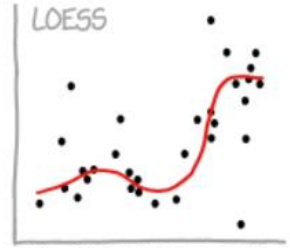
"I WANTED A CURVED LINE, SO I MADE ONE WITH MATH."



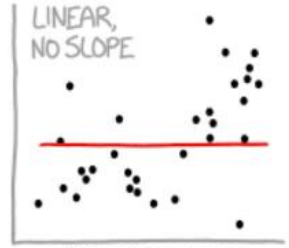
"LOOK, IT'S TAPERING OFF!"



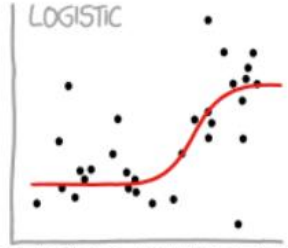
"LOOK, IT'S GROWING UNCONTROLLABLY!"



"I'M SOPHISTICATED, NOT LIKE THOSE BUMBLING POLYNOMIAL PEOPLE."

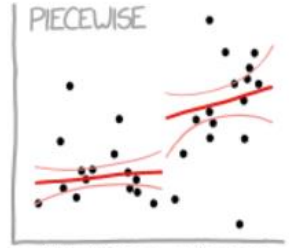


"I'M MAKING A SCATTER PLOT BUT I DON'T WANT TO."

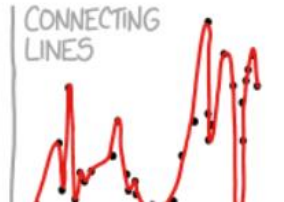


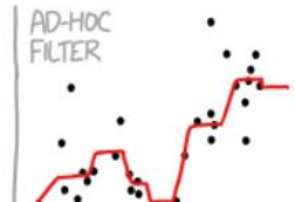
"I NEED TO CONNECT THESE TWO LINES, BUT MY FIRST IDEA DIDN'T HAVE ENOUGH MATH."

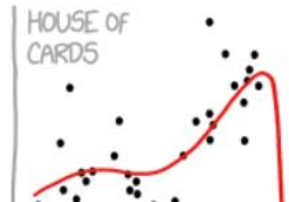
"LISTEN, SCIENCE IS HARD. BUT I'M A SERIOUS PERSON DOING MY BEST."



"I HAVE A THEORY, AND THIS IS THE ONLY DATA I COULD FIND."





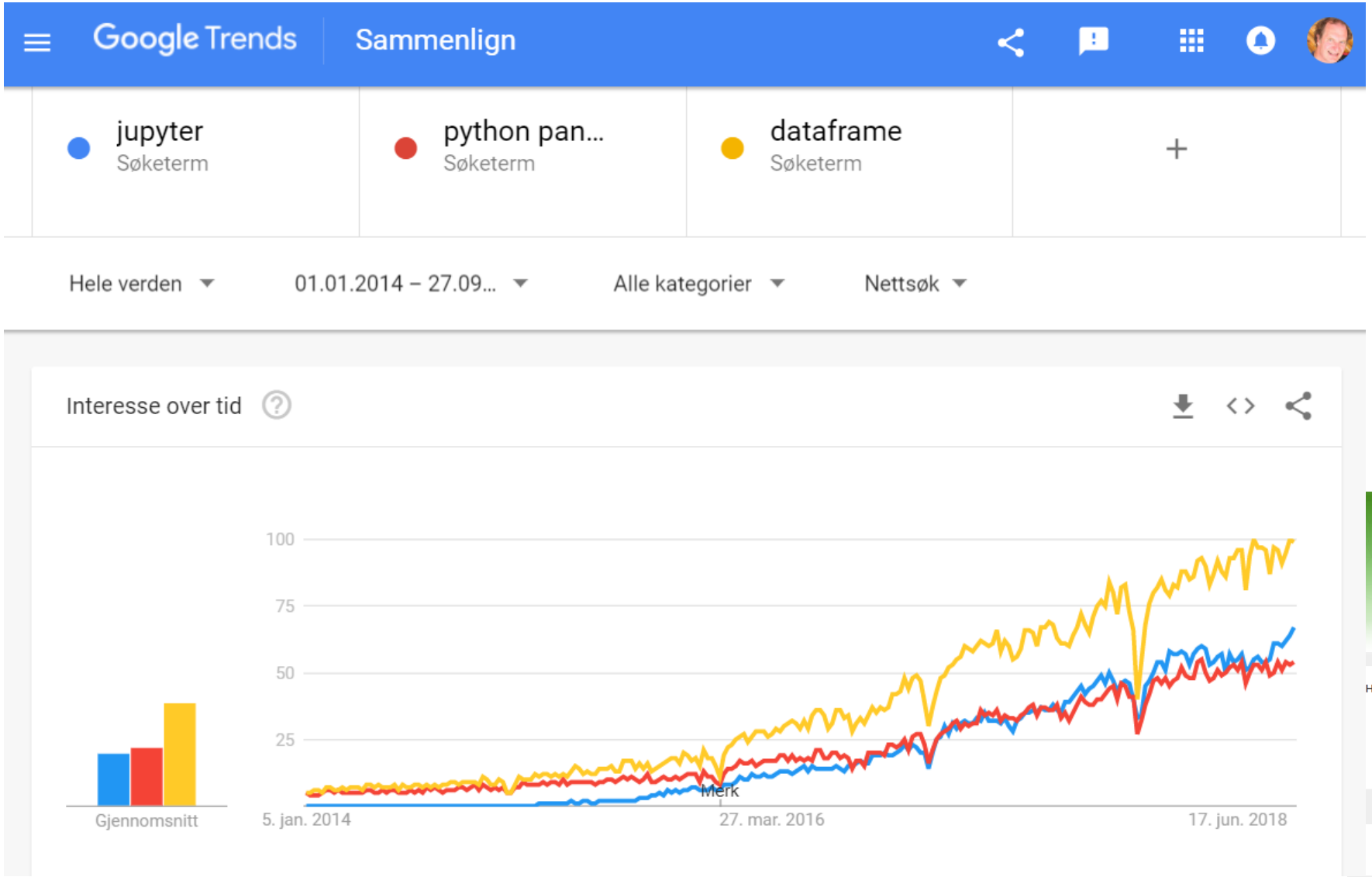


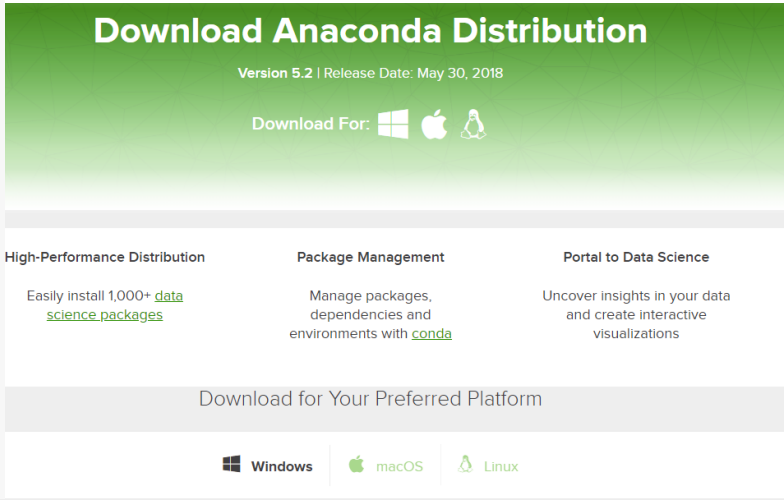
Statistisk sentralbyrå
Statistics Norway

