ENERGY DATA SCIENCE

Introduction to Time series analysis

Prof. Juri Belikov

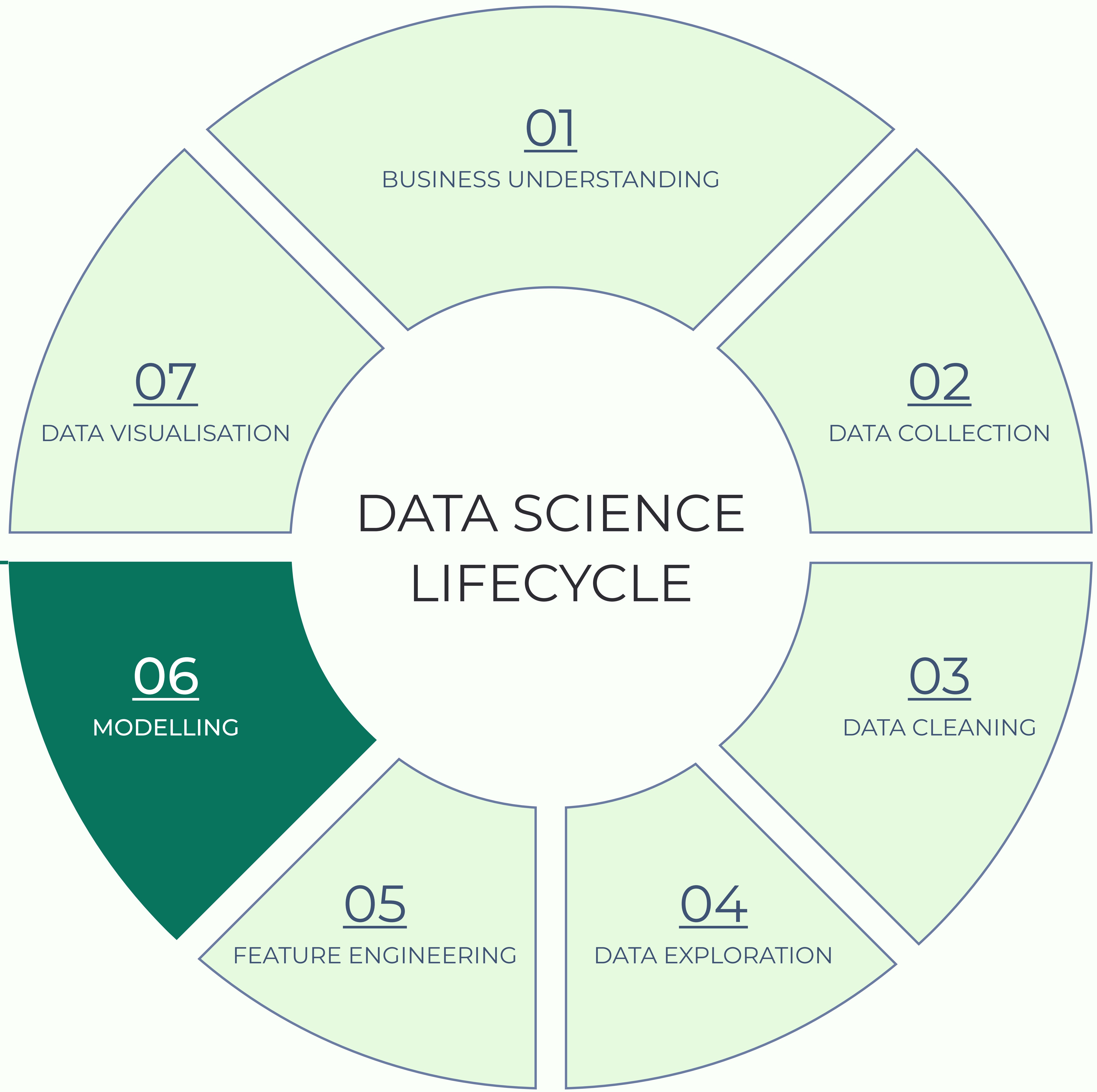
Department of Software Science
Tallinn University of Technology
juri.belikov@taltech.ee

