ENERGY DATA SCIENCE

Time series analysis: Forecasting

Prof. Juri Belikov

Department of Software Science
Tallinn University of Technology
juri.belikov@taltech.ee

PREVIOUSLY IN COURSE …

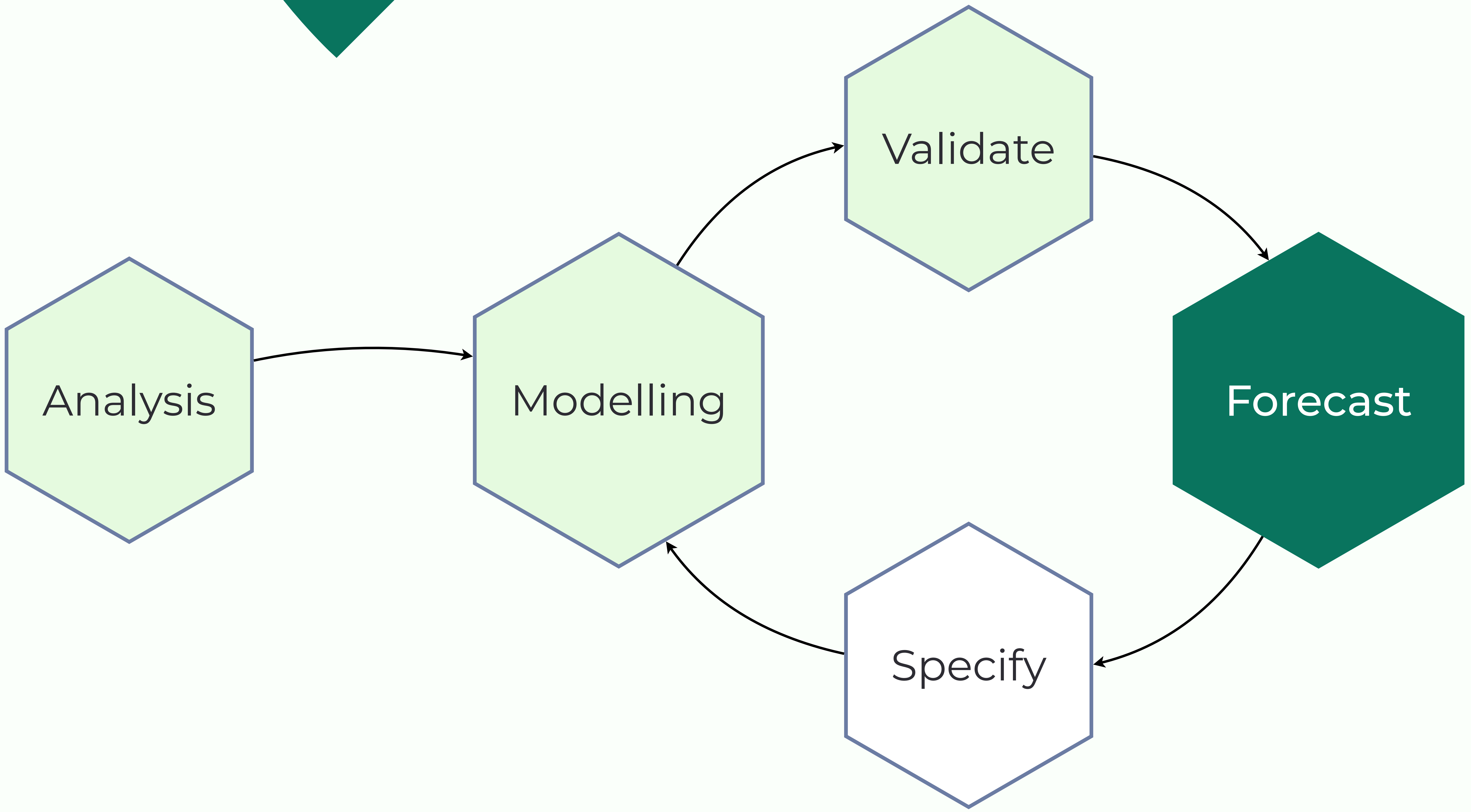
Key takeaways:

- ML and Time series
- RNN, LSTM, LightGBM

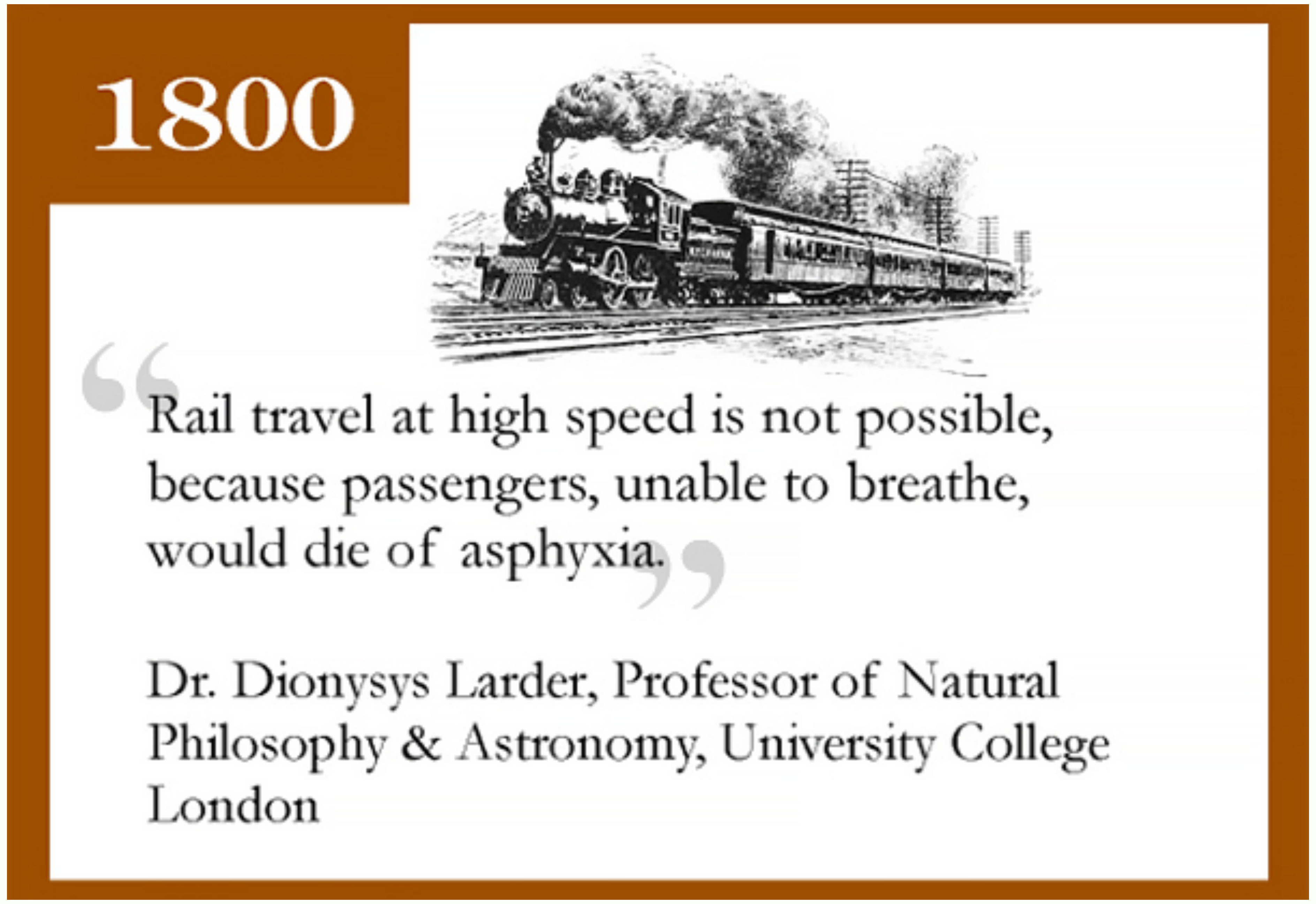
Train ML models, evaluate performance, and make predictions.

06

MODELLING

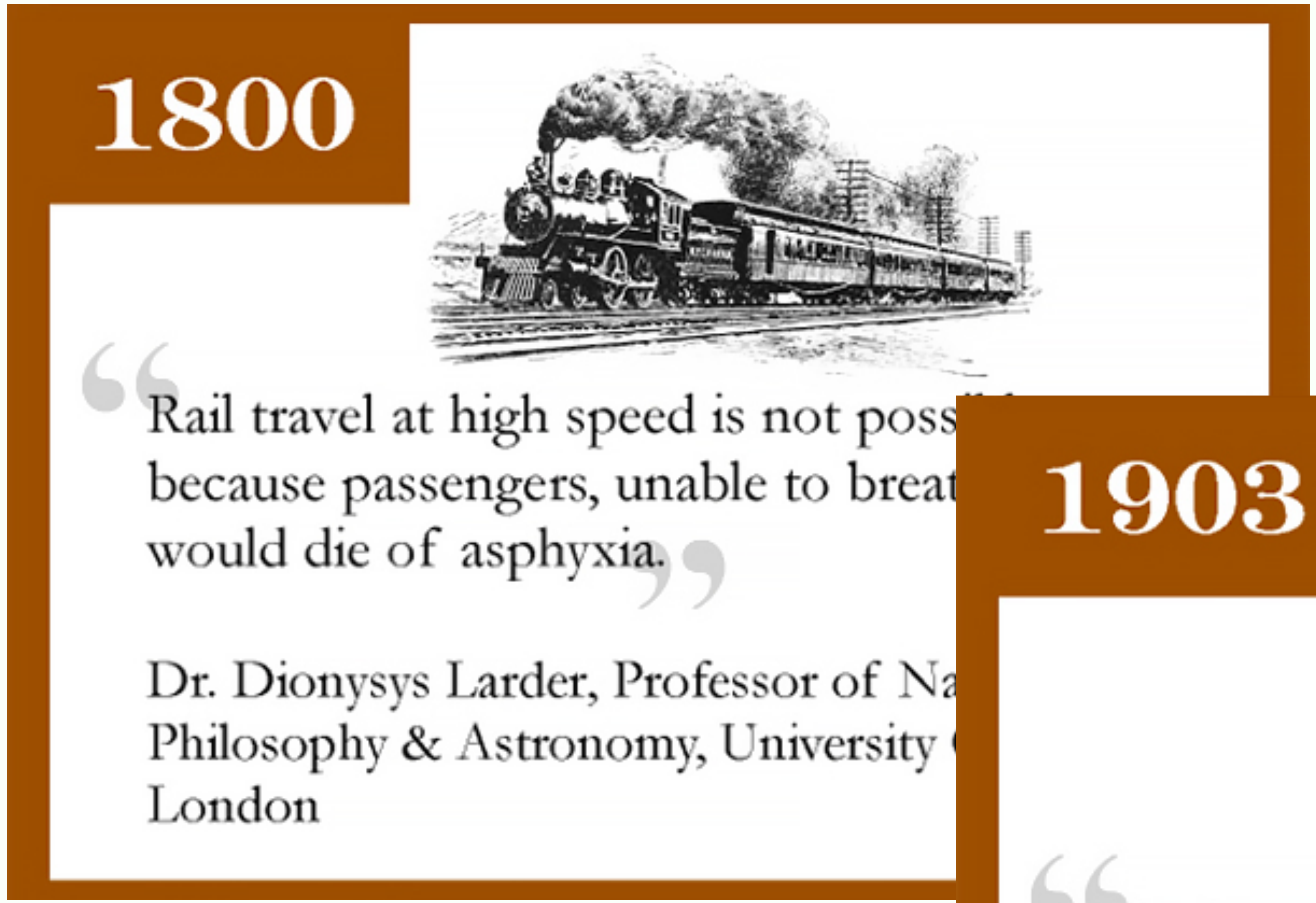


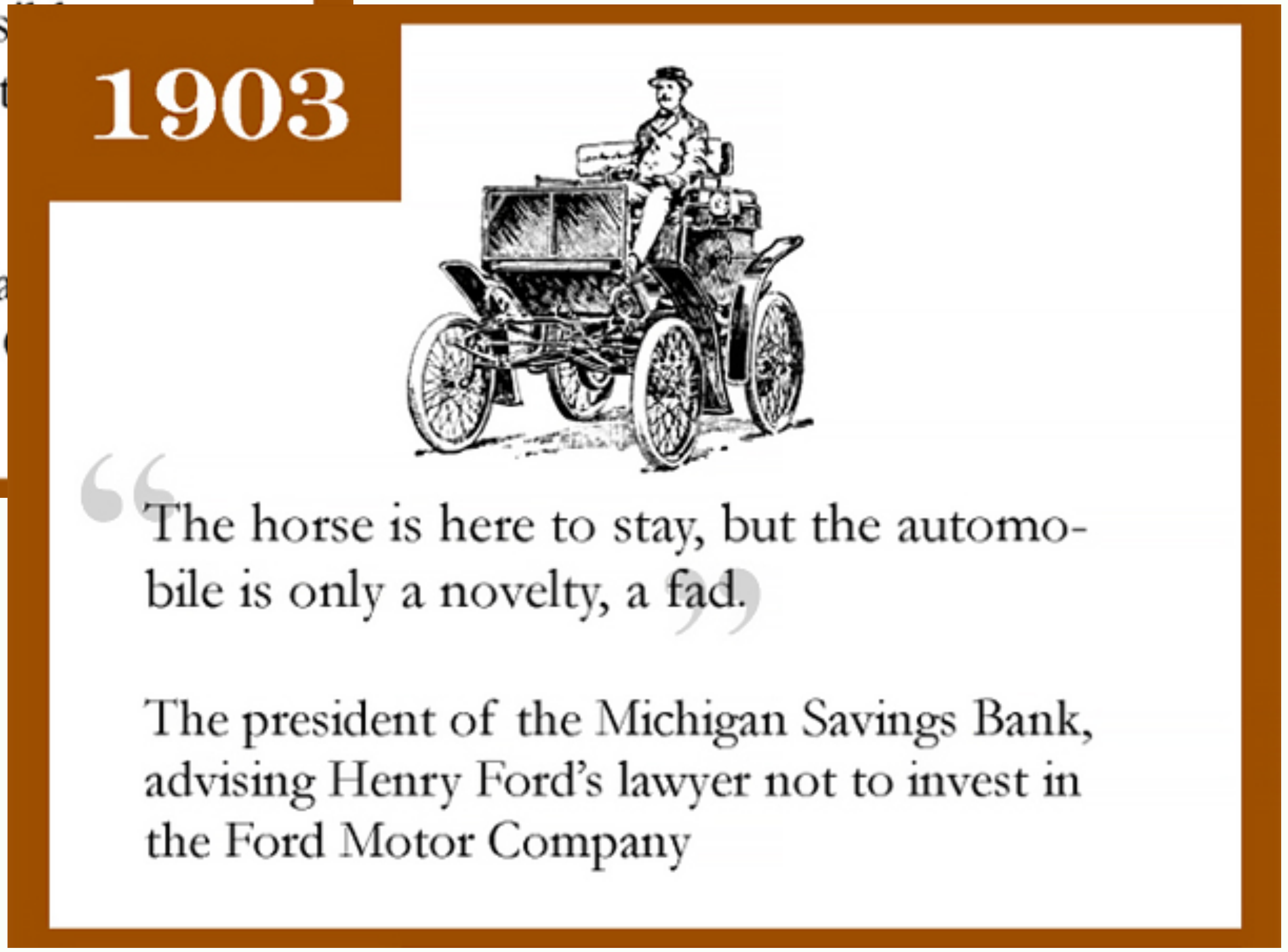
PREDICTING IS DIFFICULT?!