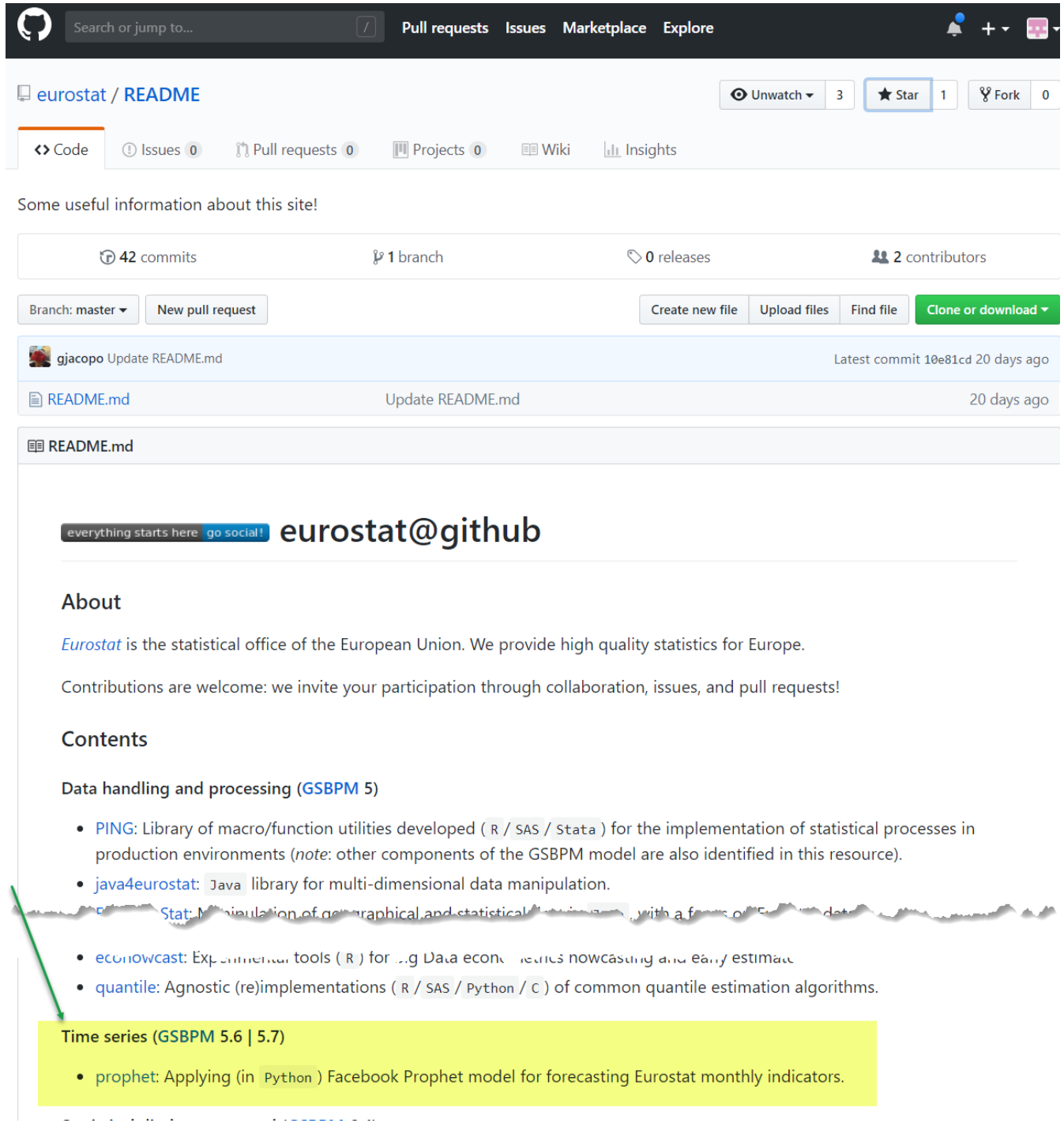
Goal

- Test Facebook Prophet forecasting model towards Statistics Norway's Statbank API
- Make a simple user interface by combining two Python packages: *fbprophet* og *stats_to_pandas*
- Learn a bit Python, Pandas and dataframes
- Experience use of Jypyter notebooks
- Result:
  Possible to make fast forecast towards all 100.000 time series in Statbank Norway

Search trends Google







Statistisk sentralbyrå
Statistics Norway

Search or jump to... / Pull requests Issues Marketplace Explore

eurostat / prophet

Watch 3 Unstar 4 Fork 1

Code Issues 0 Pull requests 0 Projects 0 Wiki Insights

Application of Facebook Prophet model (Python) for forecasting Eurostat monthly indicators https://github.com/eurostat/prophet

prophet time-series eurostat-data python

16 commits 1 branch 0 releases 1 contributor EUPL-1.1

prophet

Applying *Facebook* `Prophet` model for forecasting *Eurostat* monthly indicators

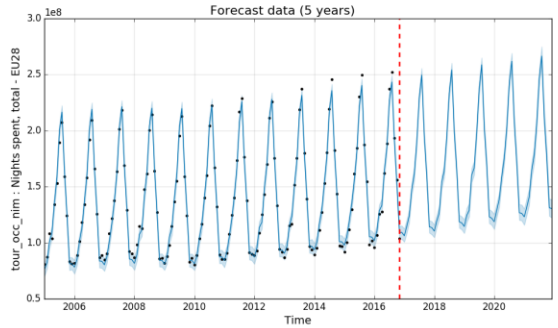
About

This is a **blind/dummy** (no assumption whatsoever) application of `Prophet` automatic procedure for forecast estimates of Eurostat *tour_occ_nim* time-series on the number of *"nights spent at tourist accommodation establishments"* per month.

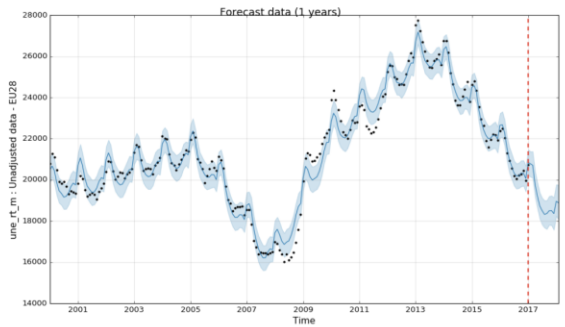
status	since 2017 – closed
contributors	
license	EUPL

Facebook has open sourced `Prophet software`, a forecasting project with an interface available in `Python`. We use this resource.

Run the `tour_forecast.py` source code or explore the `run_forecast.ipynb` notebook to produce the following 5-years forecast estimates of Eurostat *tour_occ_nim* monthly indicator:



Another example is provided by the 1-year prediction of unemployment *une_rt_m* monthly indicator:



Reference

- Taylor, S.J. and Letham, B. (2017): Forecasting at Scale.

jupyter

Statistisk sentralbyrå
Statistics Norway

Facebook Open Source

- Android
- Artificial Intelligence
- Data Infrastructure
- Developer Operations
- Development Tools
- Frontend
- iOS
- Languages
- Linux
- Security
- Virtual Reality

Artificial Intelligence

FastText
Library for fast text representation and classification.

GitHub Website

Detectron
FAIR's research platform for object detection research, implementing popular algorithms like Mask R-CNN and RetinaNet.

GitHub

PyTorch
Deep learning Framework.

GitHub Website

Prophet
Tool for producing high quality forecasts for time series data that has multiple seasonality with linear or non-linear growth.

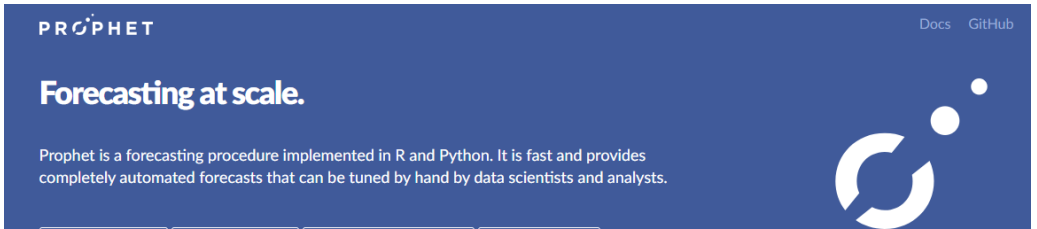
GitHub Website

Tensor Comprehensions
A domain specific language to express machine learning workloads.

GitHub Website

ONNX
The new open ecosystem for interchangeable AI models.

GitHub Website



What is Facebook prophet?

In the sphere of social media, it is important to predict the behaviour of users at particular times of a year. For this application Facebook developed an algorithm (Prophet) that can be tuned in terms of its complexity, whose uses extend beyond social media into other time series applications. The Prophet algorithm is open access and can be accessed from both the Python and R programming languages (https://facebook.github.io/prophet) and is described in detail by Taylor and Letham (2017). In brief, the algorithm works by way of an "analyst in the loop" approach, where the user can adjust parameters such as timing of regular events (e.g., holidays in business models), growth rate (e.g. linear or logistic), or the number of regressors (e.g., covariates) for an additive regressive model with four components: a piecewise linear or logistic growth curve that detects changepoints, a yearly component modeled using a Fourier series, a weekly component using dummy variables, and a user supplied list of important dates. The algorithm itself is written using 'STAN', a language commonly used by Bayesian modelers, and also includes the ability to run predictions through a Markov Chain Monte Carlo simulation. The use of the Prophet algorithm by two of the winners of our competition suggests further exploration of Prophet for ecological time series modeling and prediction is warranted, particularly in cases where rapid predictions might be useful while more mechanistically-motivated predictions are being developed.