PREVIOUSLY IN COURSE …

Key takeaways (feature):

- Creation
- Scaling
- Transformation
- Extraction
- Selection

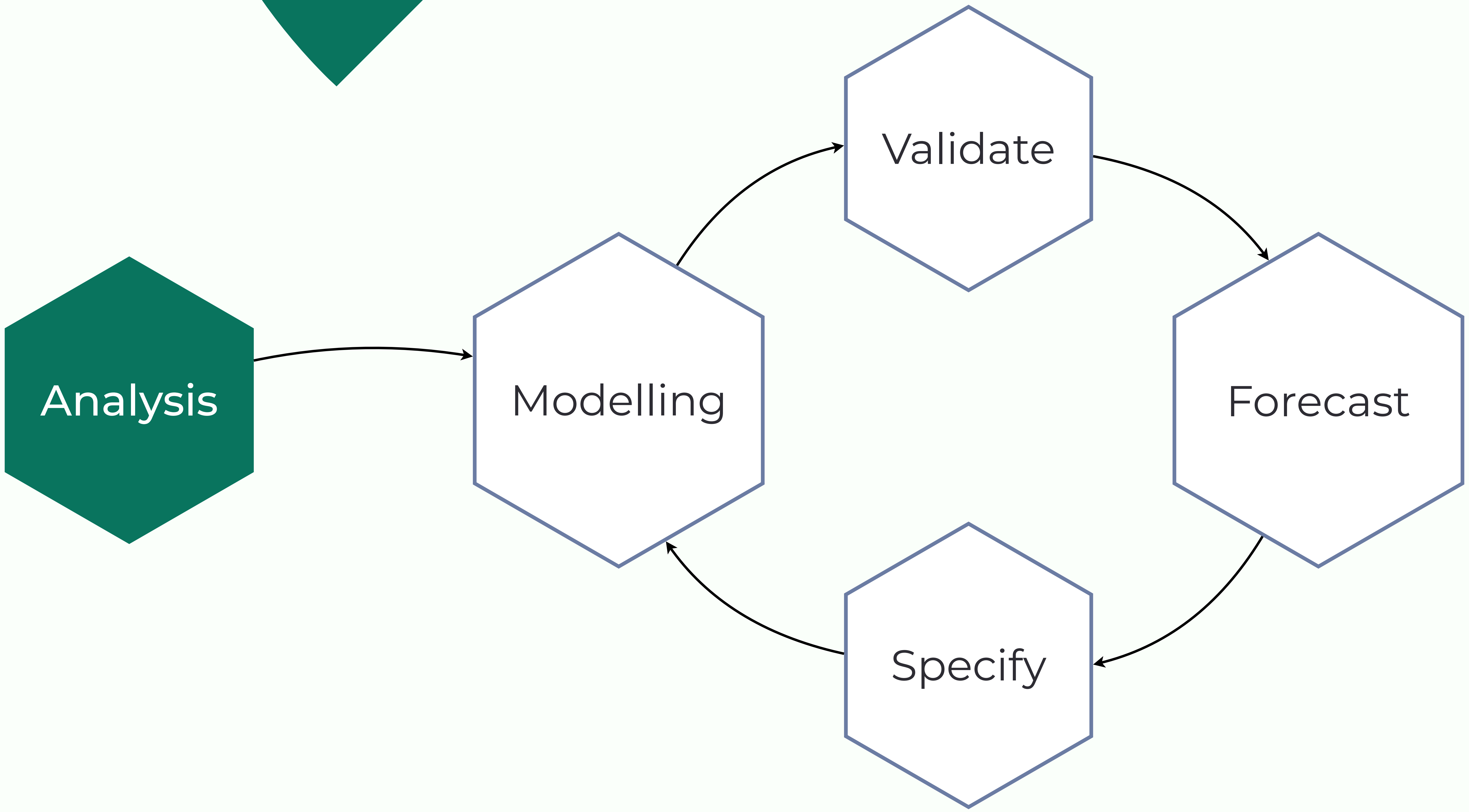
Train ML models, evaluate performance, and make predictions.