[Accessed June 6, 2024], https://ritholtz.com/2018/02/timeline-bad-future-predictions/

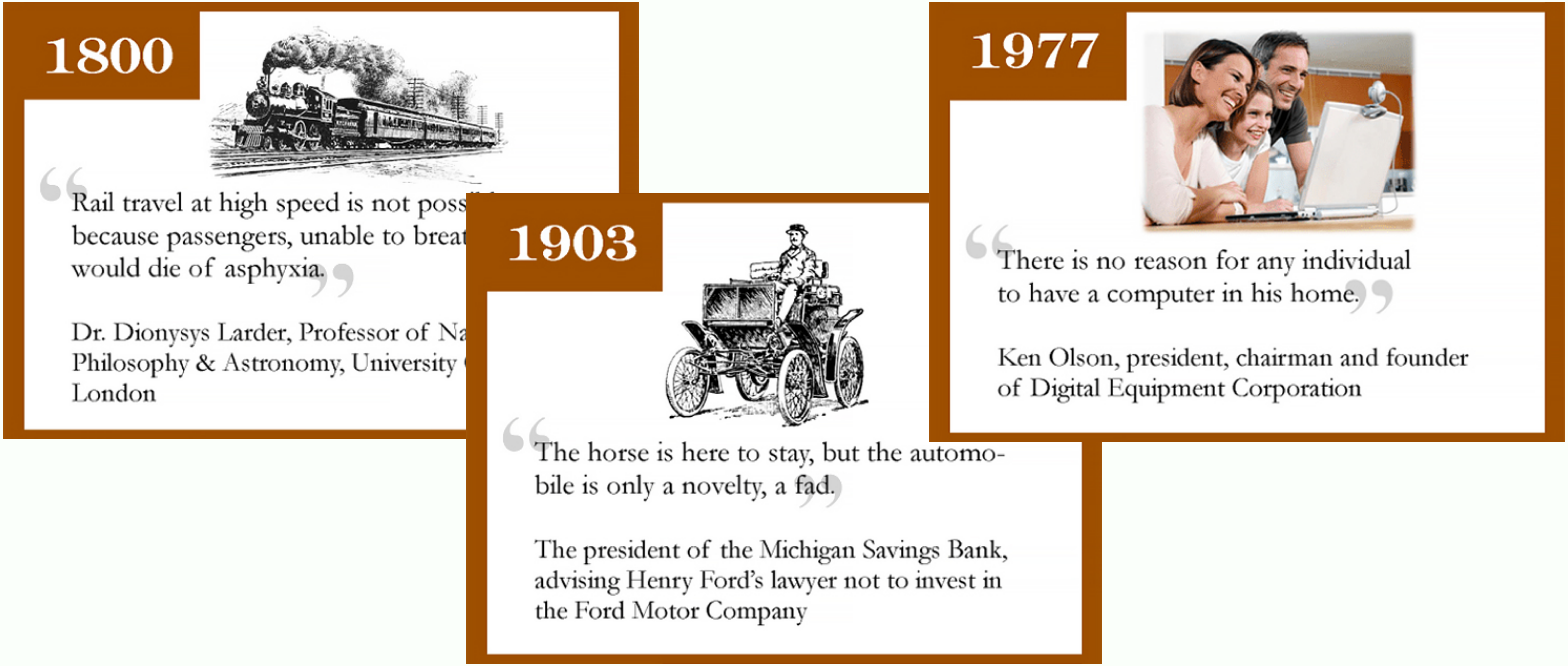
PREDICTING IS DIFFICULT?!





[Accessed June 6, 2024], https://ritholtz.com/2018/02/timeline-bad-future-predictions/

PREDICTING IS DIFFICULT?!



[Accessed June 6, 2024], https://ritholtz.com/2018/02/timeline-bad-future-predictions/

FORECASTING VS PREDICTING

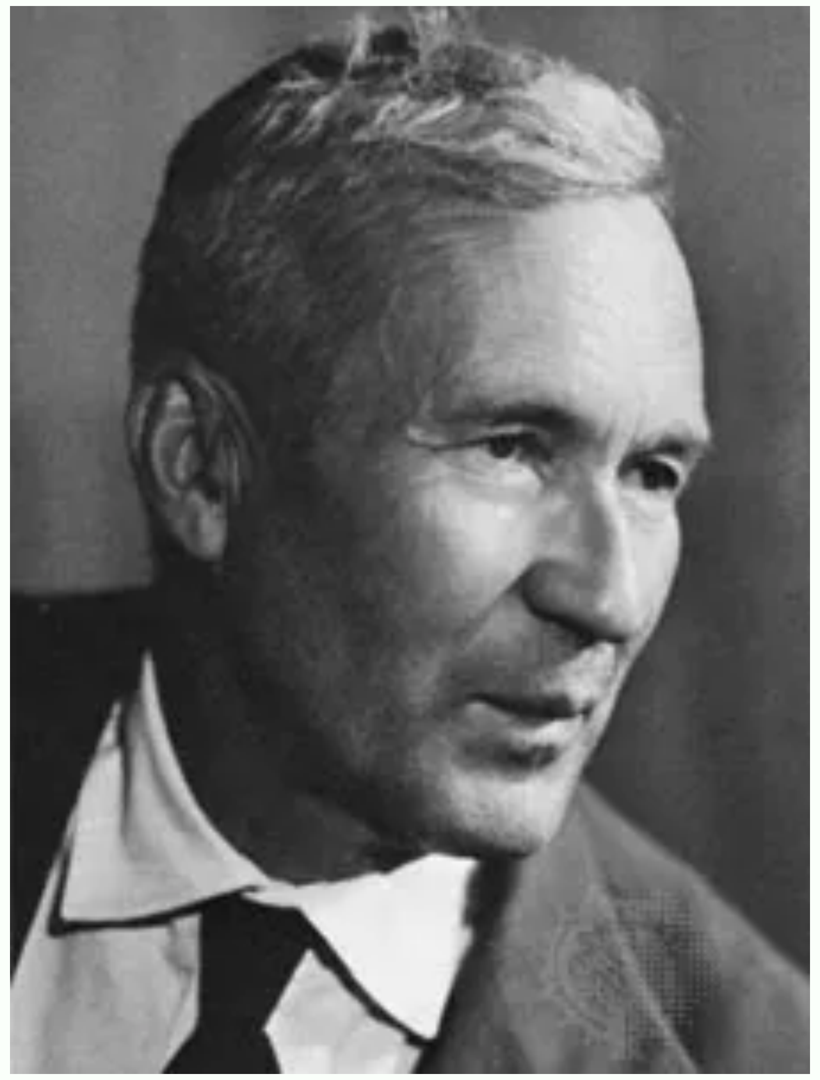
Forecast:

- ✓ systematic
- ✓ methodologically rigorous
- ✓ quantitative
- ✓ formal

Prediction:

- ✓ broad
- ✓ potentially subjective
- ✓ qualitative or quantitative
- ✓ less formal

THE ORIGINS

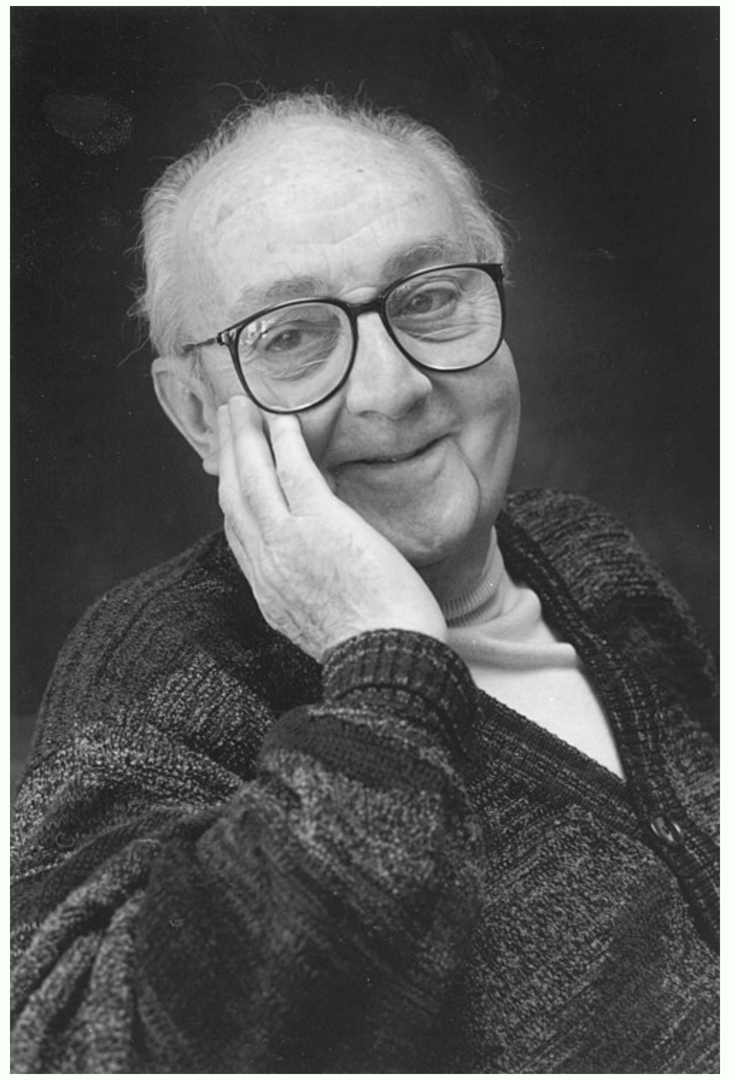


A. N. Kolmogorov

(1903-1987)

N. Wiener

(1894-1964)



G. E.P. Box

(1919-2013)

FORECASTING METHOD

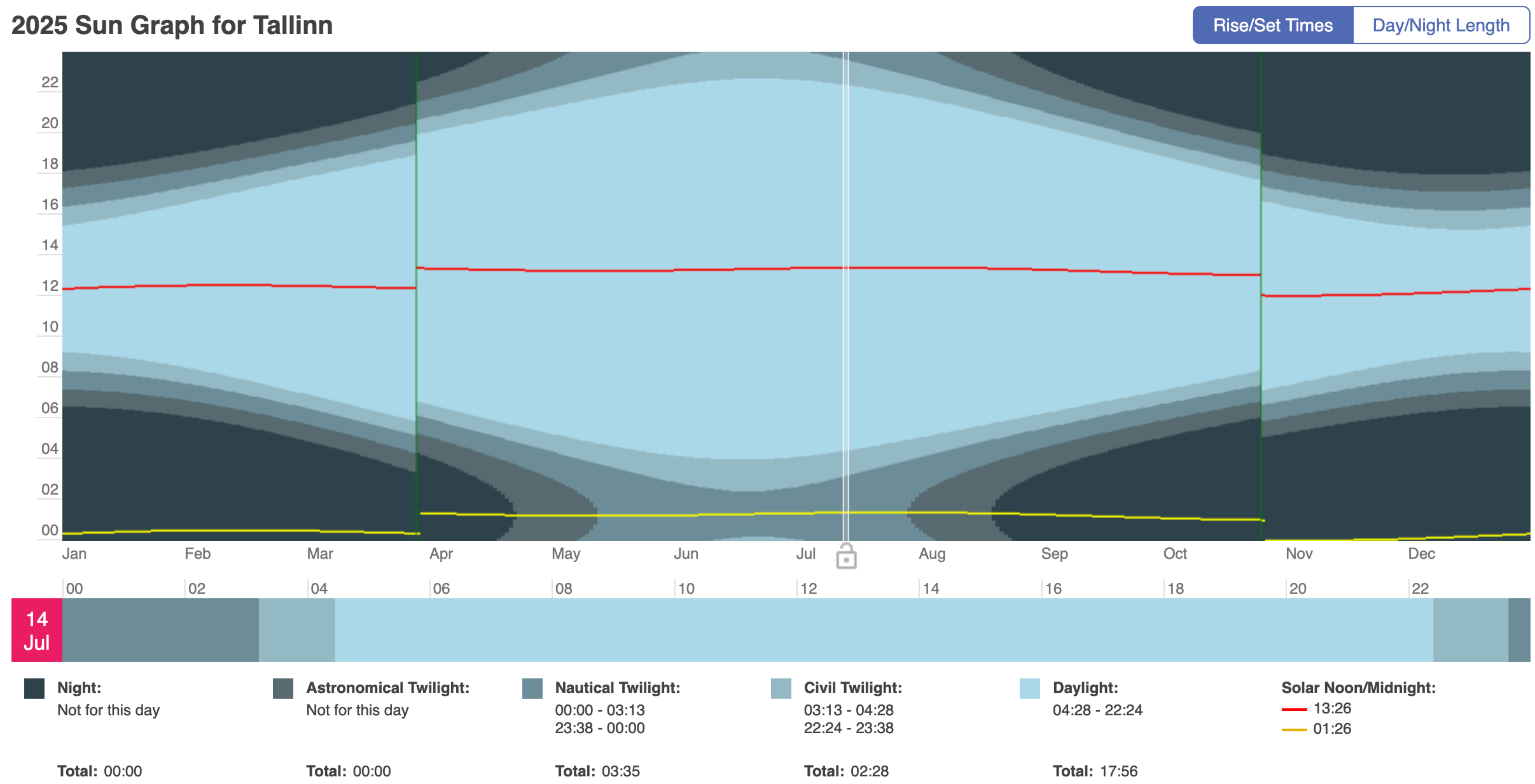
Qualitative forecasting methods are used when there are no numeric data available, or if the data available are not relevant to the forecasts.

Quantitative forecasting can be applied when two conditions are satisfied:

1. numerical information about the past is available;
2. it is reasonable to assume that some aspects of the past patterns will continue into the future.

WHAT CAN BE FORECASTED?