Source:

Predicting the future is hard and other lessons from a population time series data science competition

GRW Humphries, C Che-Castaldo, PJ Bull... - Ecological ..., 2018 - Elsevier

Population forecasting, in which past dynamics are used to make predictions of future state, has many real-world applications. While time series of animal abundance are often modeled in ways that aim to capture the underlying biological processes involved, doing so is neither ...

☆ 99

Fulltext @ SSB

Taylor & Francis

The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: http://www.tandfonline.com/loi/utas20

Forecasting at Scale

Sean J. Taylor & Benjamin Letham

To cite this article: Sean J. Taylor & Benjamin Letham (2018) Forecasting at Scale, The American Statistician, 72:1, 37-45, DOI: 10.1080/00031305.2017.1380080

Google Scholar

Forecasting at scale
SJ Taylor, B Letham - The American Statistician, 2018 - amstat.tandfonline.com
Page 1. **Forecasting** at Scale Sean J. Taylor **Facebook** and Benjamin Letham **Facebook** August 21, 2017 Abstract **Forecasting** is a common data science task that helps organizations with capacity planning, goal setting, and anomaly detection …
☆ 99 Sitert av 30 Beslektede artikler Alle 7 versjoner Web of Science: 1

[PDF] tandfonline.com
Fulltext @ SSB

[PDF] News and Consumer card payments
G Ardizzi, S Emiliozzi, J Marcucci, L Monteforte - 2018 - site.stanford.edu
We exploit a unique daily data set on debit cards' expenditure consumption to news related to economic policy uncertainty (security. Adopting big data techniques, we construct indexes
☆ 99 Alle 2 versjoner »

Predicting the future is hard and other lessons from a population time series data science competition Fulltext @ SSB
GRW Humphries, C Che-Castaldo, PJ Bull… - Ecological …, 2018 - Elsevier
Population forecasting, in which past dynamics are used to make predictions of future state, has many real-world applications. While time series of animal abundance are often modeled in ways that aim to capture the underlying biological processes involved, doing so is neither …
☆ 99

Large Scale Automated Forecasting for Monitoring Network Safety and Security [PDF] arxiv.org
R Naveiro, S Rodríguez, DR Insua - arXiv preprint arXiv:1802.06678, 2018 - arxiv.org
Real time large scale streaming data pose major challenges to forecasting, in particular defying the presence of human experts to perform the corresponding analysis. We present here a class of models and methods used to develop an automated, scalable and versatile …
☆ 99 Beslektede artikler Alle 3 versjoner »

Predicting the future is hard and other lessons from a population time series data science competition Fulltext @ SSB
GRW Humphries, C Che-Castaldo, PJ Bull… - Ecological …, 2018 - Elsevier
Population forecasting, in which past dynamics are used to make predictions of future state, has many real-world applications. While time series of animal abundance are often modeled in ways that aim to capture the underlying biological processes involved, doing so is neither …
☆ 99

Statistisk sentralbyrå
Statistics Norway

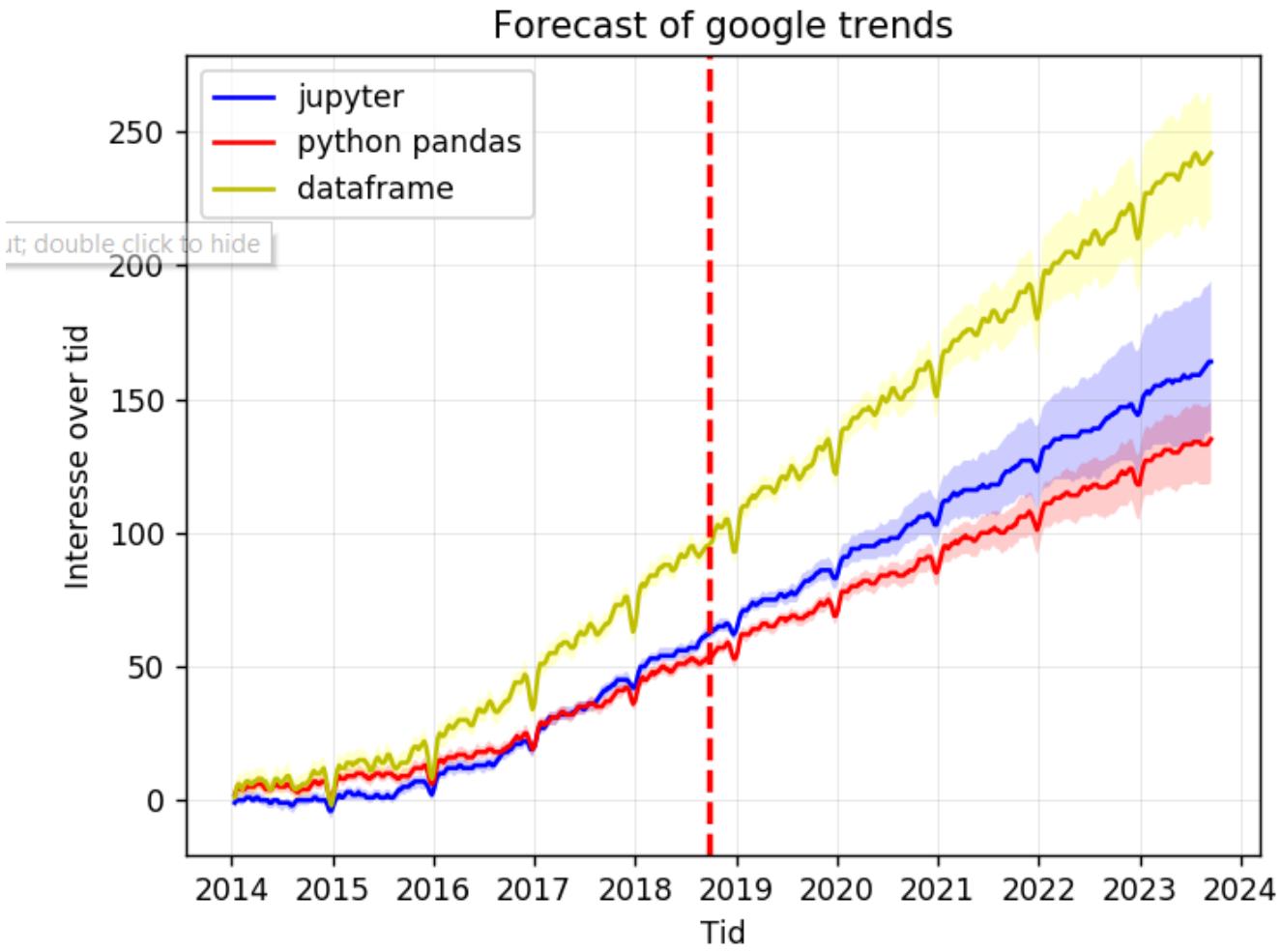
Forecasting *at scale*

1. Forecasting for most people – the prophet model is easy to use
2. Handles varied input data and allows easy customization of model
3. Forecasts at scale – evaluation at scale