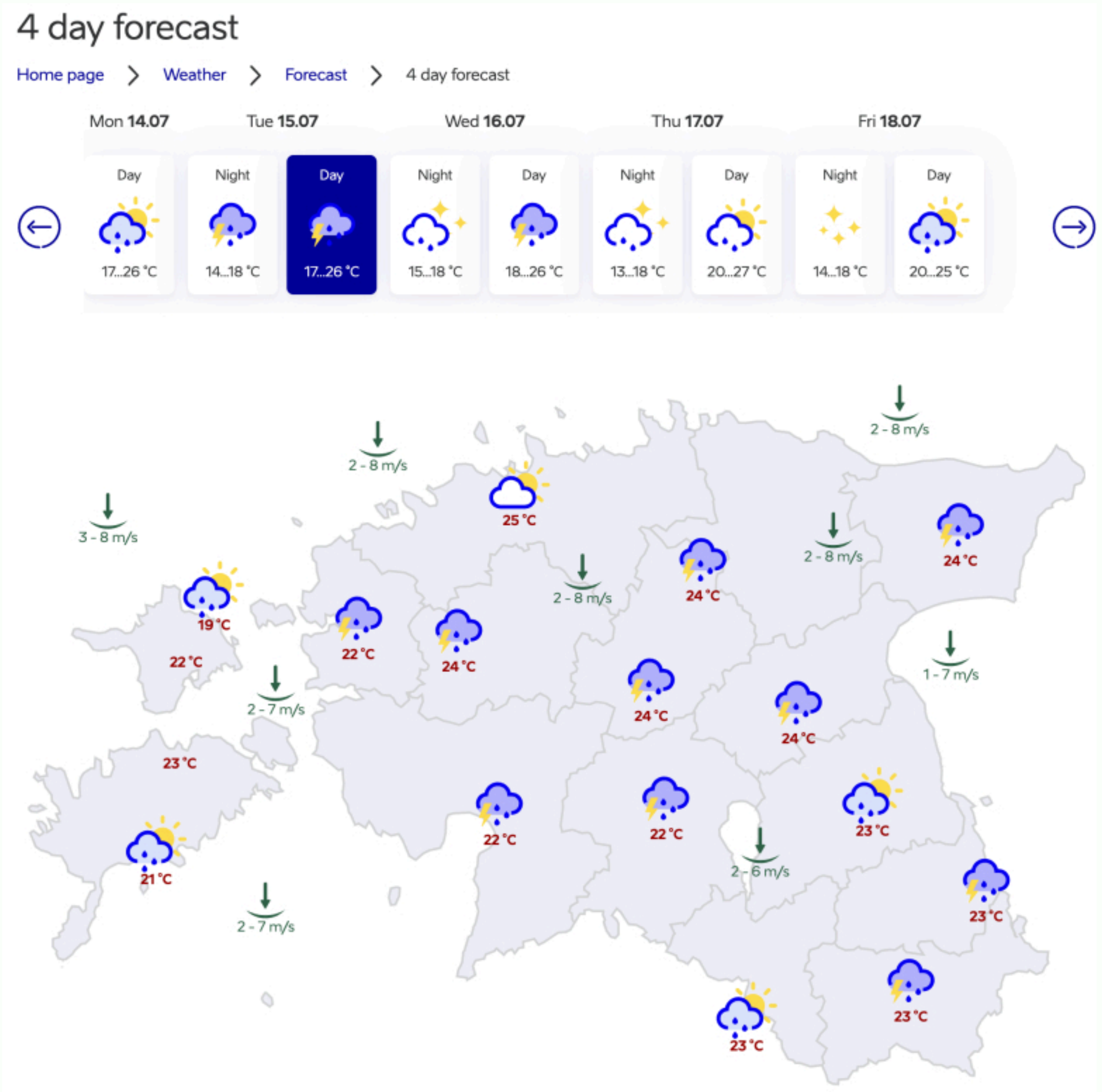
Long-term, no uncertainties

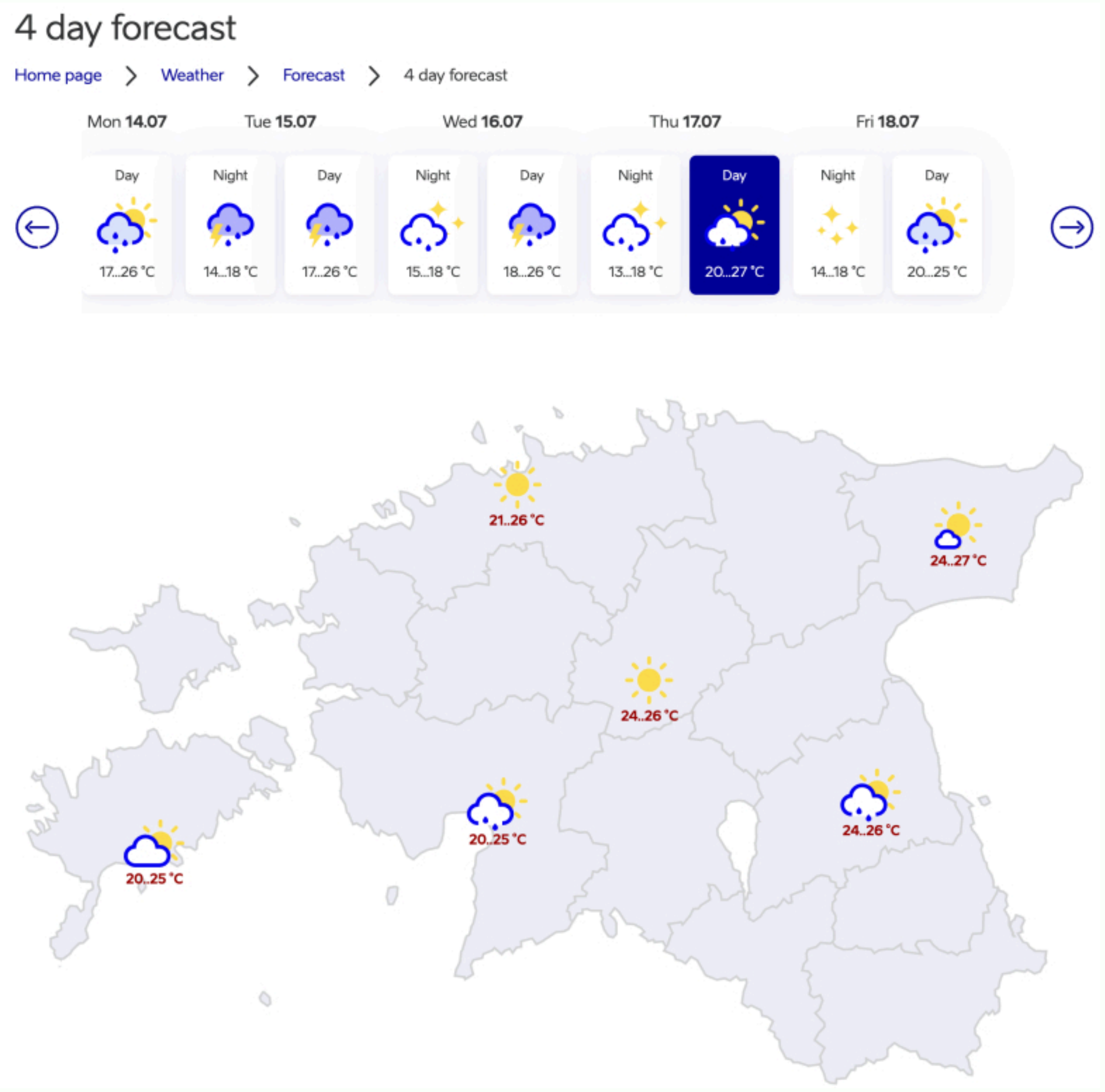


[Accessed July 15, 2025], https://www.timeanddate.com/sun/estonia/tallinn

WHAT CAN BE FORECASTED? (2)

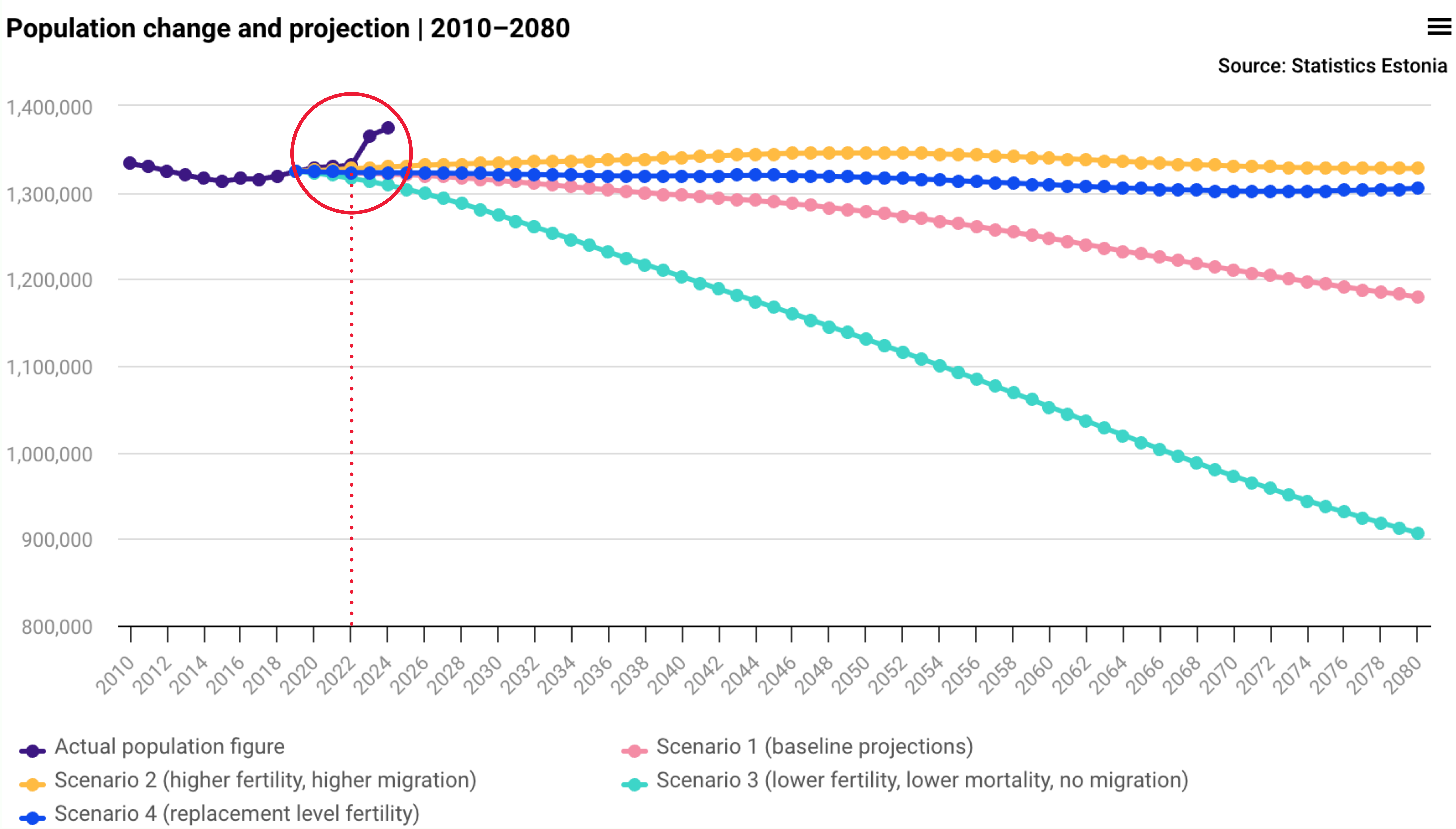
Short-term, many uncertainties





WHAT CAN BE FORECASTED? (3)

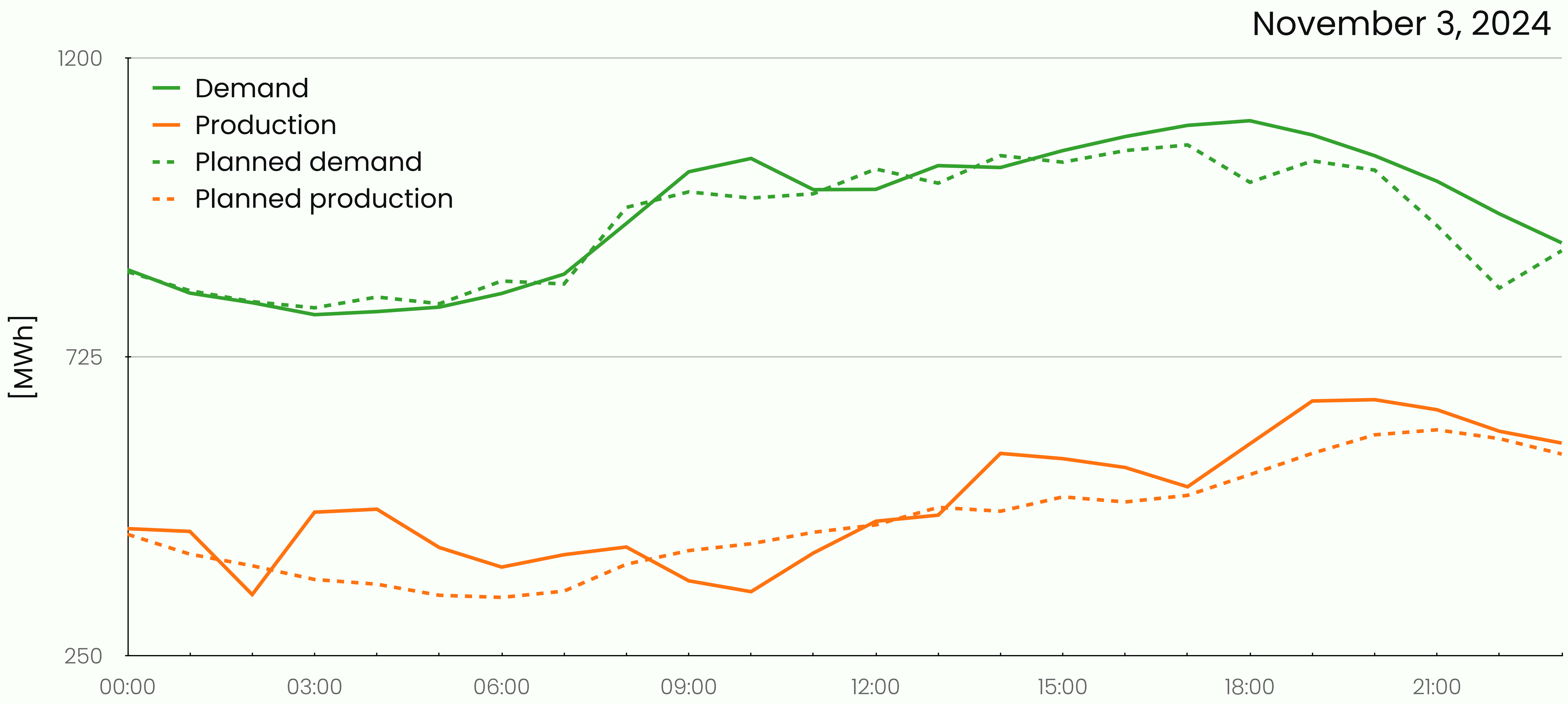
Long-term, many uncertainties



[Accessed June 4, 2024], https://www.stat.ee/en/find-statistics/statistics-theme/population/population-projection

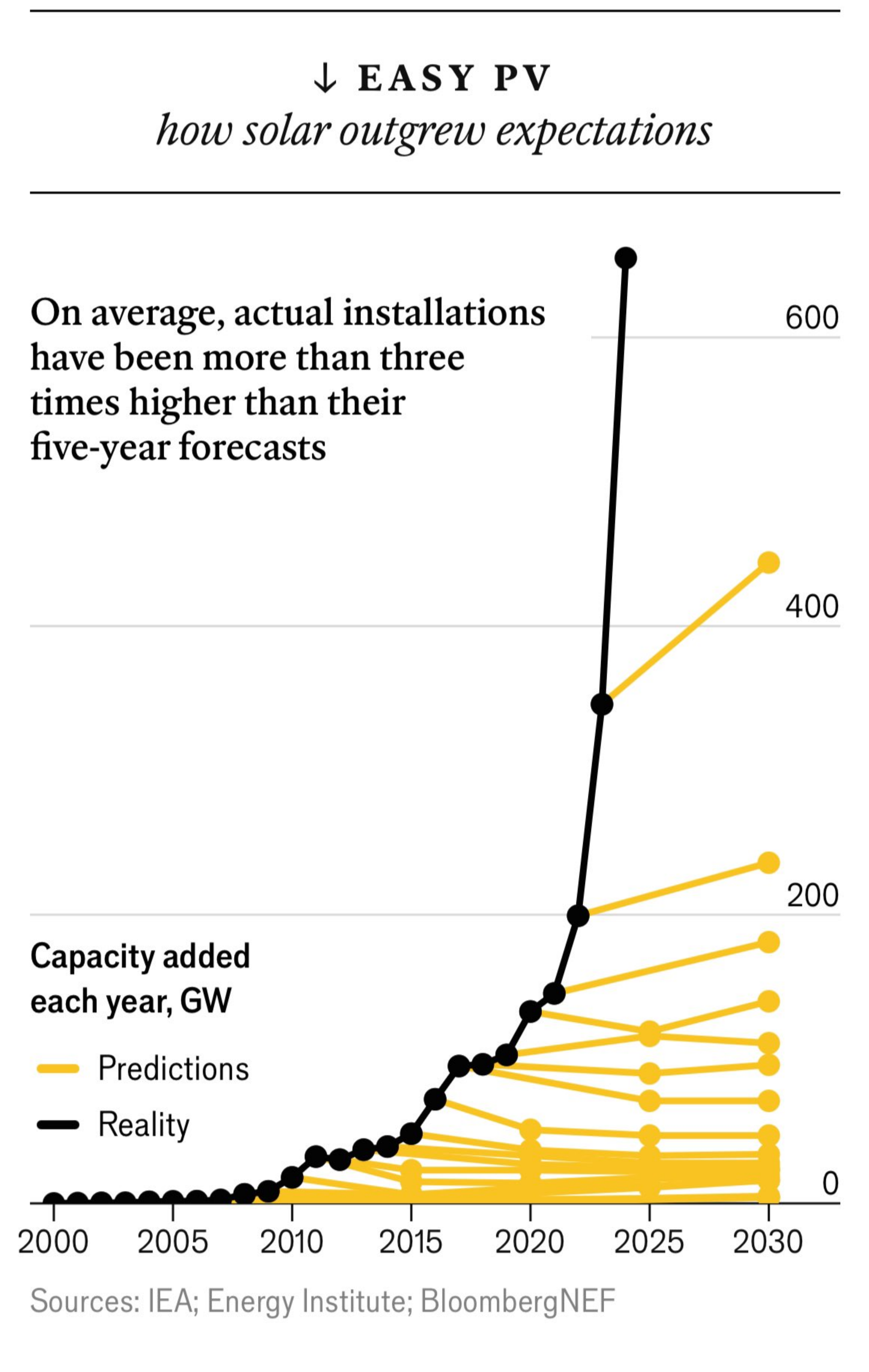
WHAT CAN BE FORECASTED? (4)

Short-term, few uncertainties, independent forecasts



WHAT CAN BE FORECASTED? (5)

Long-term, few uncertainties, dependent forecast



HOW EASY?

Easier to forecast if:

- We understand the key factors.
- Lots of data.
- There similarities with the past behaviour (patterns).
- The forecast cannot affect the thing we are trying to forecast.

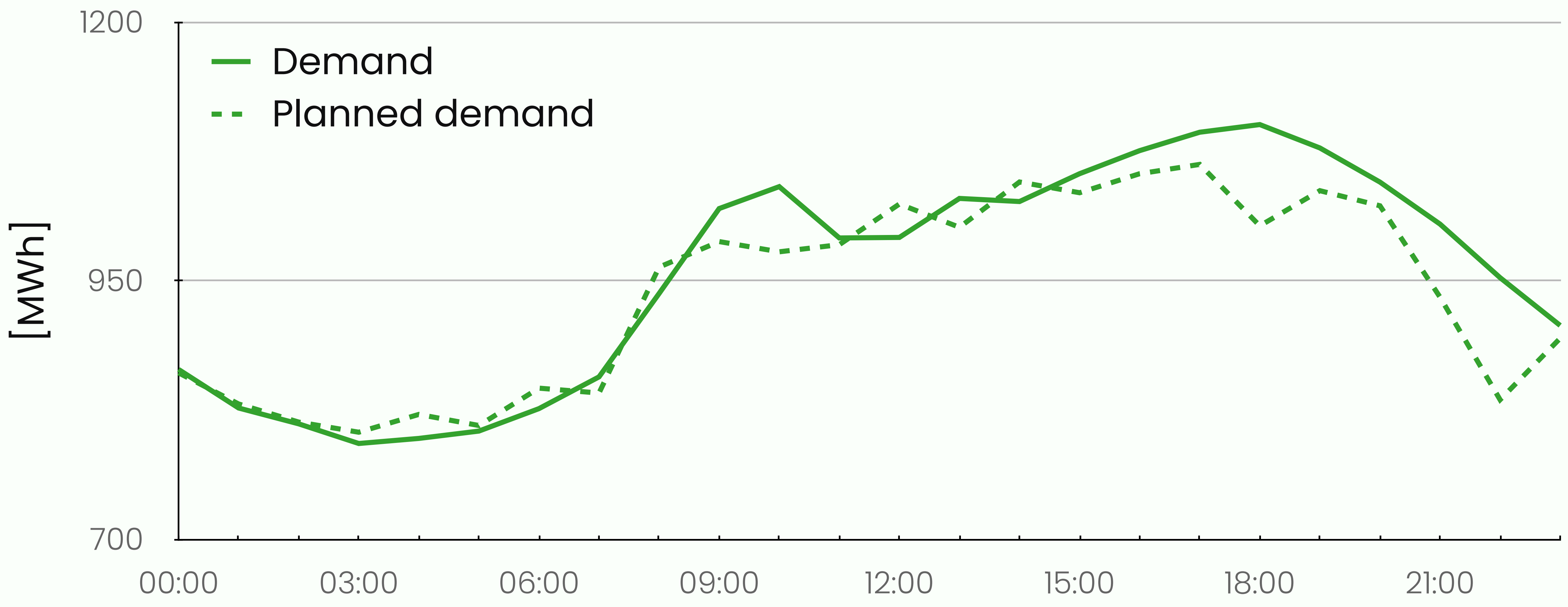
EXAMPLE: *EASY*

Important factors: temperature and calendar.