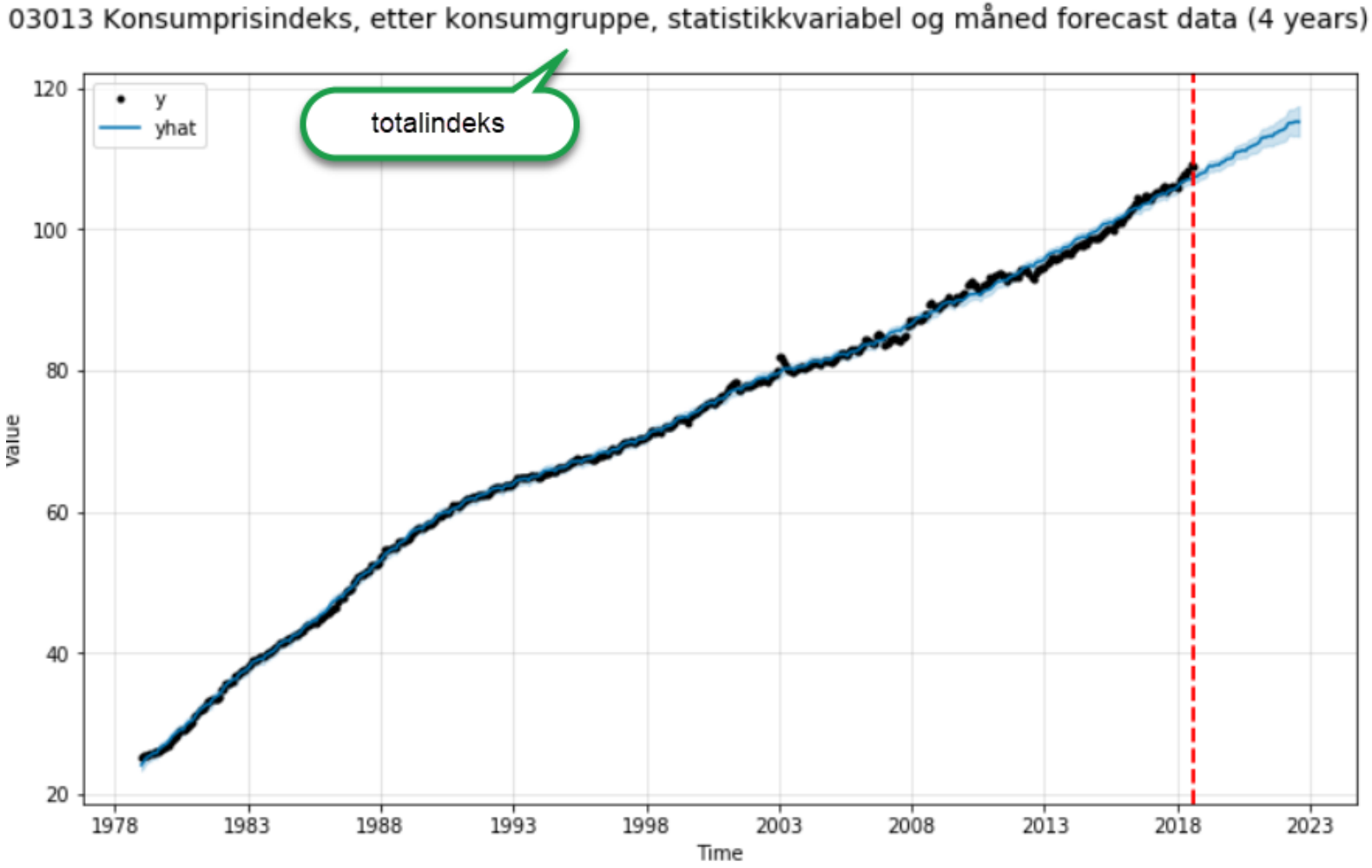
What did we test? Prophet's base is daily data

- Weekly timeseries: Google trends, salmon statstistics
- Monthly timeseries: CPI, credit indicator, foreign trade, regional tax income, GDP
- Quarterly timeseries : GDP
- Yearly timeseries : GDP, population

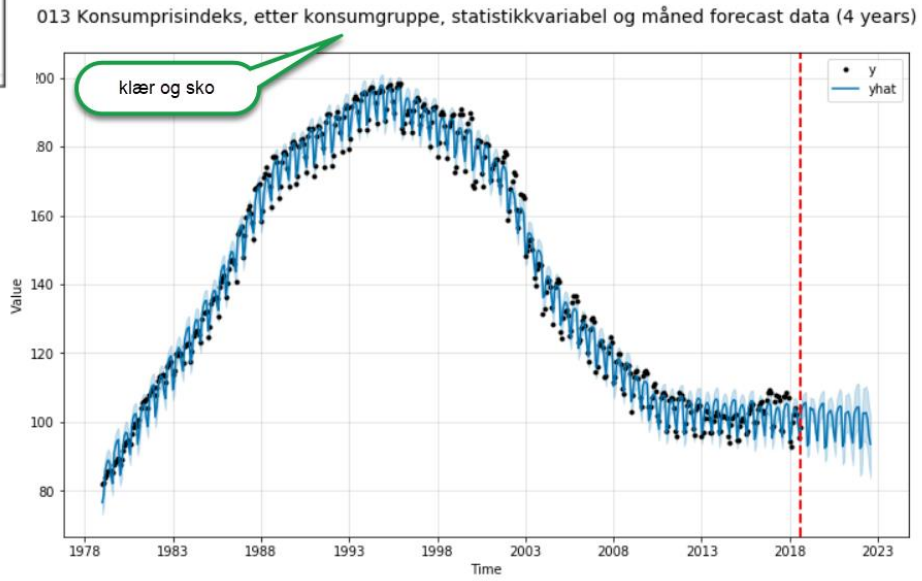
Google trends plot, Prophet 5 years forecast



Statistisk sentralbyrå
Statistics Norway

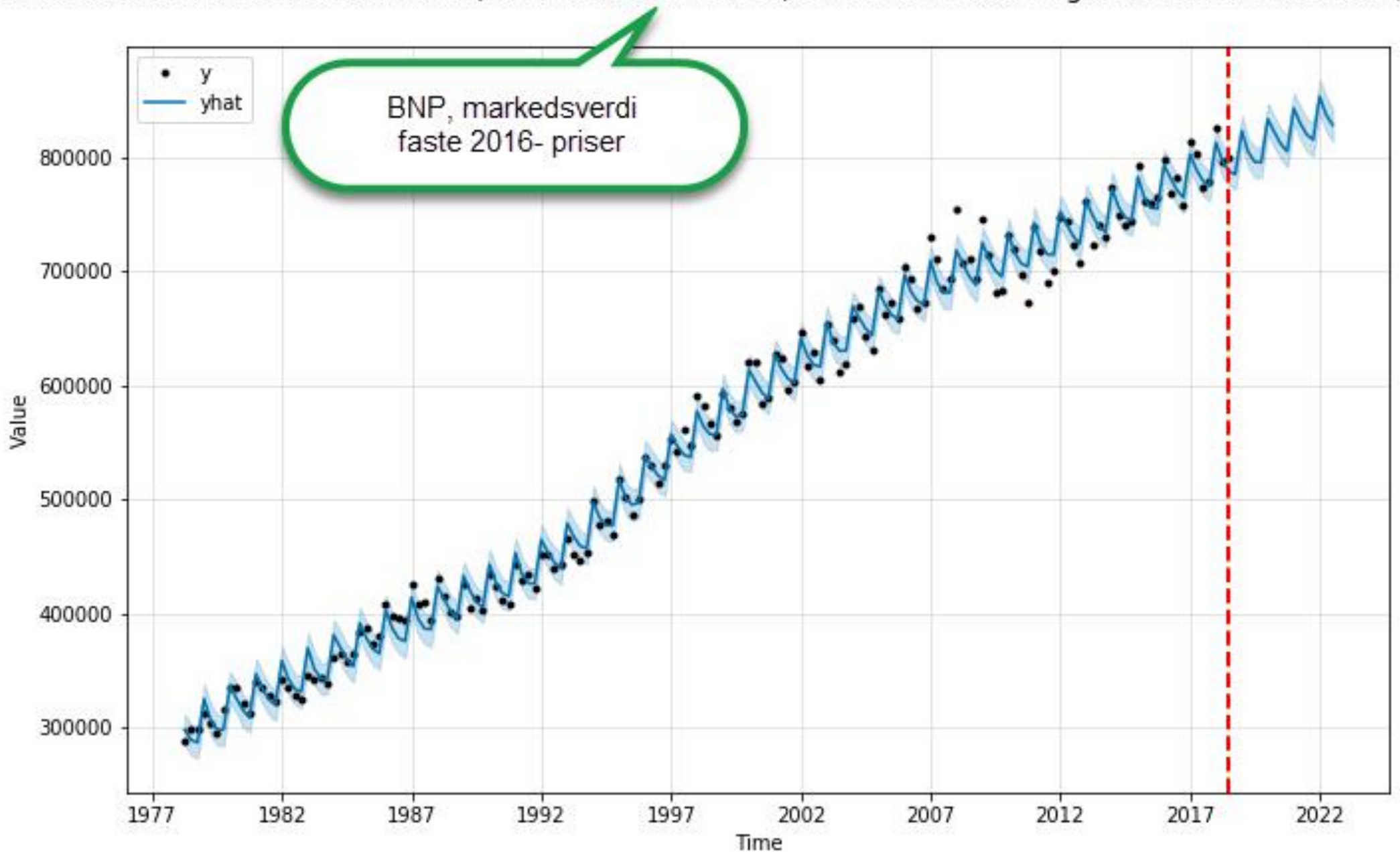

03013 Konsumprisindeks, etter konsumgruppe, statistikkvariabel og måned forecast data (4 years)
totalindeks
y
yhat
Value
Time
120
100
80
60
40
20
1978
1983
1988
1993
1998
2003
2008
2013
2018
2023