Data: long history of past demand and weather measurements.

Patterns: seasonality in the data.

Effect: fixed-fee clients are not affected. But for open market clients …



EXAMPLE: *HARD*

Important factors: too many.

Data: long history is available.

Patterns: economy cycles, technical analysis (indicators).

Effect: buy low / sell high.

[Accessed June 6, 2024], https://www.zacks.com/stock/quote/SEDG
https://www.zacks.com/stock/quote/NVDA

THE IDEA

Good forecasts capture **patterns** and **relationships** which exist in the historical data, but do not replicate past events that will not occur again.

FORECASTING FOR DECISION MAKING

Forecasting:

the future, given all of the information available.

Goals:

what has to happen? Linked to forecasts and plans.

Planning:

response to forecasts and goals. Determine actions to make forecasts match goals.

FORECASTING FOR DECISION MAKING (2)

Forecast:

Increasing number of electric vehicles.

Goals:

Building new charging stations.

Planning:

Planning infrastructural developments - where and how many.

FORECASTING: THE BEGINNING

- Decide what should be forecasted.
- Consider forecasting horizon: short- medium-, or long-term.
- Frequency of the forecast.
- Collect the data.

DATA WINDOWS: ROLLING VS RECURSIVE

The **rolling** approach makes use of fixed windows of data to *re-estimate* the parameters over the out-of-sample period, whereas the **recursive** approach makes use of an increasing window to *re-estimate* the models.

Training data

Forecast

Recursive window

	1	...	T-1	T	T+1	T+2	...
Step 1							
Step 2							
...							

Rolling window

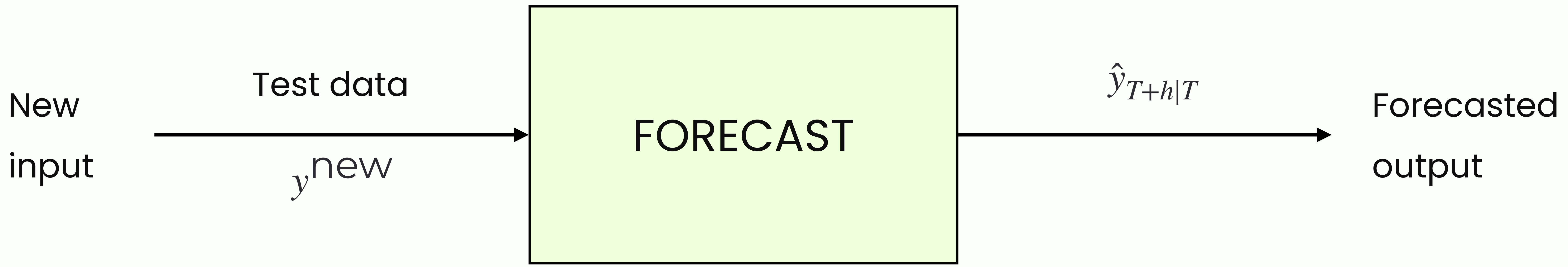
	1	2	...	T-1	T	T+1	T+2	...
Step 1								
Step 2								
...								

LENGTH OF THE TIME SERIES

Rule of thumb: more data is better than less data.

But do we always need all of the data available?

MATHEMATICALLY SPEAKING



Notation $\hat{y}_{T+h|T}$ means forecast of y_{T+h} taking account $(y_1, \ldots, y_T)$ (i.e., an h-step forecast using all observations up to time T).

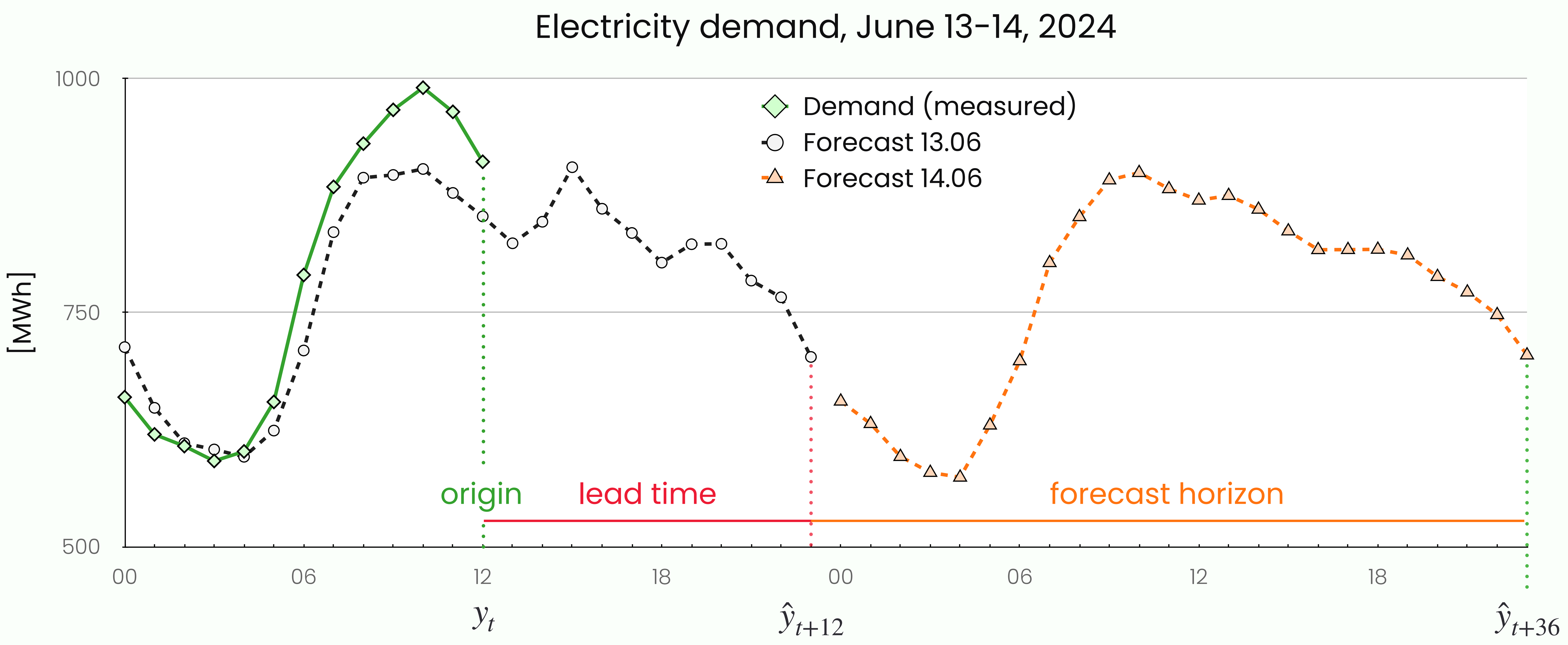
For the sake of simplicity, sometimes we will be using y_T and y_t interchangeably.

TERMINOLOGY

The forecast **origin** is time at which you are making a forecast.

The forecast **horizon** is the time for which you are making a forecast.

The time between the origin and the horizon is the **lead** time (or latency) of the forecast.



TYPES OF FORECAST

Forecasts can be either *in-sample* or *out-of-sample*.

In-sample (training/validation data)					Out-of-sample (test data)			
1	...	T-6	T-5	T-4	T-3	T-2	T-1	T

Out-of-sample forecasts are a better test of how well the model works, as the forecast uses data not included in the estimation of the model.

TYPES OF FORECAST (2)

A **one**-step-ahead is a forecast for the next observation only (i.e., $h = 1$).

A **multi**-step-ahead forecast is for $h > 1$ steps ahead.

Four multi-step Forecasting strategies

G. Bontempi, S. B. Taieb, and Y.-A. Le Borgne. Machine Learning Strategies for Time Series Forecasting. *Lecture Notes in Business Information Processing*, vol 138. Springer, Berlin, 2013.

DIRECT

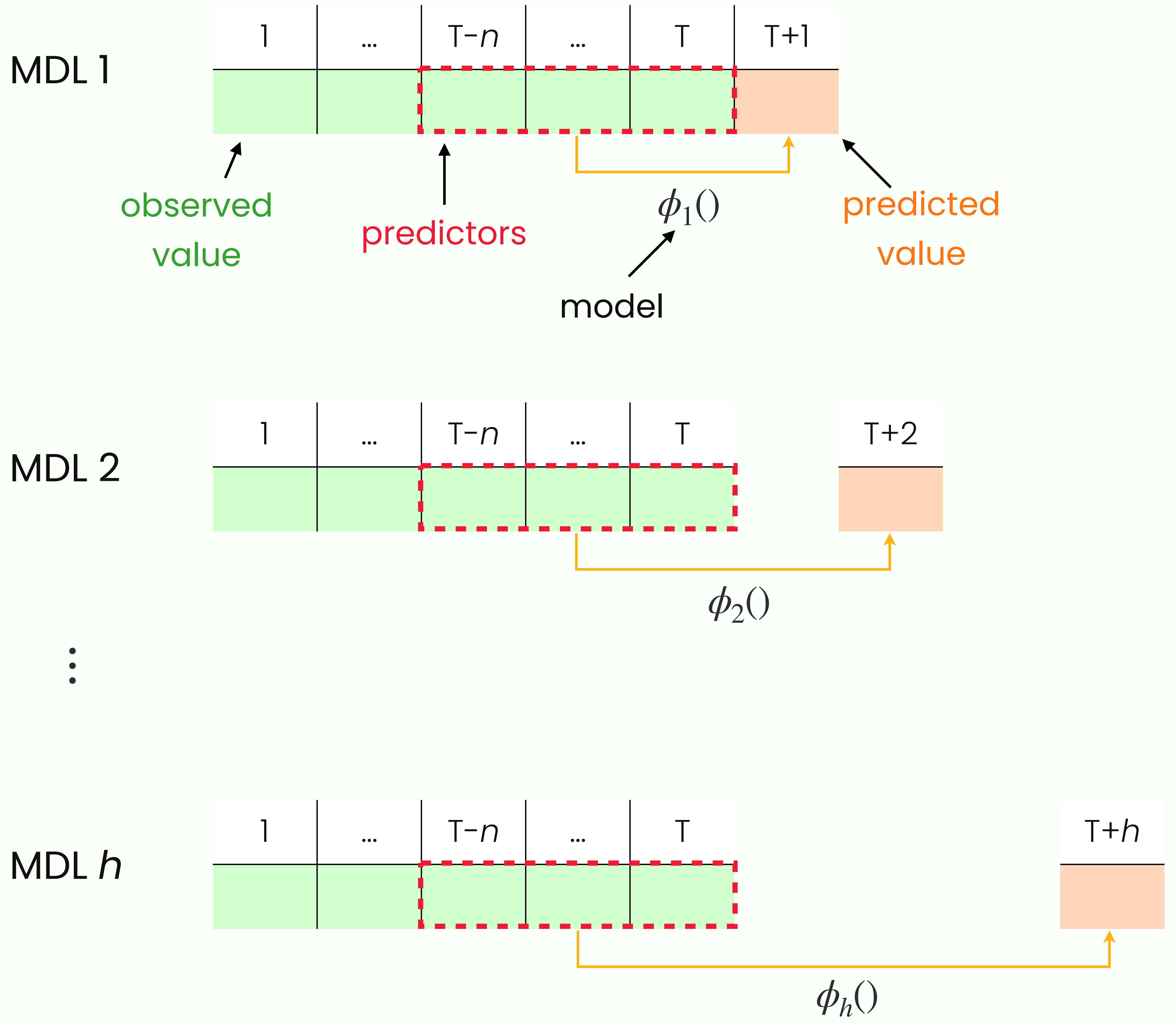
Train a separate model for each step in the horizon:

- model $\phi_1()$ forecasts 1-step ahead,
- model $\phi_2()$ forecasts 2-steps ahead,
- …
- model $\phi_h()$ forecasts h-steps ahead.

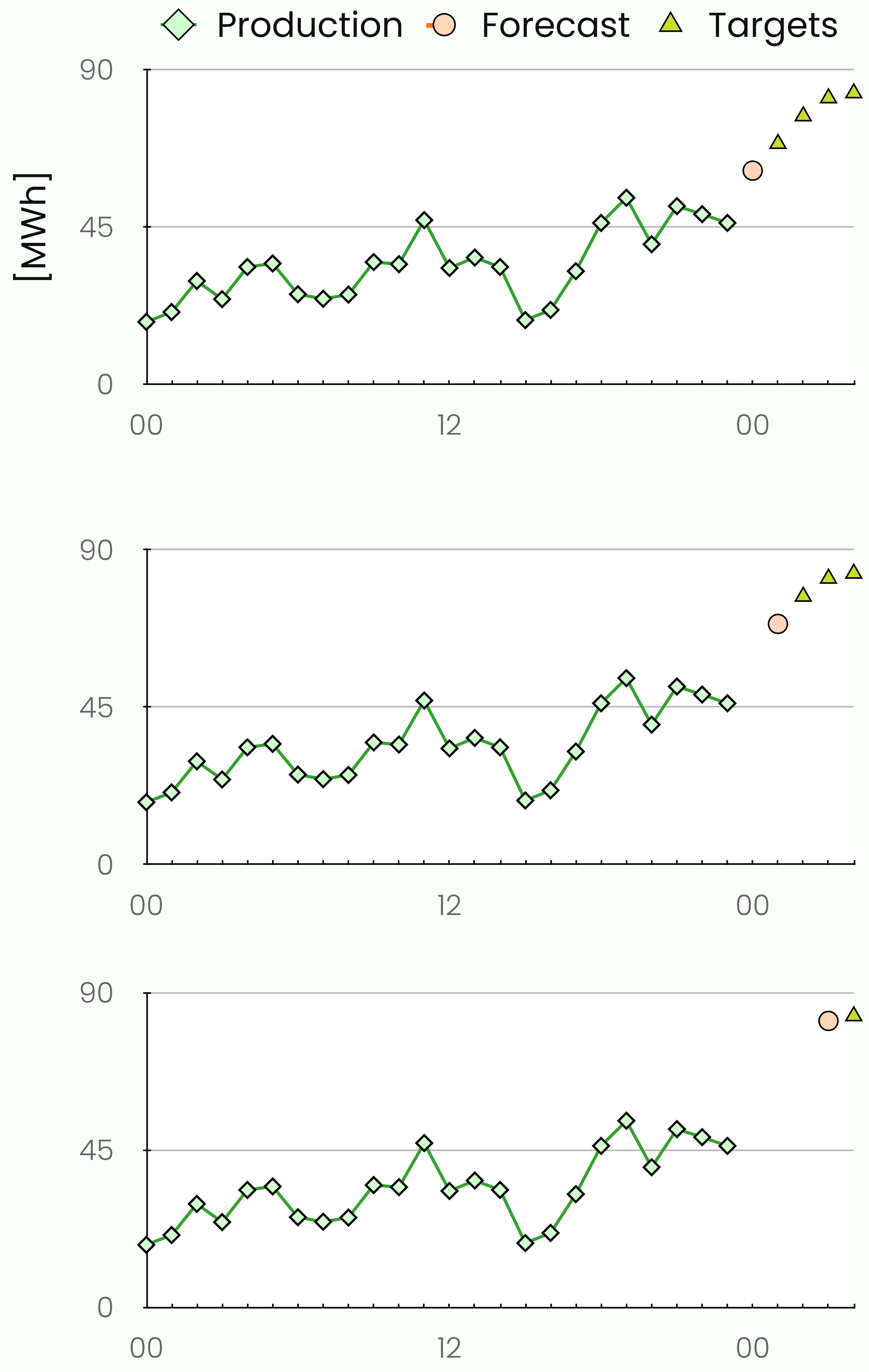
$$
\begin{aligned}
\hat{y}_{t+1} &= \phi_1(y_t, y_{t-1}, \ldots, y_{t-n}) \\
\hat{y}_{t+2} &= \phi_2(y_t, y_{t-1}, \ldots, y_{t-n}) \\
&\vdots \\
\hat{y}_{t+h} &= \phi_h(y_t, y_{t-1}, \ldots, y_{t-n})
\end{aligned}
$$

NB! Training lots of models can be computationally expensive.

DIRECT (2)



MDL 1
1
...
T-n
...
T
T+1
observed value
predictors
$\phi_1()$
model
predicted value
MDL 2
1
...
T-n
...
T
T+2
$\phi_2()$
⋮
MDL h
1
...
T-n
...
T
T+h
$\phi_h()$



Production
Forecast
Targets
90
45
0
[MWh]
00
12
00
90
45
0
00
12
00
90
45
0
00
12
00

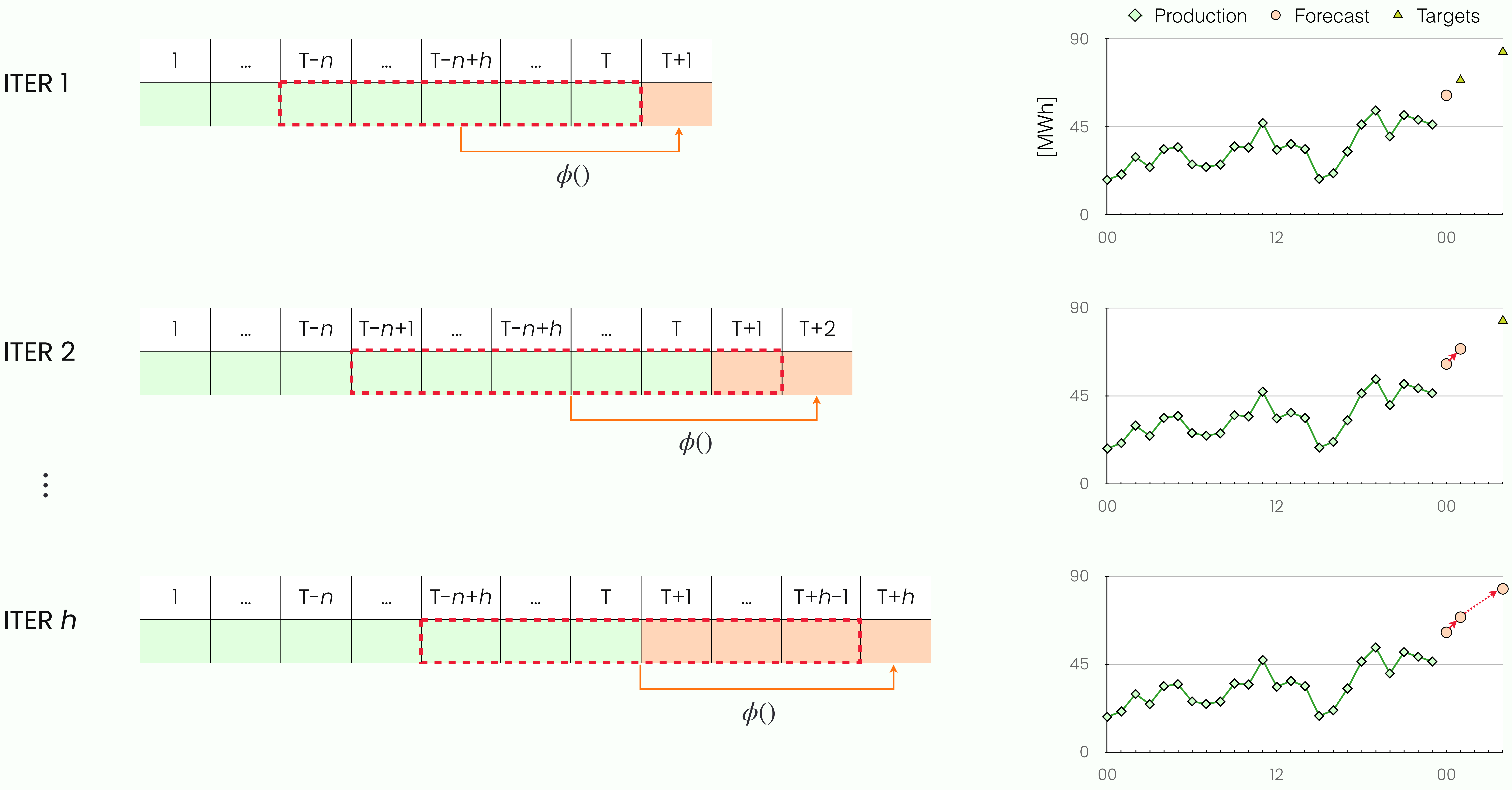
RECURSIVE

Train a single 1-step model and use its forecasts to update the lag features for the next step.

NB! We only need to train one model $\phi()$, but since errors will propagate from step to step, forecasts can be inaccurate for long horizons.

$$
\begin{aligned}
\hat{y}_{t+1} &= \phi(y_t, y_{t-1}, \ldots, y_{t-n}), \\
\hat{y}_{t+2} &= \phi(\hat{y}_{t+1}, y_t, \ldots, y_{t-n+1}), \\
&\vdots \\
\hat{y}_{t+h} &= \phi(\hat{y}_{t+h-1}, \ldots, y_{t-n+h})
\end{aligned}
$$

RECURSIVE (2)


ITER 1
1
...
T-n
...
T-n+h
...
T
T+1
ϕ()
ITER 2
1
...
T-n
T-n+1
...
T-n+h
...
T
T+1
T+2
ϕ()
ITER h
1
...
T-n
...
T-n+h
...
T
T+1
...
T+h-1
T+h
ϕ()
Production
Forecast
Targets
[MWh]
90
45
0
00
12
00

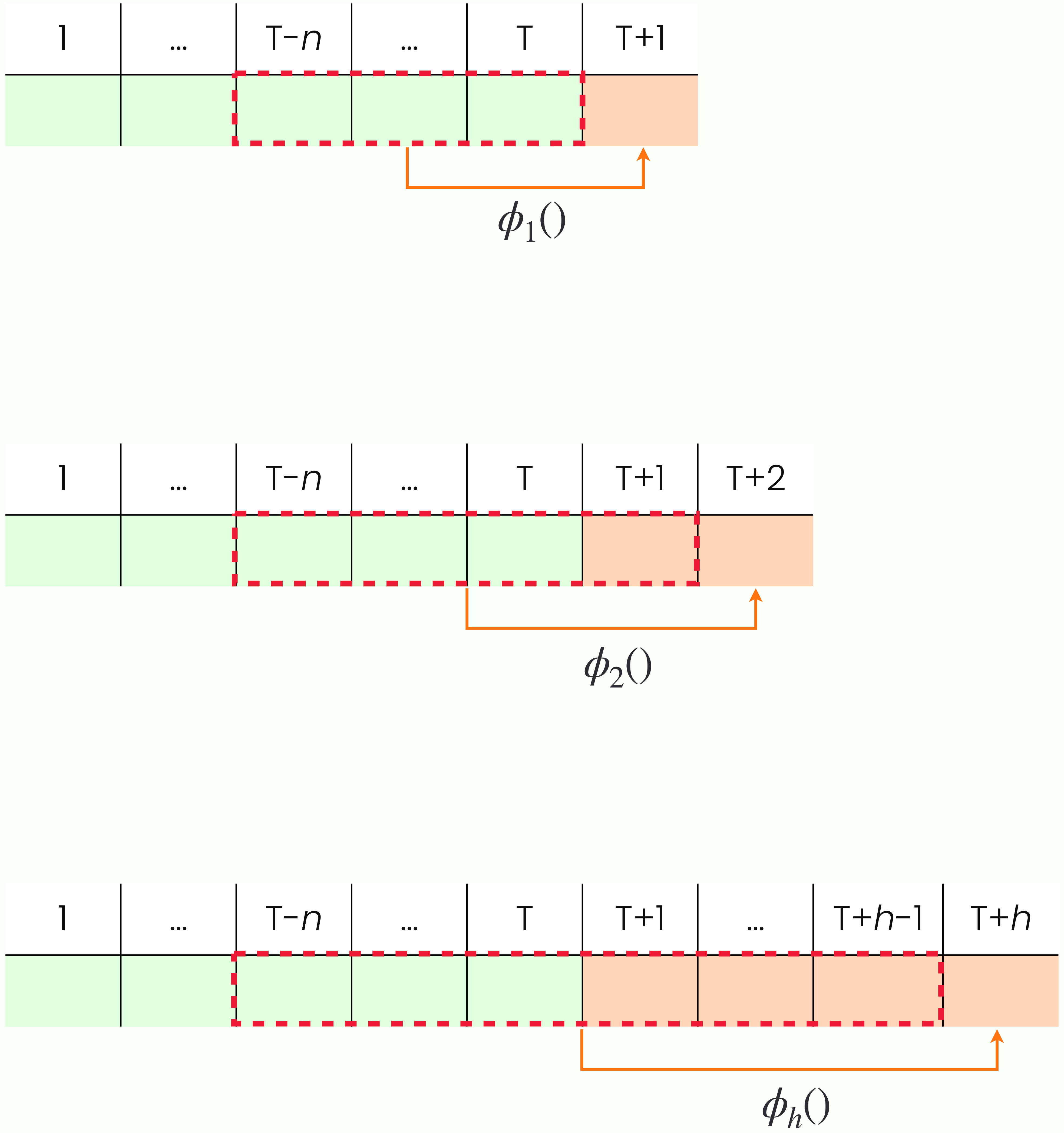
DIRECT + RECURSIVE

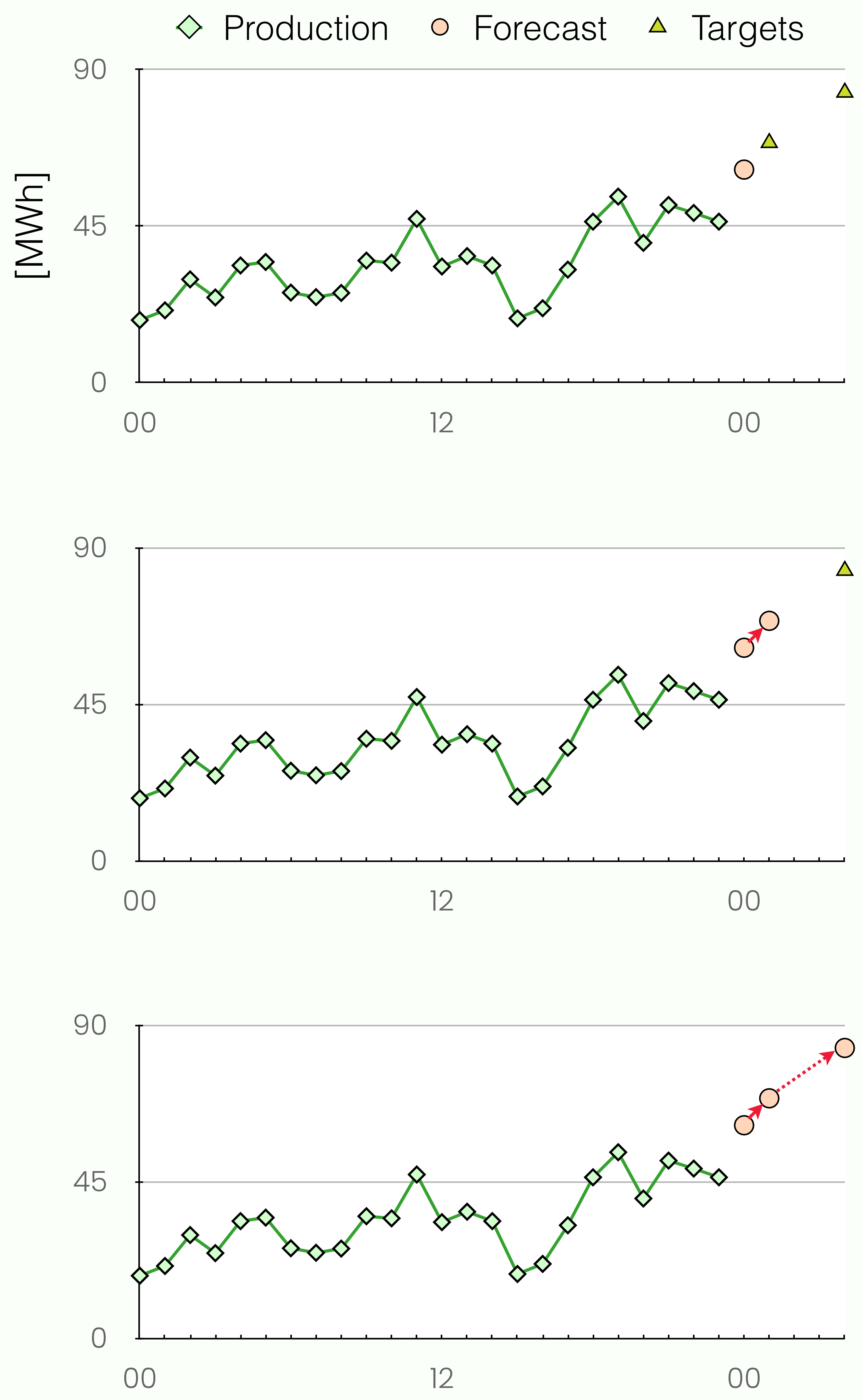
Train a model for each step and use forecasts from previous steps as new lag features.

NB! The model may still suffer from error propagation like Recursive.

$$\begin{aligned}
\hat{y}_{t+1} &= \phi_1(y_t, y_{t-1}, \ldots, y_{t-n}),\\
\hat{y}_{t+2} &= \phi_2(\hat{y}_{t+1}, y_t, \ldots, y_{t-n}),\\
&\vdots\\
\hat{y}_{t+h} &= \phi_h(\hat{y}_{t+h-1}, \ldots, \hat{y}_{t+1}, y_t, \ldots, y_{t-n})
\end{aligned}$$

DIRECT + RECURSIVE (2)





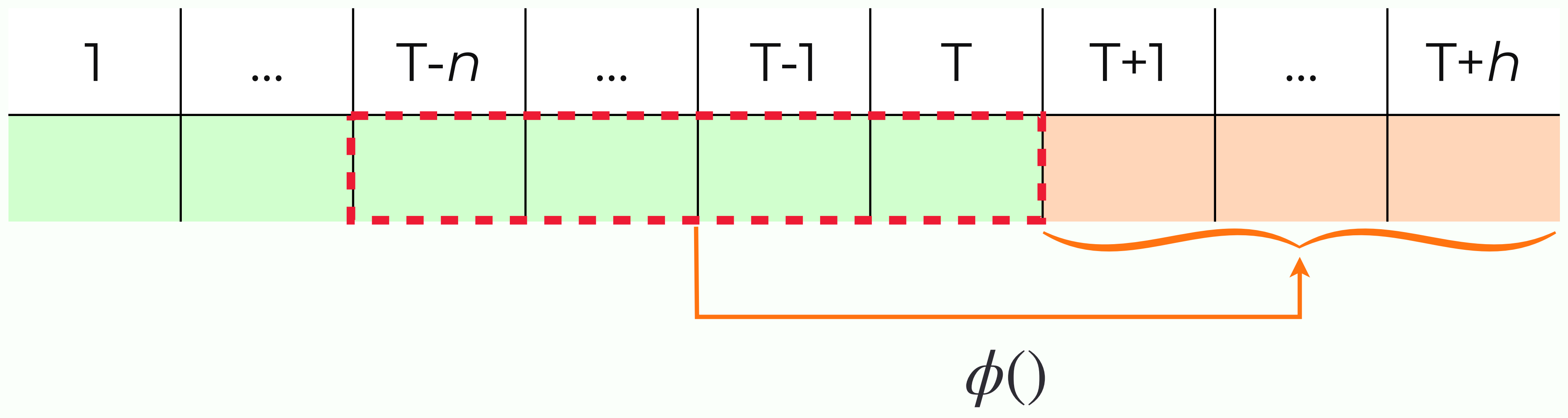
MULTI-OUTPUT

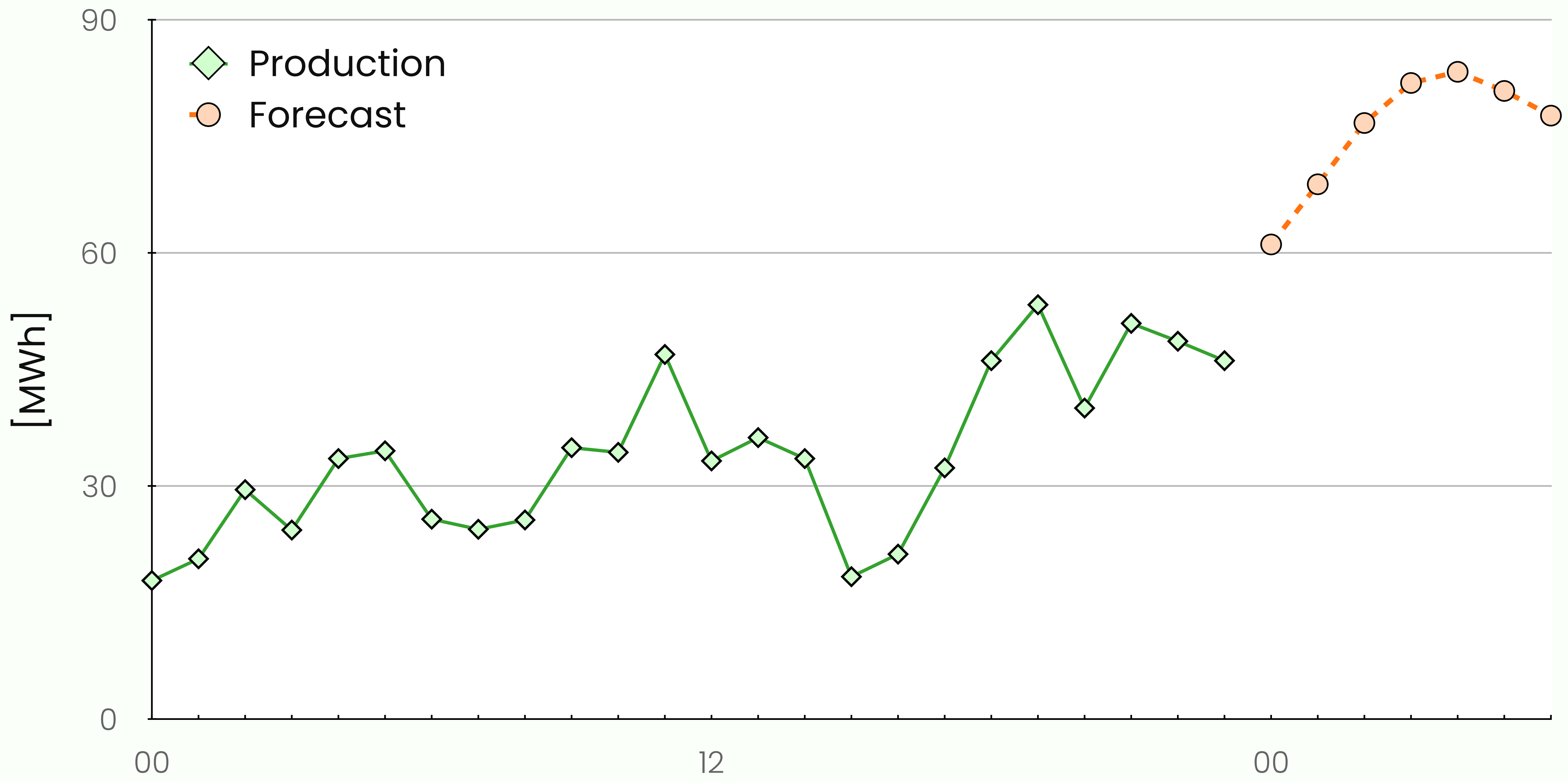
Create a model $\phi()$ that can produce multiple outputs $\hat{y} \in \mathbb{R}^h$.