CPI

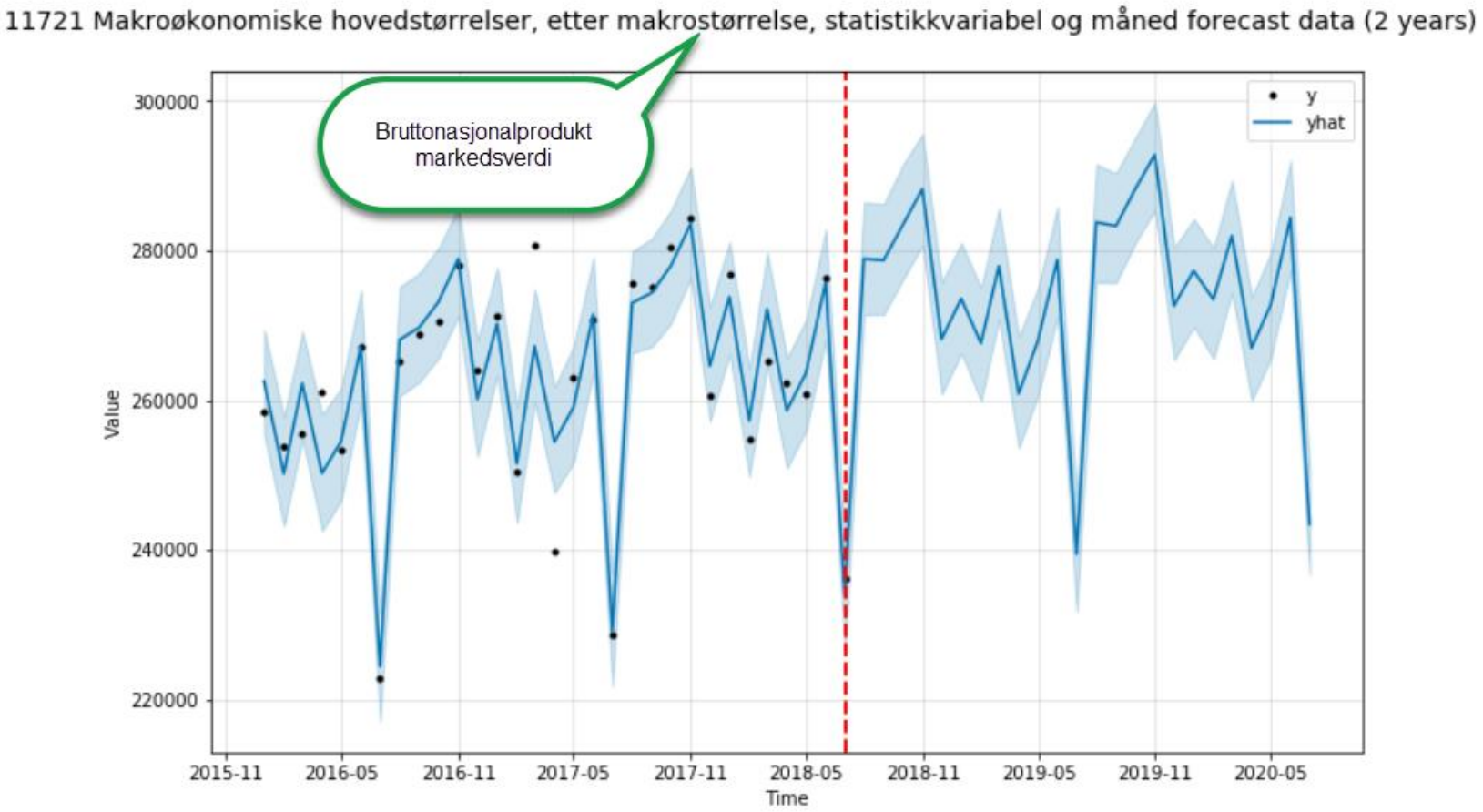

013 Konsumprisindeks, etter konsumgruppe, statistikkvariabel og måned forecast data (4 years)
klær og sko
y
yhat
Value
Time
160
140
120
100
80
1978
1983
1988
1993
1998
2003
2008
2013
2018
2023