NB! This strategy is simple and efficient, but may require more data and be slower, in general.

$$\hat{y} = \begin{bmatrix} \hat{y}_{t+1} \\ \vdots \\ \hat{y}_{t+h} \end{bmatrix} = \phi(y_t, y_{t-1}, \ldots, y_{t-n})$$

MULTI-OUTPUT (2)





SUMMARY

	Direct	Recursive	DirRec	MO
Can be used with scalar models	+	+	+	-
Immune to accumulation of errors	+	-	+	+
Dependency between forecasts	-	+	+	+
# of models	h	1	h	1

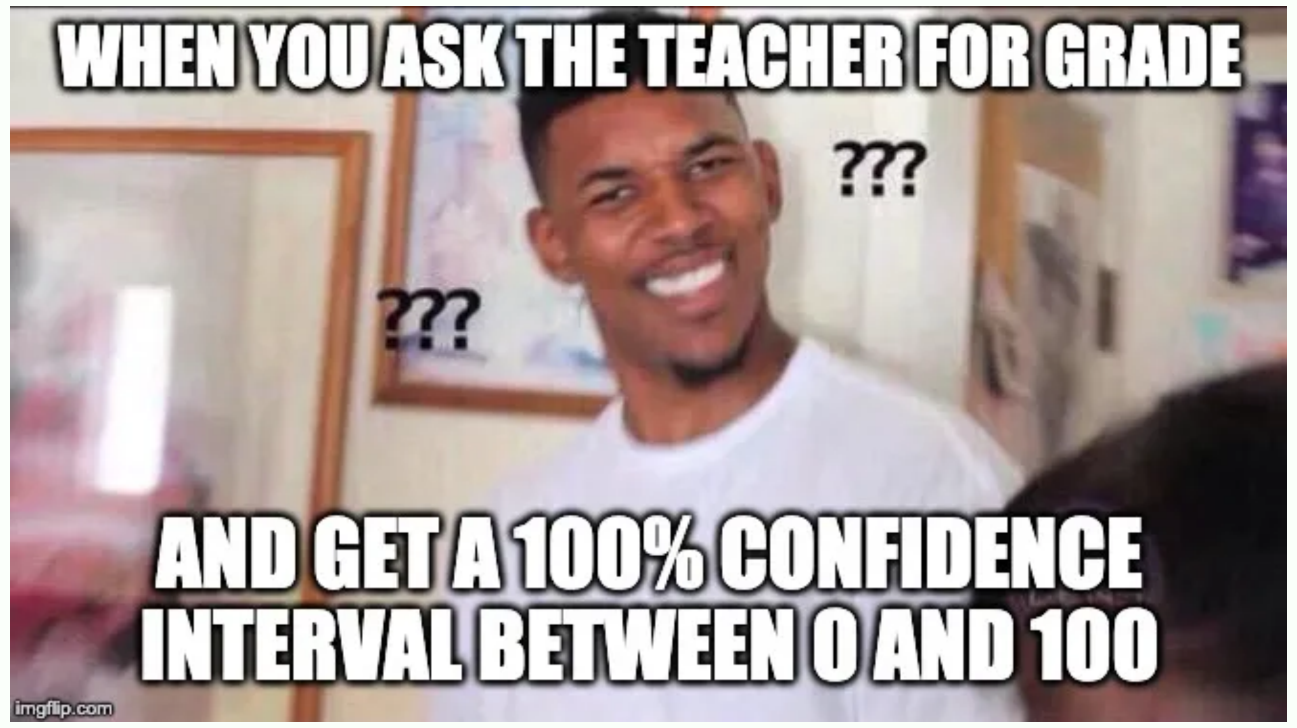
*scalar models refer to models that handle univariate time series data, meaning they work with a single variable over time, e.g., ARMA-family models.

STATISTICAL FORECAST: CI VS PI

- **Confidence** intervals display the probability that a parameter will fall between a pair of values around the mean.
- **Prediction** intervals tell you where you can expect to see the next data point.

A prediction interval predicts an individual number, whereas a confidence interval predicts the mean value.

NB! PI is wider than CI.



PREDICTION INTERVALS

A prediction interval gives an interval within which we expect y_t to lie with a specified probability:

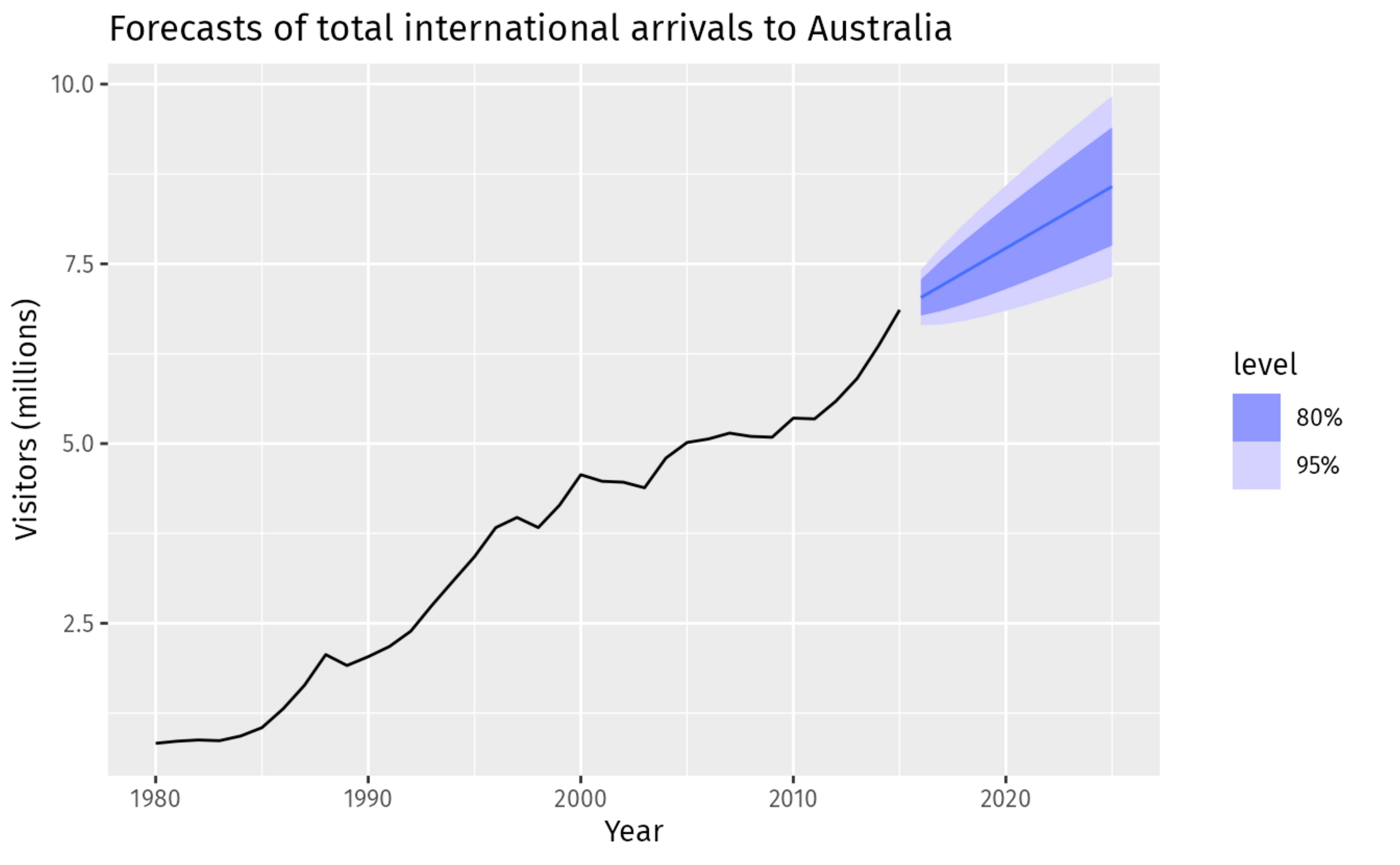
$$\hat{y}_{T+h|T} \pm c\hat{\sigma}_h,$$

where $\hat{\sigma}_h$ is an estimate of the standard deviation of the h-step forecast distribution, and c depends on the coverage probability.

Multipliers (c) used for prediction intervals

Percentage	Multiplier
50	0.67
80	1.28
90	1.64
95	1.96
99	2.58

PREDICTION INTERVALS (2)

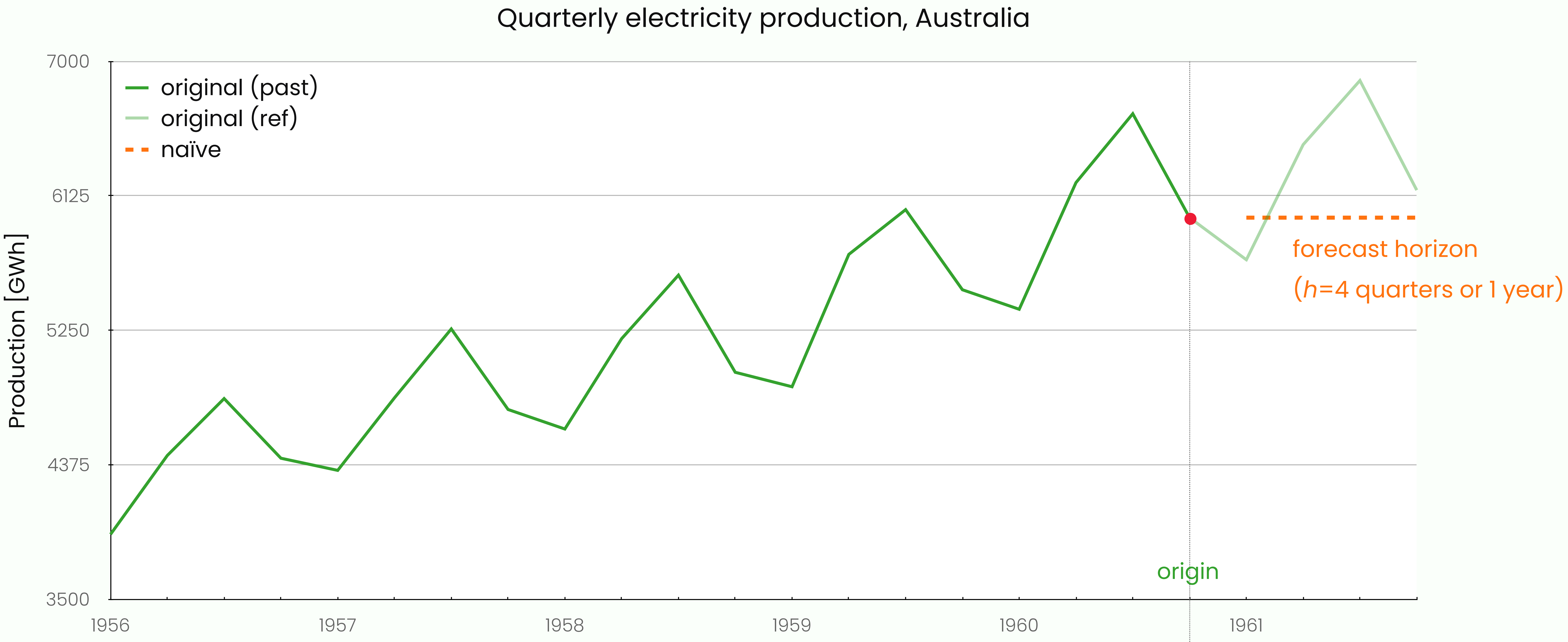


Set a baseline.

Use simple models!

METHOD 1: NAÏVE

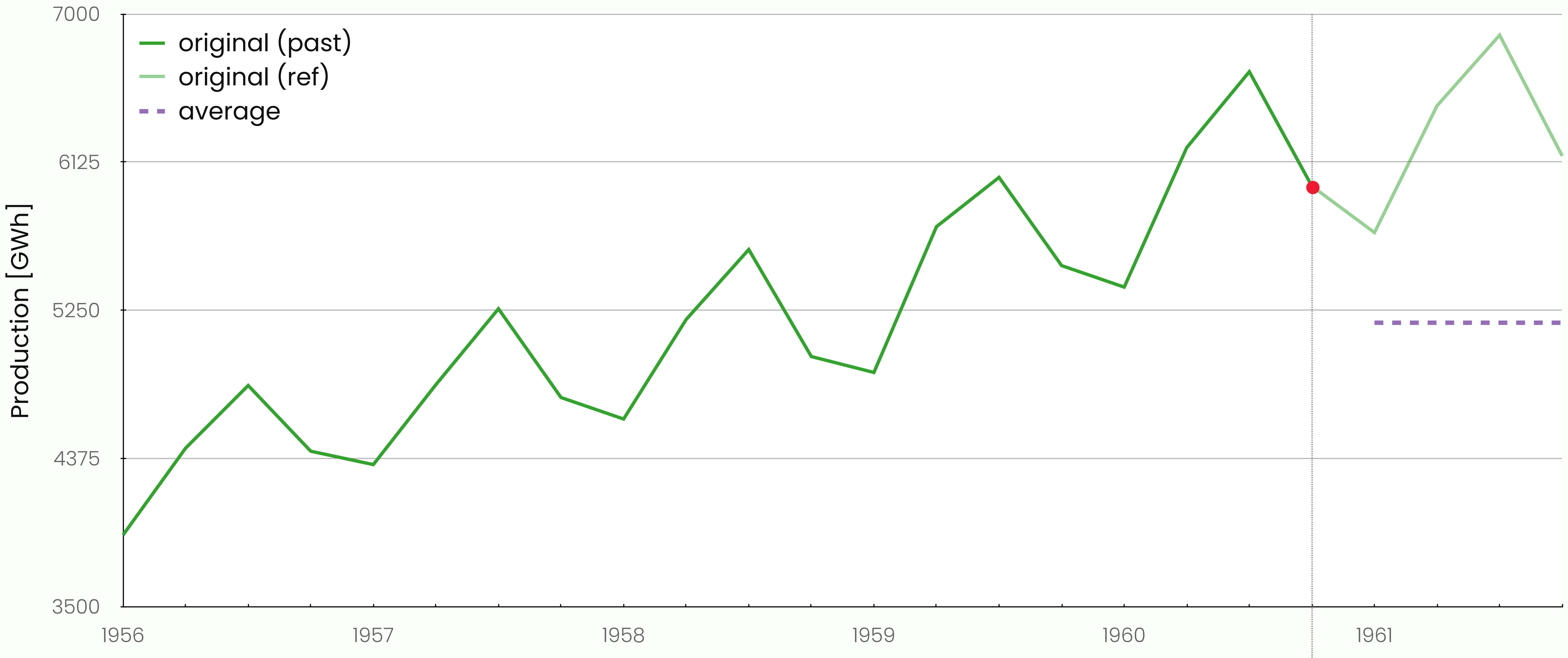
All forecasts are equal to the value of the last observation: $\hat{y}_{T+h|T} = y_T$.



1	...	T	T+1	...	T+4
3923	...	5983	5983	...	5983

METHOD 2: AVERAGE (OR MEAN)

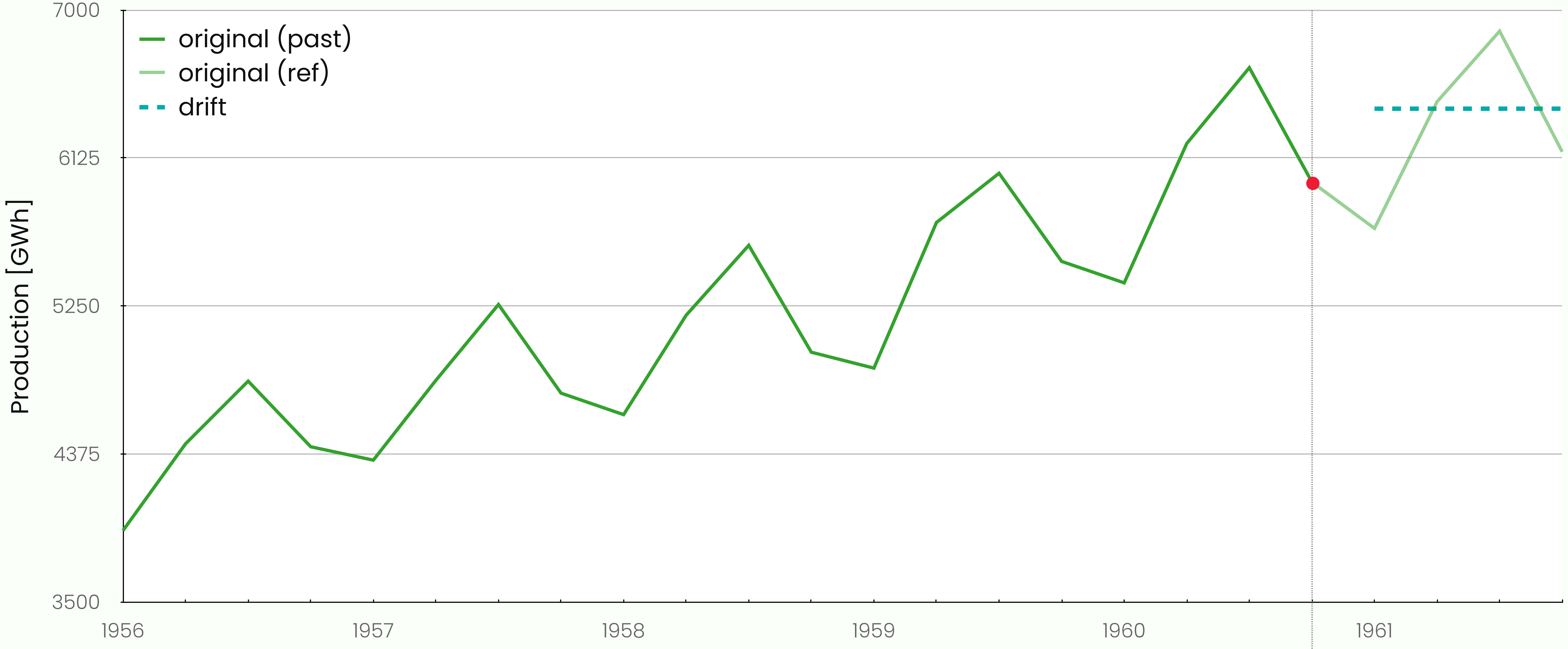
Forecasts are equal to the average of the historical data: $\hat{y}_{T+h|T} = \dfrac{y_1 + \cdots + y_T}{T}$.



1	...	T	T+1	...	T+4
3923	...	5983	5177	...	5177

METHOD 3: DRIFT

A variation of the naïve method, forecasts may change (drift) over time: $\hat{y}_{T+h|T} = y_T + \frac{h}{T-1}(y_T - y_1)$.

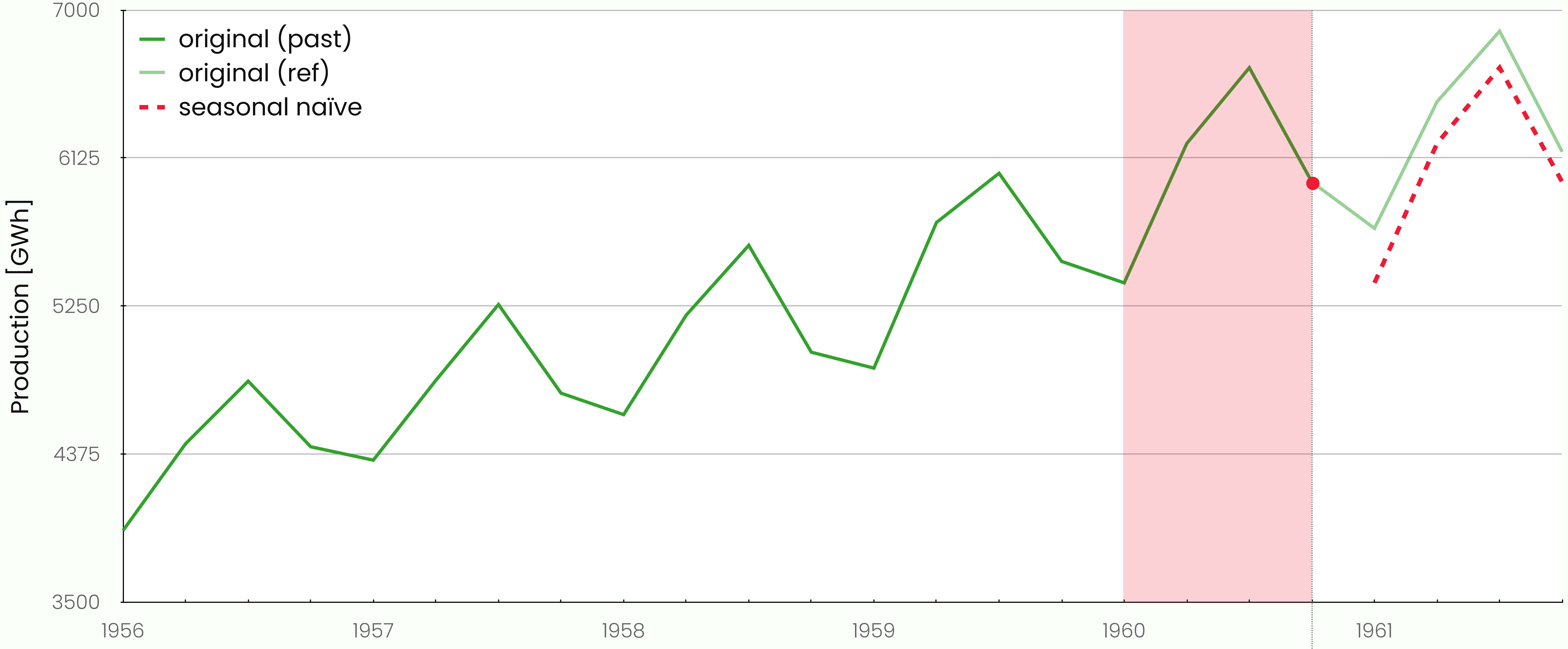


1	...	T	T+1	...	T+4
3923	...	5983	6417	...	6417

METHOD 4: SEASONAL NAÏVE

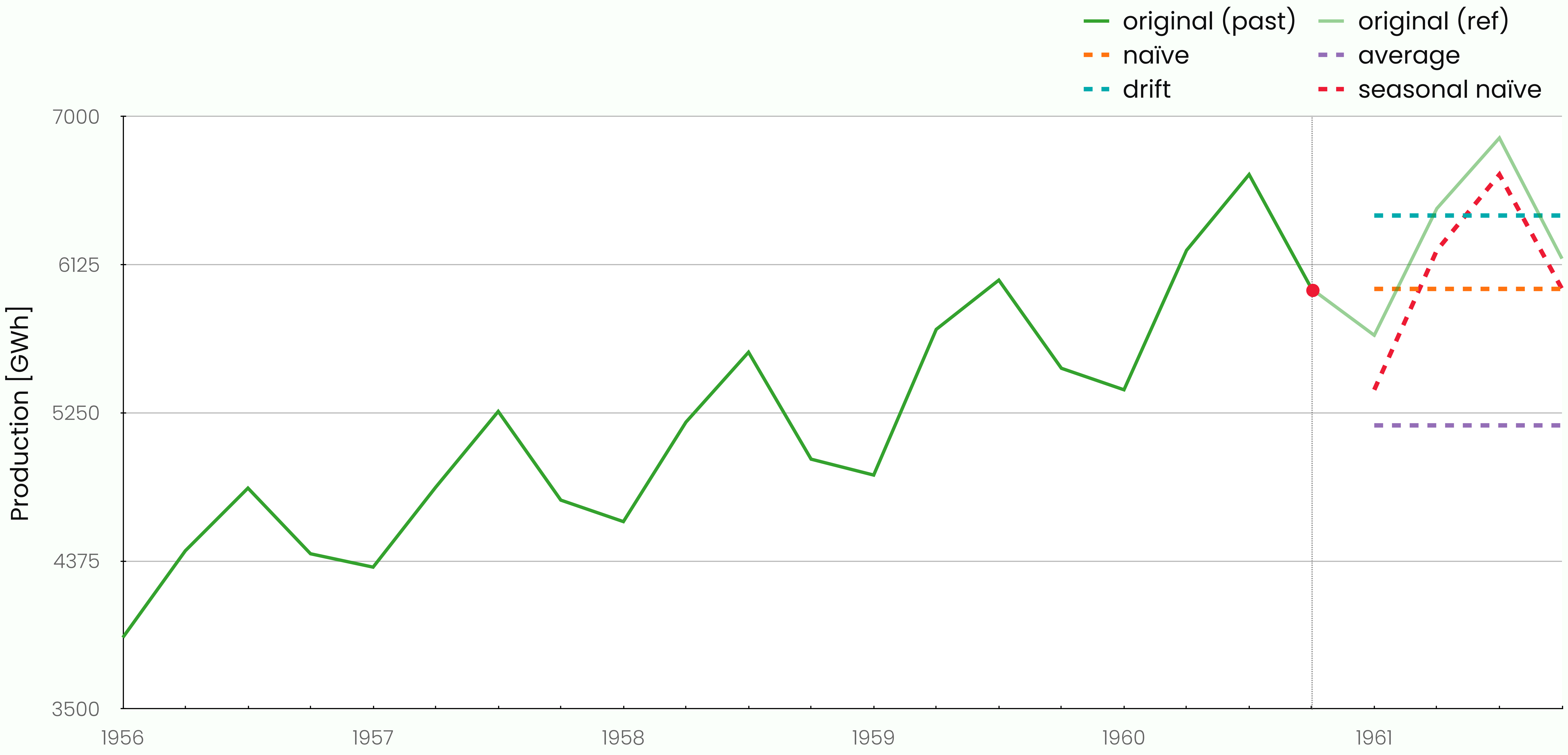
Each forecast is equal to the last observed value from the same season: $\hat{y}_{T+h|T} = y_{T+h-m(k+1)}$.

m is the seasonal period; k is the integer part of $(h-1)/m$, i.e., the number of complete seasons prior to $T+h$



1	...	T	T+1	...	T+4
3923	...	5983	5387	...	5983

FORECAST (OUT-OF-SAMPLE)