09190 Makroøkonomiske hovedstørrelser, etter makrostørrelse, statistikkvariabel og kvartal forecast data (4 years)

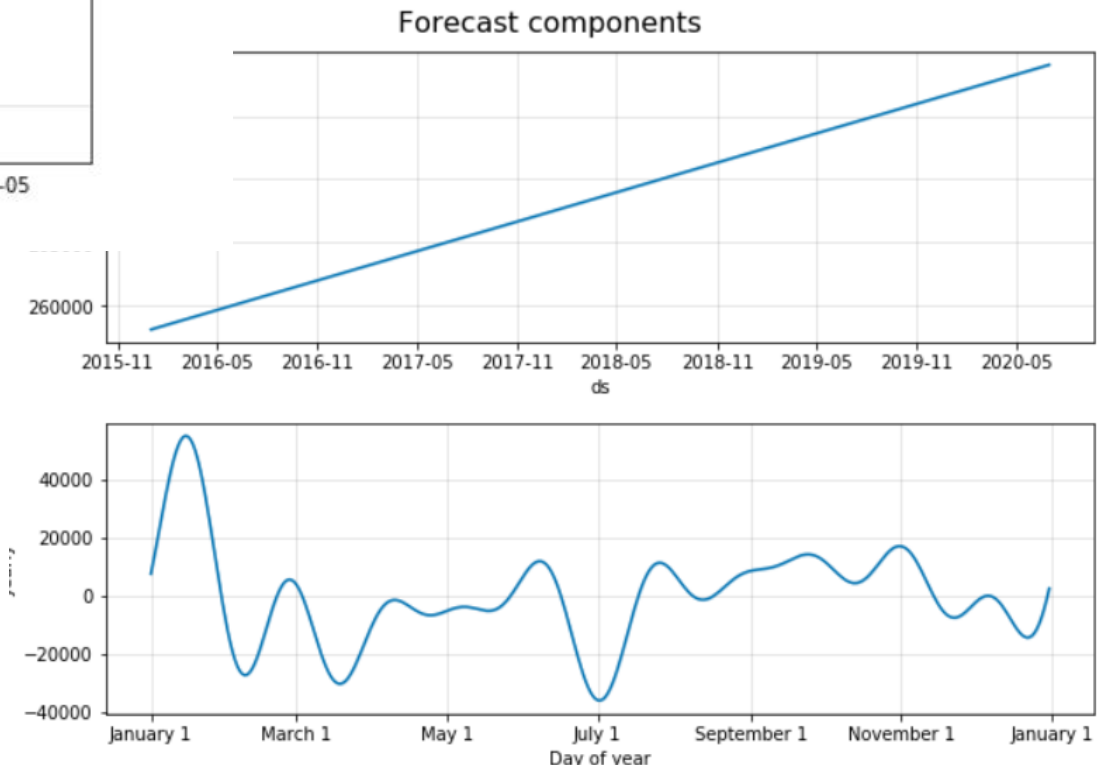
GDP



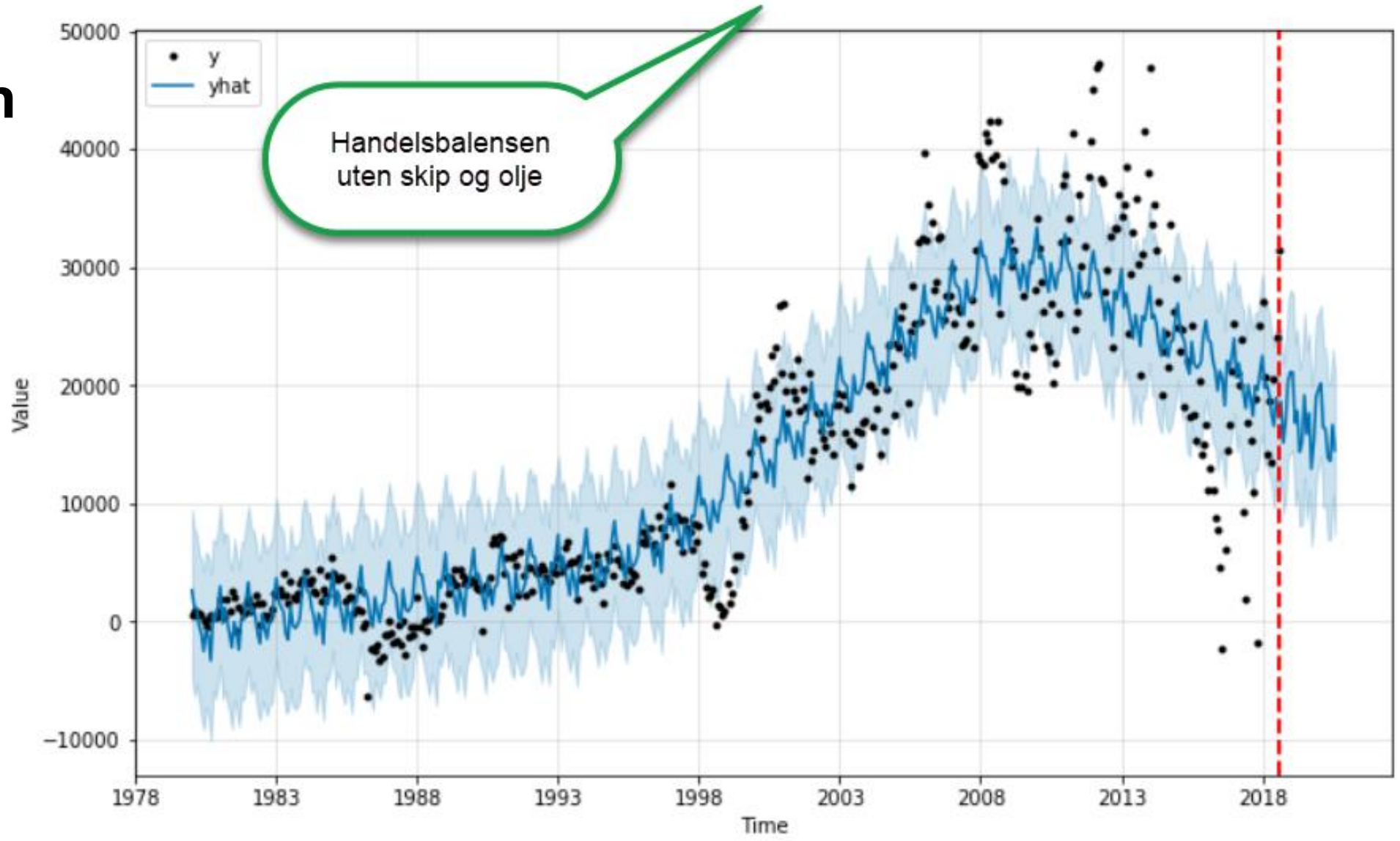
GDP



	ds	yhat	yhat_lower	yhat_upper
50	2020-02-29	281993.367207	274394.741163	289397.493044
51	2020-03-31	266983.008689	259946.423086	273812.569863
52	2020-04-30	272744.977169	265610.818552	279742.131272
53	2020-05-31	284432.692429	277421.952231	292036.552845
54	2020-06-30	243465.805639	236750.900723	250966.361777

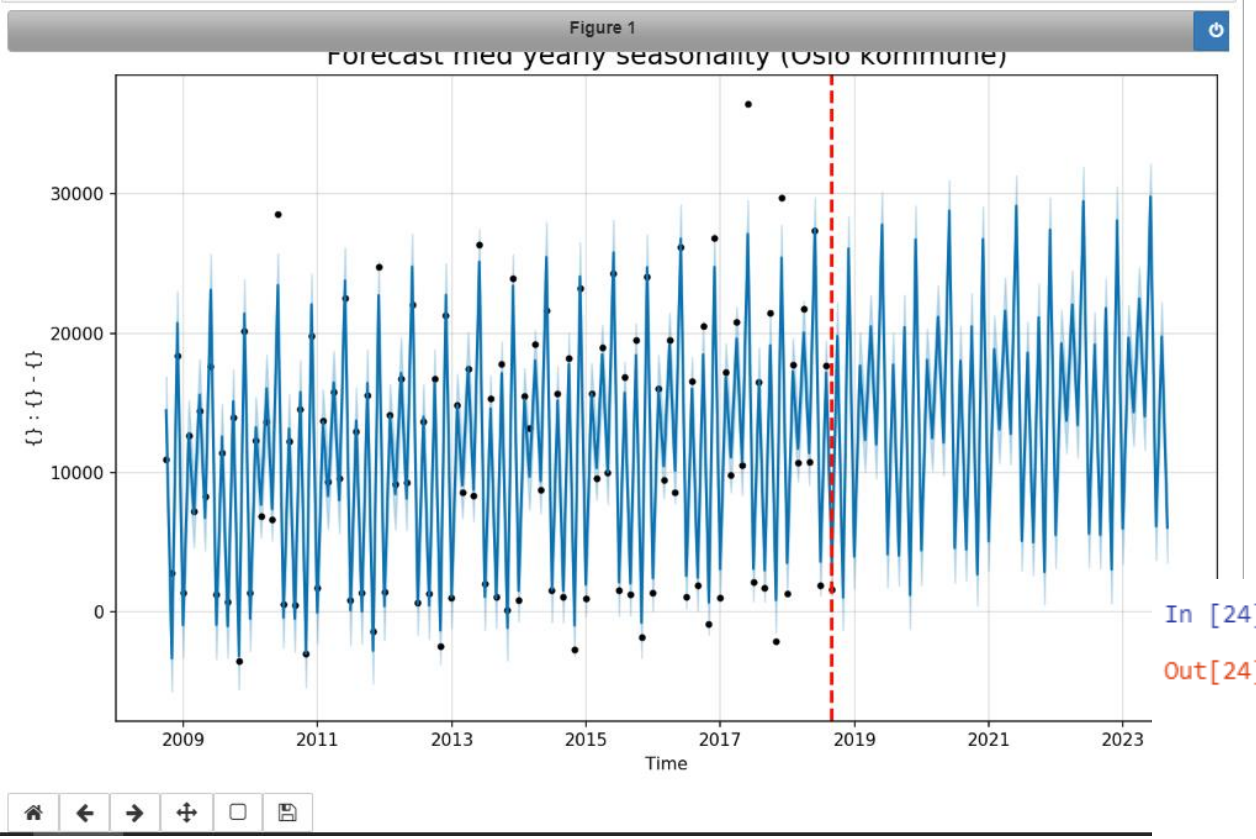


Foreign trade



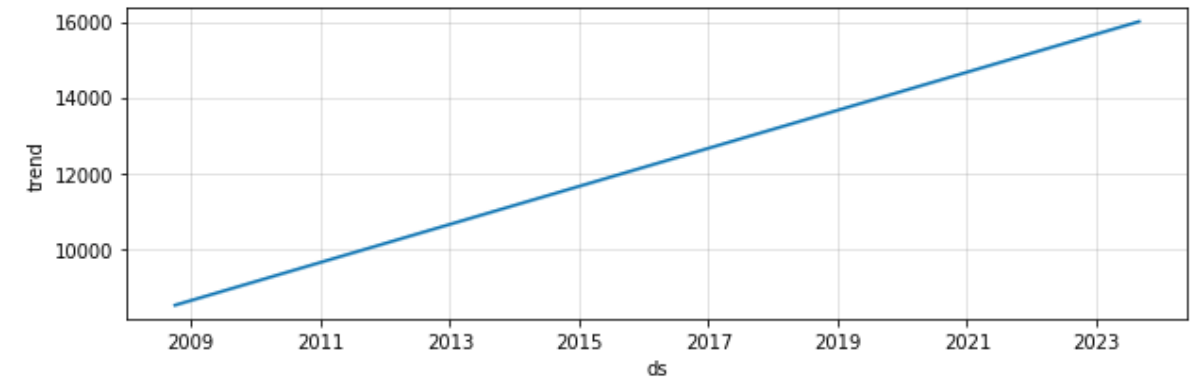
Statistisk sentralbyrå
Statistics Norway

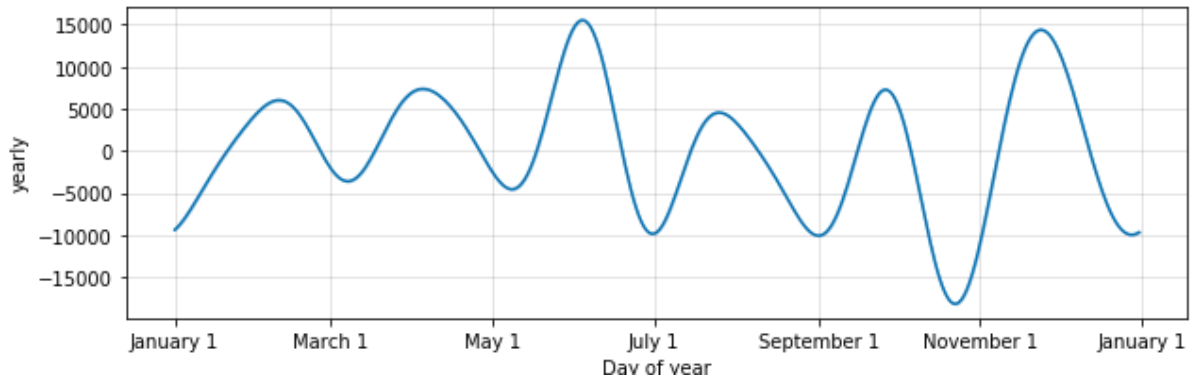
Tax income



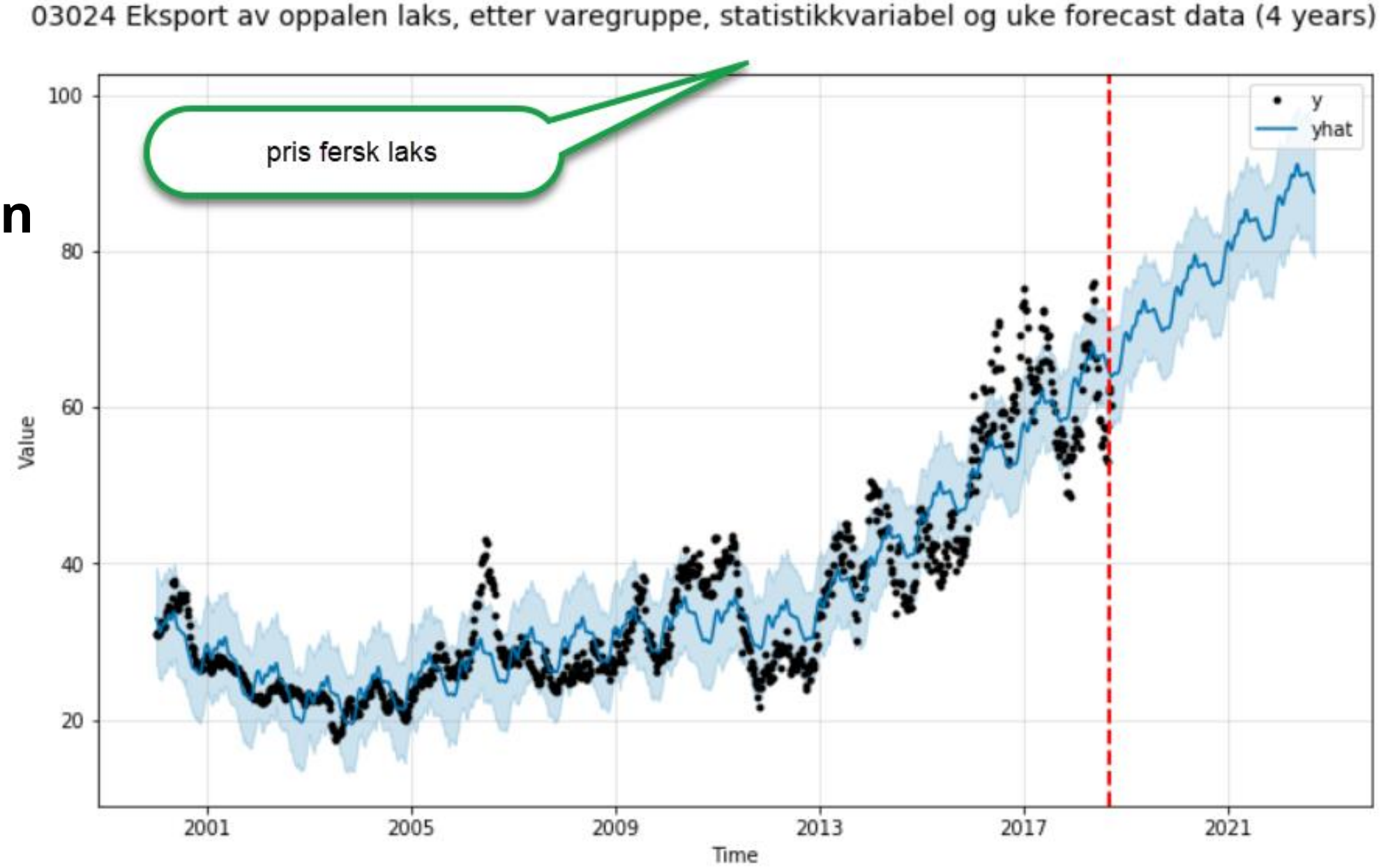
In [24]:
```
m.plot_components(fcst)
```

Out[24]:



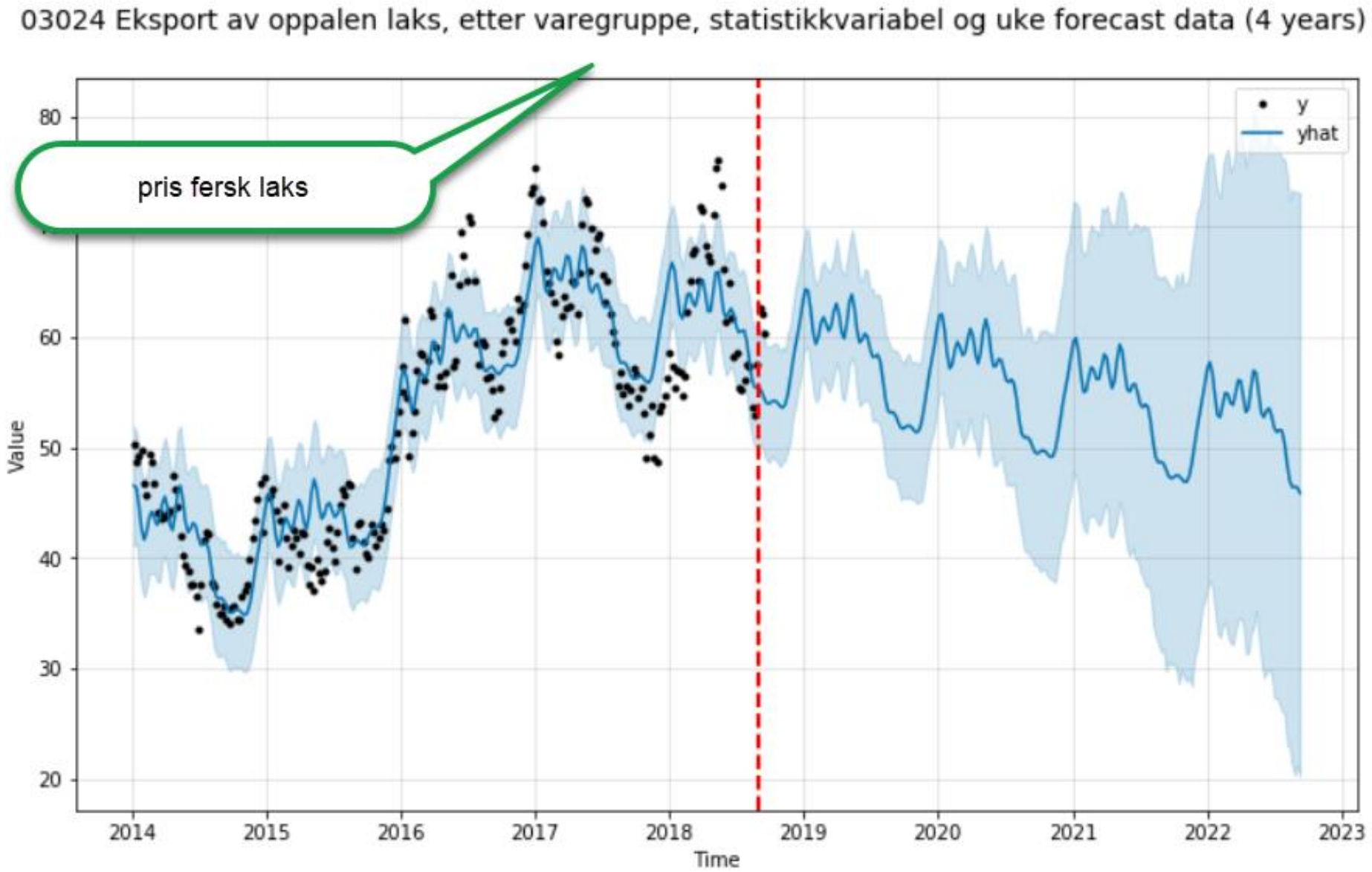


Fresh salmon price



Statistisk sentralbyrå
Statistics Norway

Fresh salmon price



Statistisk sentralbyrå
Statistics Norway

Demo

In [32]:
```
from fbprophet.diagnostics import cross_validation
df_cv5y_1y = cross_validation(m, initial='1825 days', period='365 days', horizon = '365 days')
df_cv5y_1y.head()
```

```
INFO:fbprophet.diagnostics:Making 34 forecasts with cutoffs between 1984-08-09 00:00:00 and 2017-08-01 00:00:00
```

Out[32]:

	ds	yhat	yhat_lower	yhat_upper	y	cutoff
0	1984-09-01	42.384634	42.297872	42.464116	42.3	1984-08-09
1	1984-10-01	42.461943	42.370141	42.554605	42.5	1984-08-09
2	1984-11-01	42.508974	42.407537	42.625729	42.6	1984-08-09
3	1984-12-01	42.523835	42.402093	42.665194	42.8	1984-08-09
4	1985-01-01	43.434724	43.287077	43.612337	43.1	1984-08-09

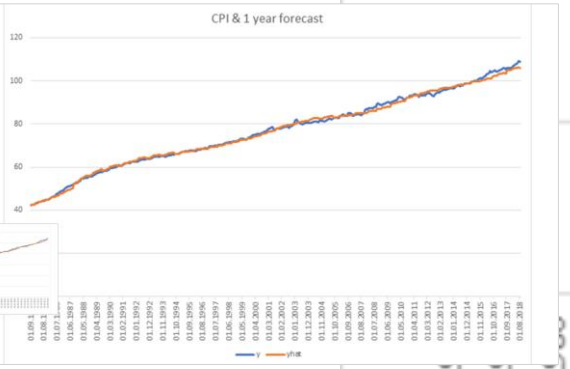
In [33]:
```
df_cv5y_1y[['ds','yhat','yhat_lower','yhat_upper','y','cutoff']].to_csv("C:/Users/anm/Documents/Python Scripts/kpi_cv5y_
```

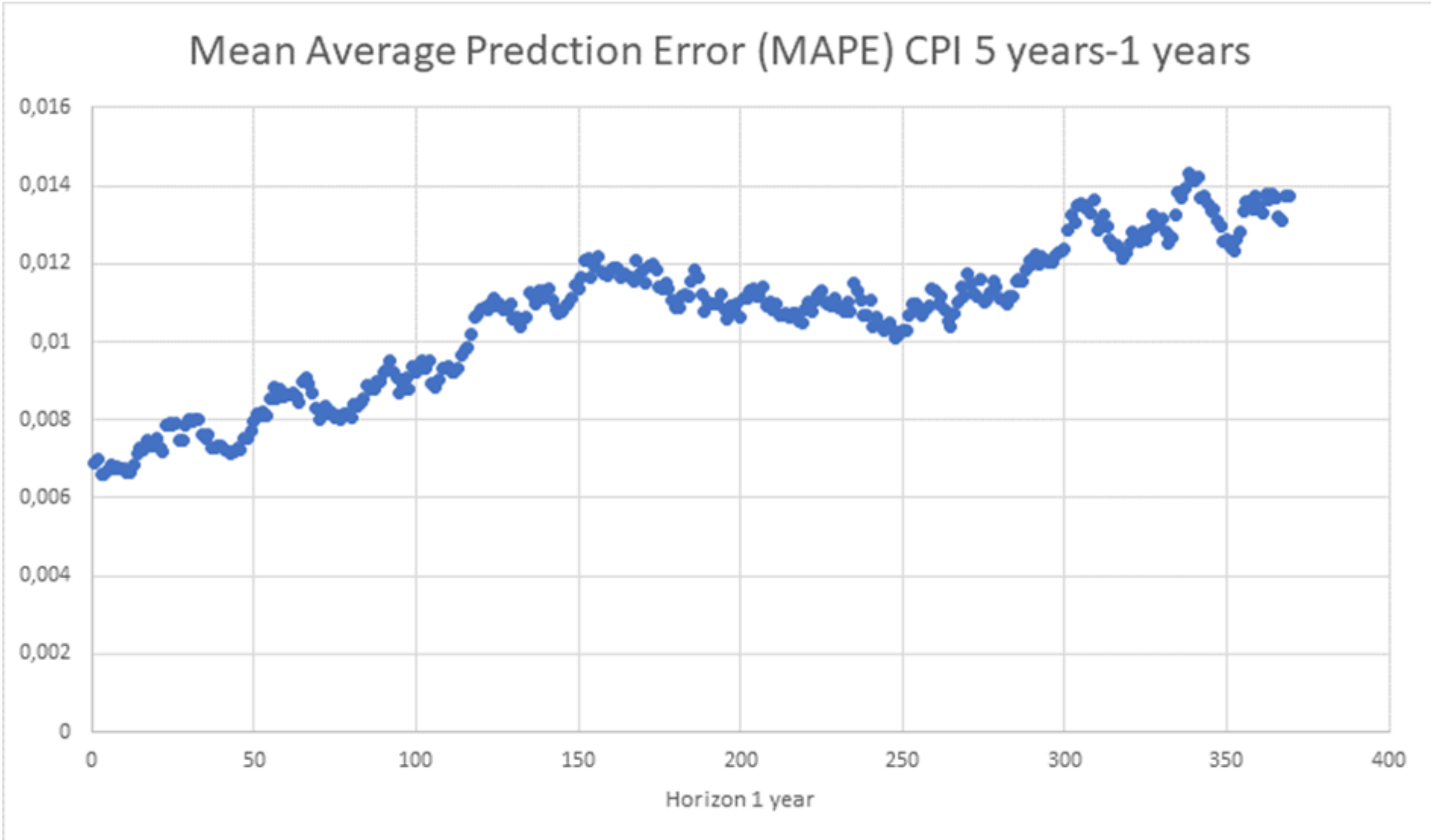
In [34]:
```
from fbprophet.diagnostics import performance_metrics
df_p5y_1y = performance_metrics(df_cv5y_1y)
df_p5y_1y.head()
```

Out[34]:

	horizon	mse	rmse	mae	mape	coverage
73	54 days	0.607188	0.779223	0.532254	0.006903	0.400
49	54 days	0.609498	0.780704	0.538027	0.006995	0.375
61	54 days	0.585702	0.765312	0.521239	0.006618	0.375
133	55 days	0.585922	0.765455	0.521964	0.006603	0.400
109	55 days	0.588183	0.766931	0.527571	0.006680	0.375

In [35]:
```
df_p5y_1y[['horizon','mse','rmse','mae','mape','coverage']].to_csv("C:/Users/anm/Documents/Python Scripts/kpi_perf5y_1y.
```


CPI & 1 year forecast
120
100
80
60
40
y
yhat


Mean Average Predction Error (MAPE) CPI 5 years-1 years
0,016
0,014
0,012
0,01
0,008
0,006
0,004
0,002
0
0
50
100
150
200
250
300
350
400
Horizon 1 year

Prophet + SSB's API-users

- Prophet + robot journalism (Kommunal rapport, NTB)
- Prophet + our competitors (Ny analyse, Oslo economics, Prognosesenteret)
- Prophet + bank/finance (Sparebank1, Swedbank)

Experiences so far

- Established a simple routine for making forecasts on a freely chosen timeseries in Statbank, and other NSI using PxWebApi.
- Prophet is fast, easy to use and easy to change parameters
- Prophet on single time series has to customizes for a good result
- Can be a tool for revision in the production of statistics

TALL SOM FORTELLER

Takk!

ssb.no

Statistisk sentralbyrå
Statistics Norway