CHECKPOINT

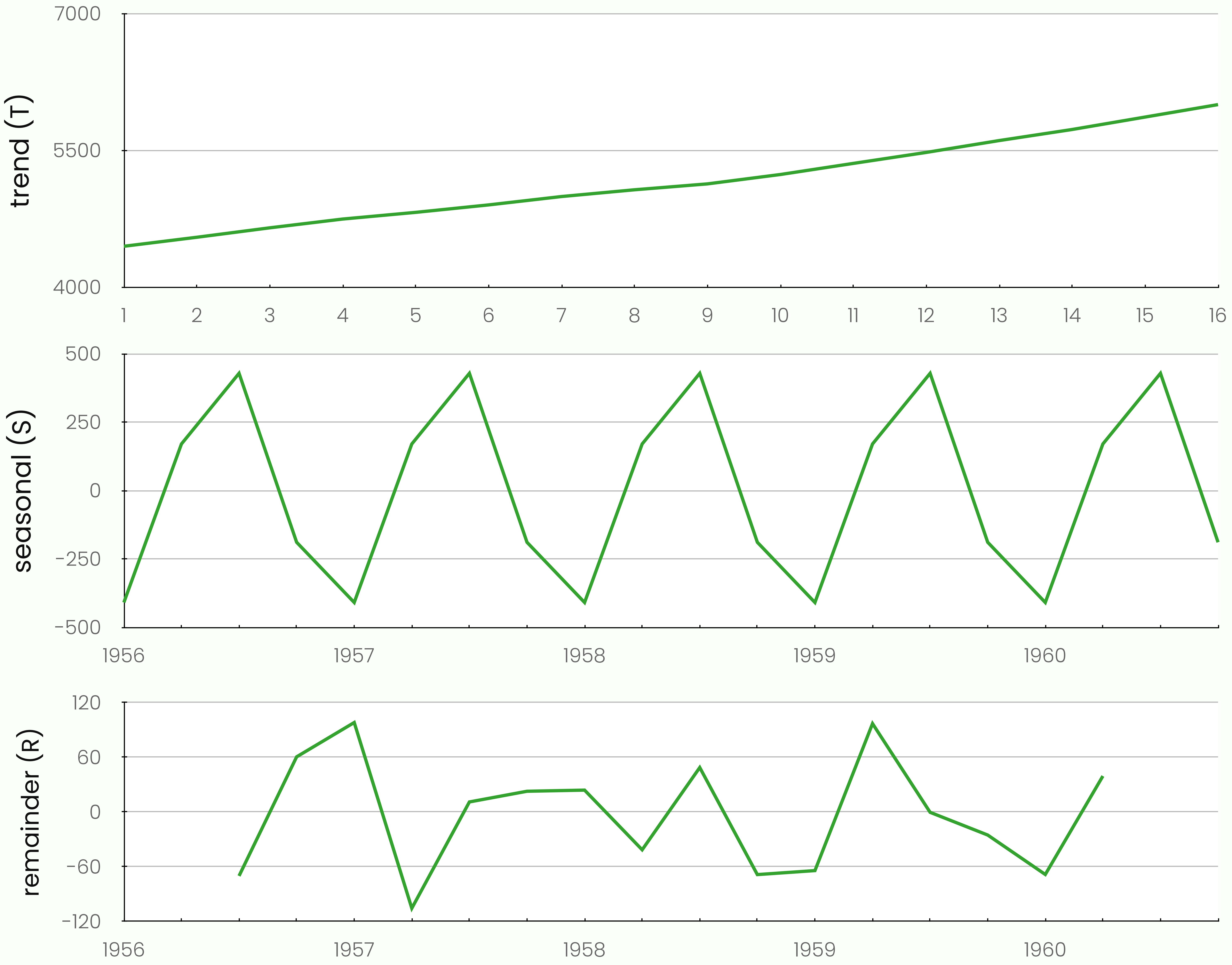
Performance of the forecasting models

	Naïve	Average	Drift	Seasonal naïve
MAE	455	1124	365	241
RMSE	531	1202	440	246

Let’s be a bit more advanced

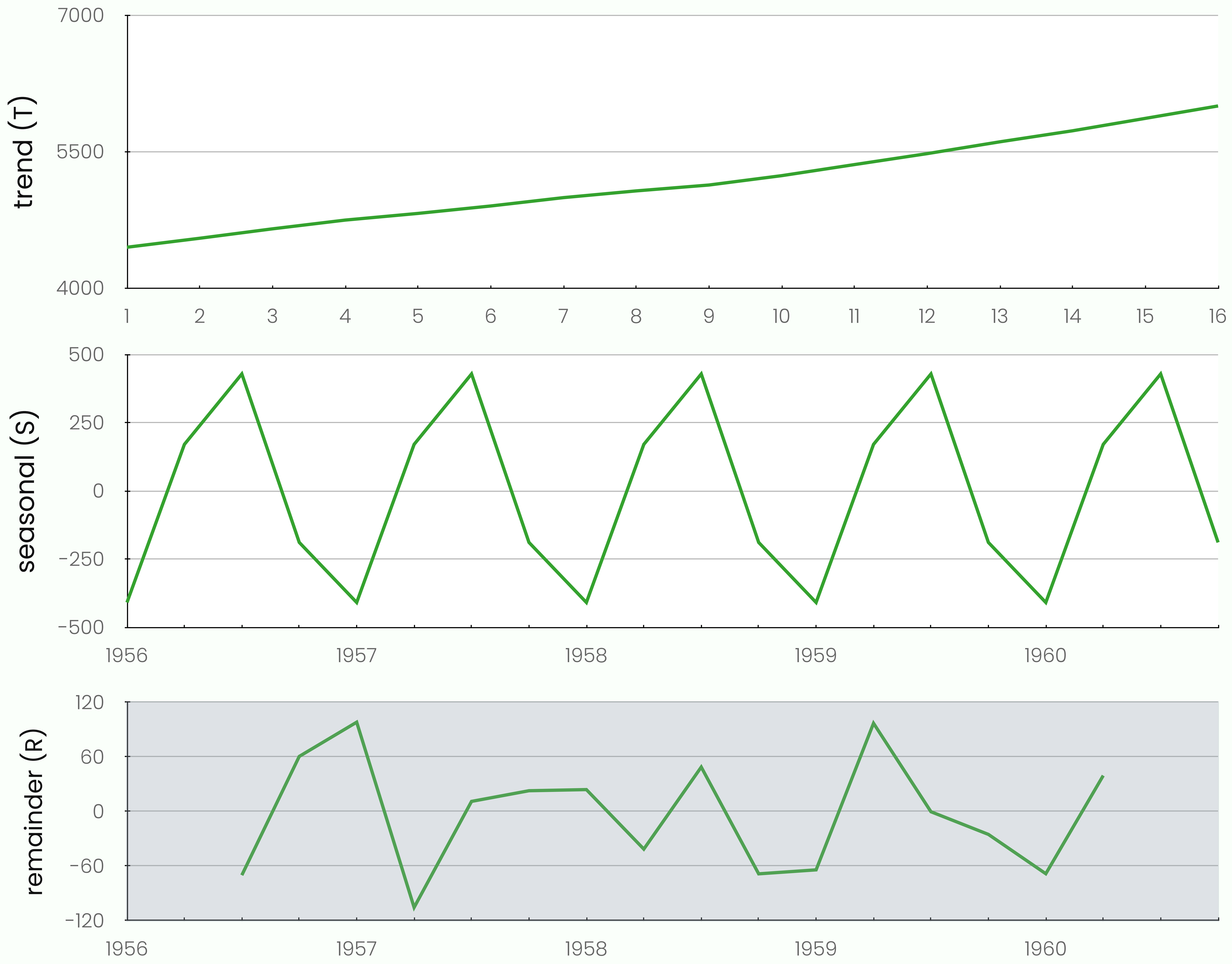
METHOD 5: CLASSICAL DECOMPOSITION, L9

$$y_t = \hat{T}_t + \hat{S}_t + \hat{R}_t$$



METHOD 5: CLASSICAL DECOMPOSITION, L9

$$y_t = \hat{T}_t + \hat{S}_t + \hat{R}_t$$



METHOD 5: CLASSICAL DECOMPOSITION, L9

$$y_t = \hat{T}_t + \hat{S}_t + \hat{R}_t$$

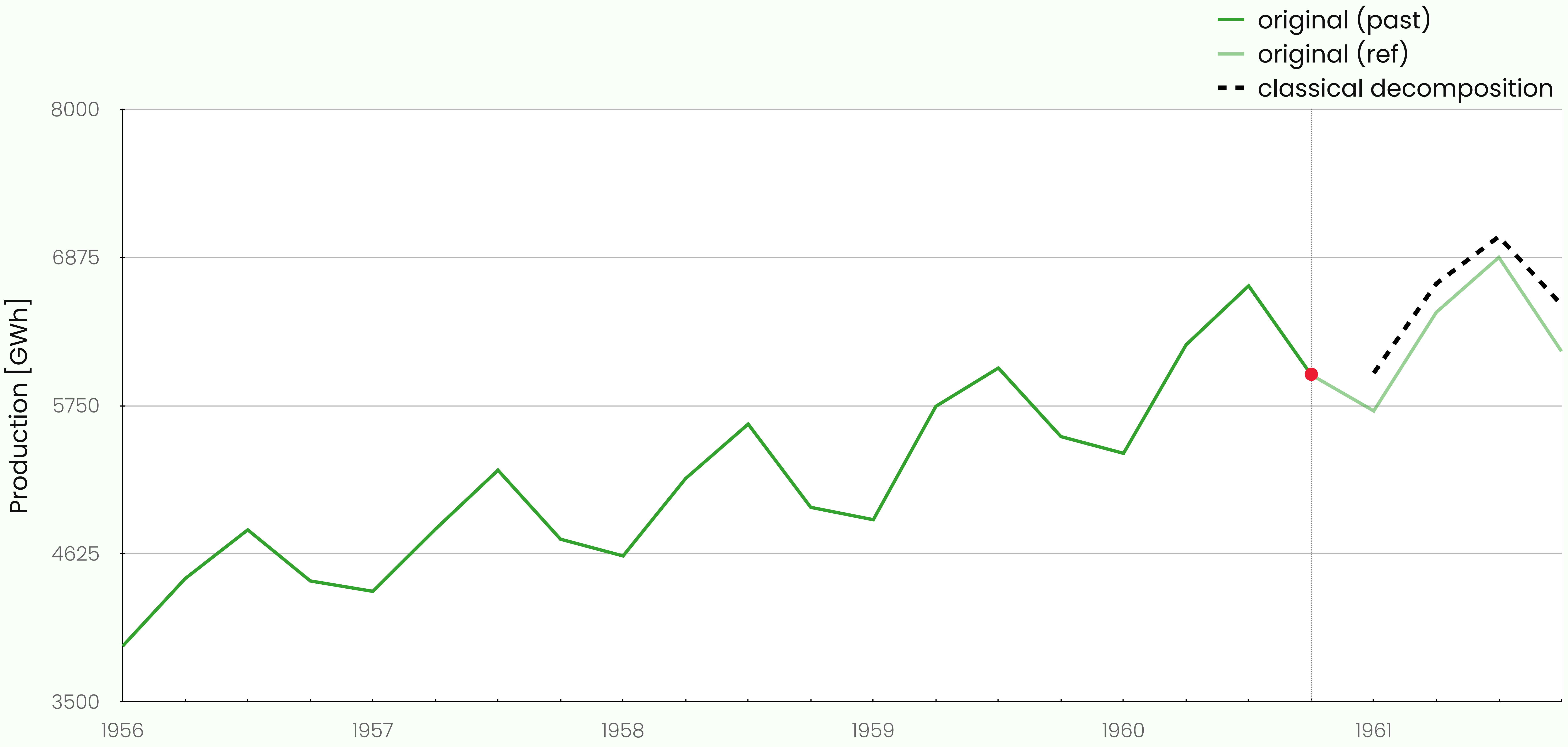
NB! x is index (not a year):

METHOD 5: CLASSICAL DECOMPOSITION, L9

$$y_t = \hat{T}_t + \hat{S}_t + \hat{R}_t$$

NB! x is index (not a year):

CLASSICAL DECOMPOSITION (2)



METHOD 6: SEASONAL ARIMA MODEL

SARIMA (p, d, q) $(P, D, Q)_m$

Non-seasonal part of the model

Seasonal part of the model

SARIMA (2)

```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                      electricity   No. Observations:                   20
Model:             SARIMAX(0, 1, 1)x(1, 0, [], 4)   Log Likelihood                -122.022
Date:                          Tue, 15 Jul 2025   AIC                            252.044
Time:                                  08:22:44   BIC                            255.821
Sample:                              03-01-1956   HQIC                           252.683
                                   - 12-01-1960
Covariance Type:                            opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      4.2971      6.424      0.669      0.504      -8.294      16.888
ma.L1         -0.6389      0.202     -3.157      0.002      -1.036      -0.242
ar.S.L4        0.9575      0.047     20.470      0.000       0.866       1.049
sigma2      1.002e+04   4191.820      2.390      0.017    1801.054    1.82e+04
===================================================================================
Ljung-Box (L1) (Q):                   0.33   Jarque-Bera (JB):                 0.80
Prob(Q):                              0.57   Prob(JB):                         0.67
Heteroskedasticity (H):               0.76   Skew:                             0.30
Prob(H) (two-sided):                  0.74   Kurtosis:                         2.20
===================================================================================
```

$$\underbrace{\left(1 - 0.9575B^4\right)}_{\text{Seasonal AR(1)}} \underbrace{(1 - B)}_{\text{Non-seasonal difference}} \hat{y}_t = \underbrace{\left(1 - 0.6389B\right)}_{\text{Non-seasonal MA(1)}} \epsilon_t$$

SARIMA (2)

```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                        electricity   No. Observations:                   20
Model:             SARIMAX(0, 1, 1)x(1, 0, [], 4)   Log Likelihood                -122.022
Date:                            Tue, 15 Jul 2025   AIC                            252.044
Time:                                    08:22:44   BIC                            255.821
Sample:                                03-01-1956   HQIC                           252.683
                                     - 12-01-1960
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      4.2971      6.424      0.669      0.504      -8.294      16.888
ma.L1         -0.6389      0.202     -3.157      0.002      -1.036      -0.242
ar.S.L4        0.9575      0.047     20.470      0.000       0.866       1.049
sigma2      1.002e+04   4191.820      2.390      0.017    1801.054    1.82e+04
===================================================================================
Ljung-Box (L1) (Q):                   0.33   Jarque-Bera (JB):                 0.80
Prob(Q):                              0.57   Prob(JB):                         0.67
Heteroskedasticity (H):               0.76   Skew:                             0.30
Prob(H) (two-sided):                  0.74   Kurtosis:                         2.20
===================================================================================
```

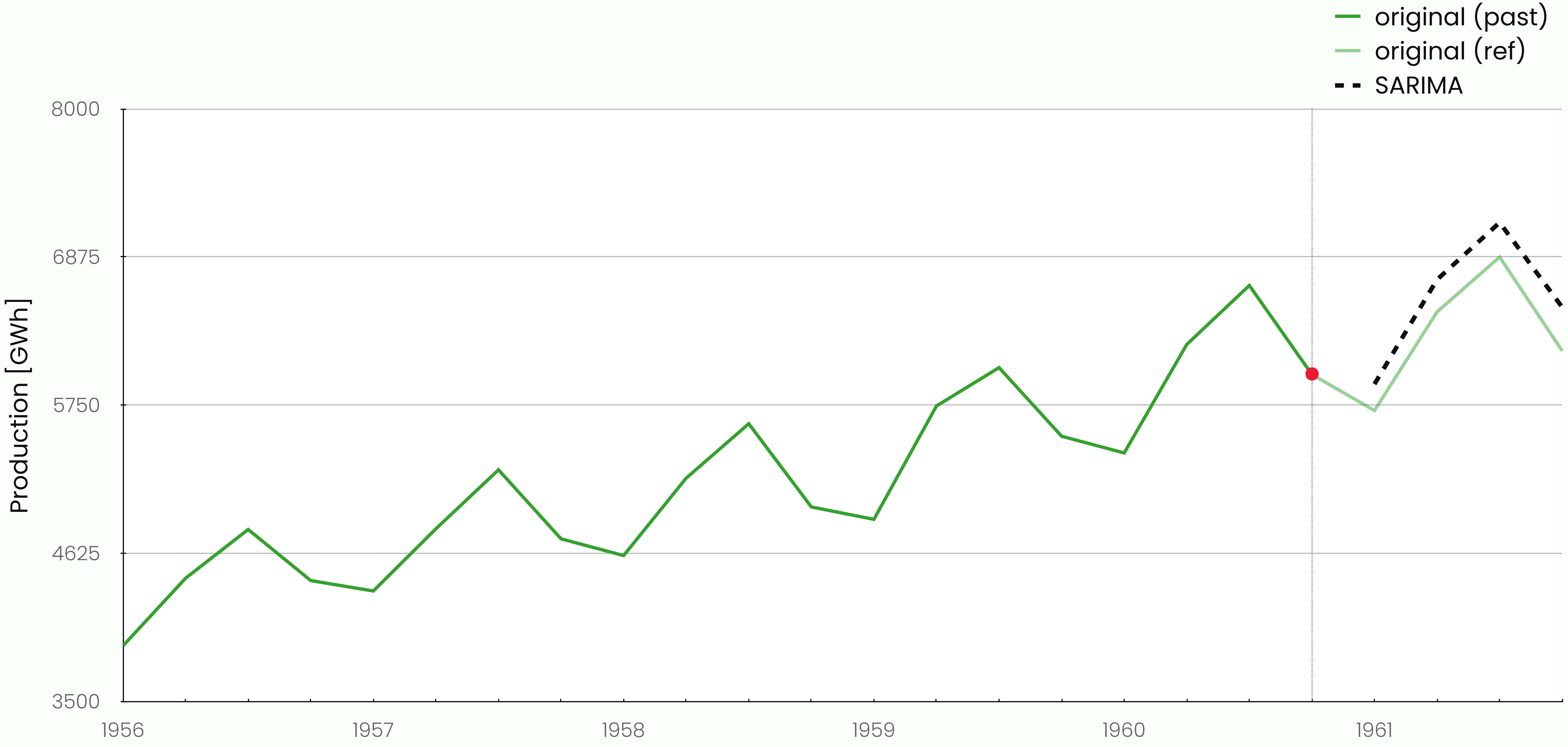
$$\left(1 - 0.9575B^4\right)(1 - B)\,\hat{y}_t = \left(1 - 0.6389B\right)\epsilon_t$$

Seasonal AR(1)

Non-seasonal difference

Non-seasonal MA(1)

SARIMA (3)


original (past)
original (ref)
SARIMA
8000
6875
5750
4625
3500
Production [GWh]
1956
1957
1958
1959
1960
1961

CHECKPOINT 2

Performance of the forecasting models

	Naïve	Average	Drift	Seasonal naïve	Classical decomposition	SARIMA
MAPE	7.01	17.47	5.96	3.88	4.11	4.13
RMSE	531	1202	440	246	264	264

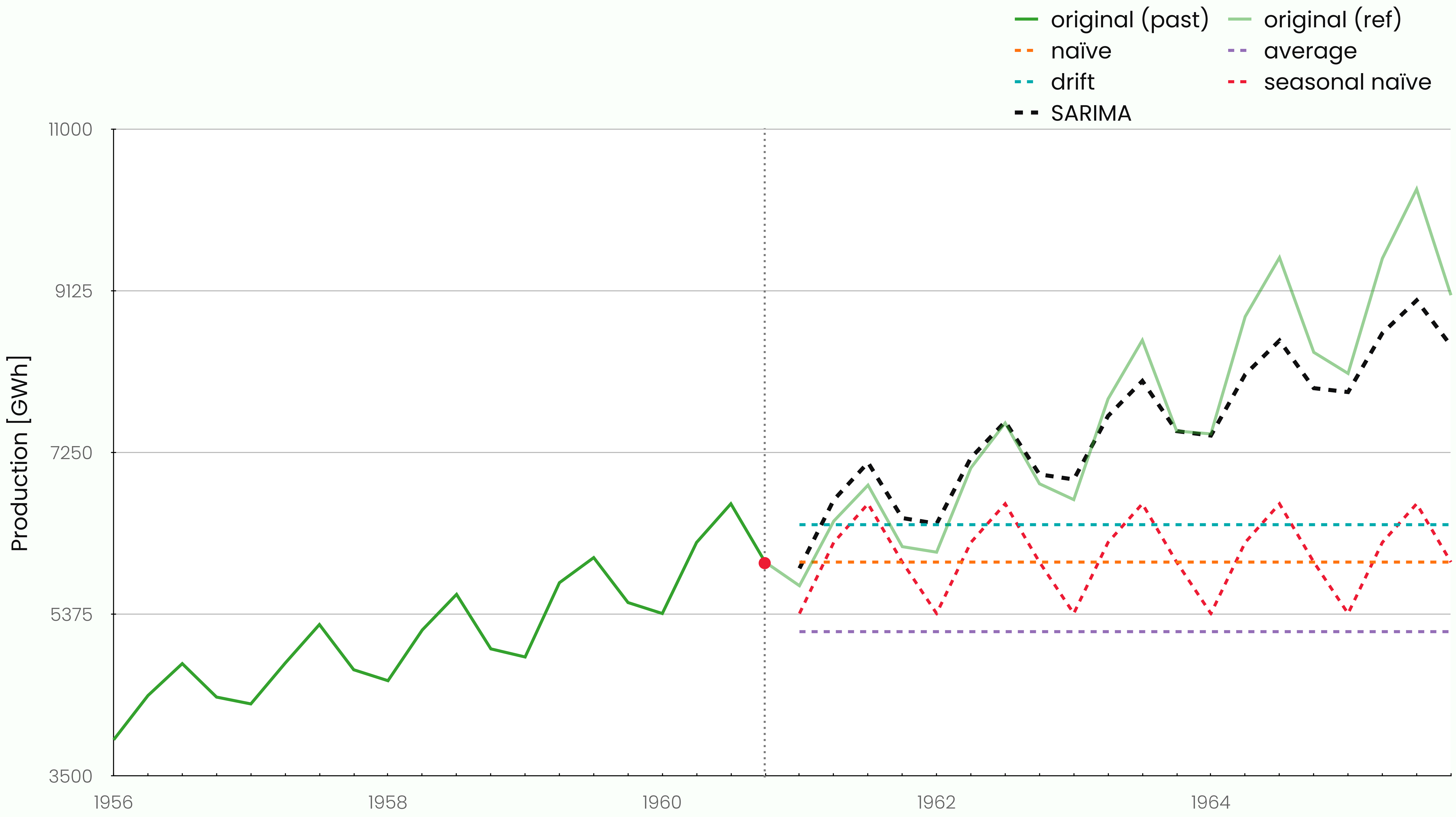
CHECKPOINT 2

Performance of the forecasting models

	Naïve	Average	Drift	Seasonal naïve	Classical decomposition	SARIMA
MAPE	7.01	17.47	5.96	3.88	4.11	4.13
RMSE	531	1202	440	246	264	264

Any remarks?

LONGER FORECAST!?



HOME ACTIVITIES

- ✳Try to improve the forecast from seasonal decomposition by using a more accurate method to estimate the trend.
- ✳Experiment with the window length and see how it affects the quality of the forecast.
- ✳Try other metrics to validate the forecast. Are they more informative?

Thank you!

Questions?