Train ML models, evaluate performance, and make predictions.

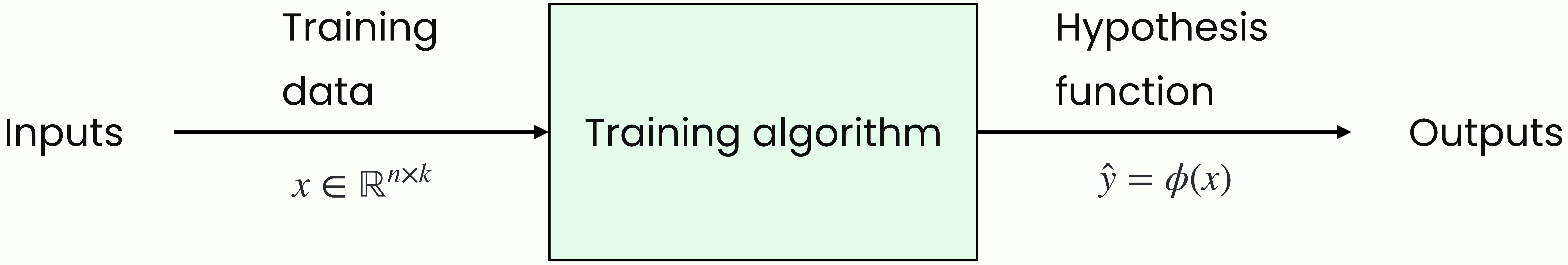
06

MODELLING



MODELLING

Sketchy formalisation from most of ML courses …



TYPICAL MODELS

Suppose our goal is to forecast the hourly electricity demand (y).

Explanatory model may be given as:

$$y = \phi(\text{temperature}, \text{hour}, \text{day of week}, \text{error}).$$

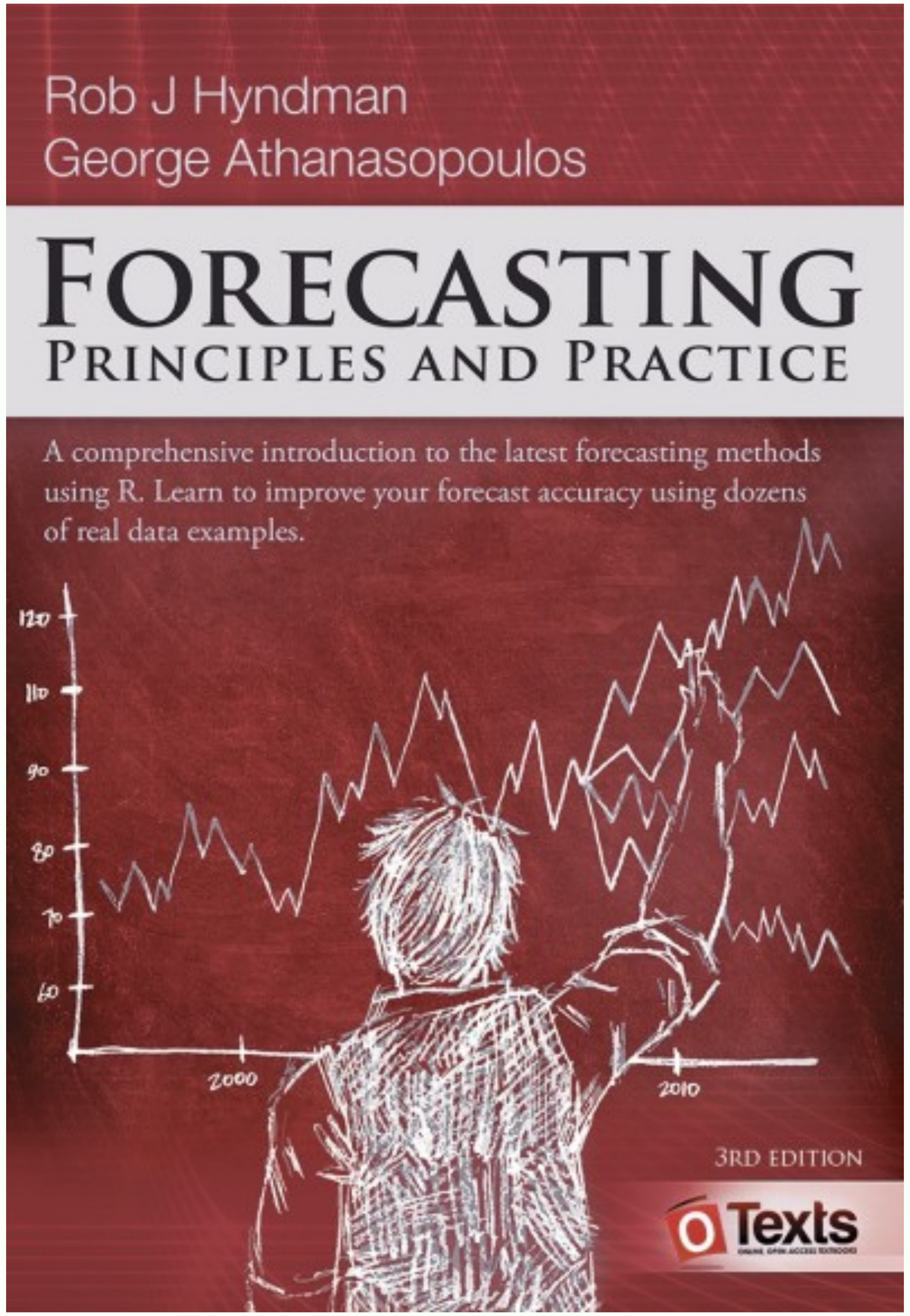
A **time series** model can be written as:

$$y_t = \phi(y_{t-1}, y_{t-2}, \dots, \text{error}).$$

Mixed models are of the form:

$$y_t = \phi(y_{t-1}, \text{temperature}, \text{hour}, \text{day of week}, \text{error}).$$

Time series are different …



“Forecasting: Principles and Practice”

by R. J. Hyndman and G. Athanasopoulus, 3rd ed, OTexts, 2021

WHAT ARE TIME SERIES?

Time series:

Set of ordered data values consistently collected over an interval of time.

Time series depend on time *t*

- Weather data (temperature, wind, humidity, etc)
- Number of sold cars
- Stock prices

WHAT ARE TIME SERIES?

Time series:

Set of ordered data values consistently collected over an interval of time.

- Weather data (temperature, wind, humidity, etc)
- Number of sold cars
- Stock prices

Time series depend on time *t*

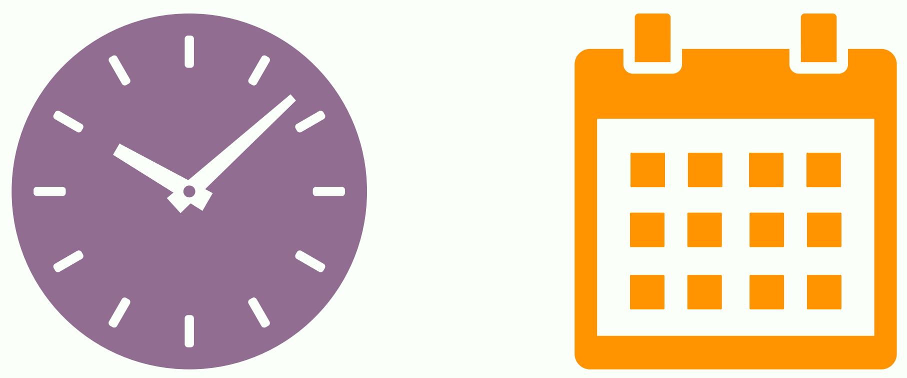
Cross-sectional:

Set of data values collected at a fixed point in time, or where time is of no significance.

WHAT ARE TIME SERIES?

Time series:

Set of ordered data values consistently collected over an interval of time.



Time series depend on time *t*

- Weather data (temperature, wind, humidity, etc)
- Number of sold cars
- Stock prices

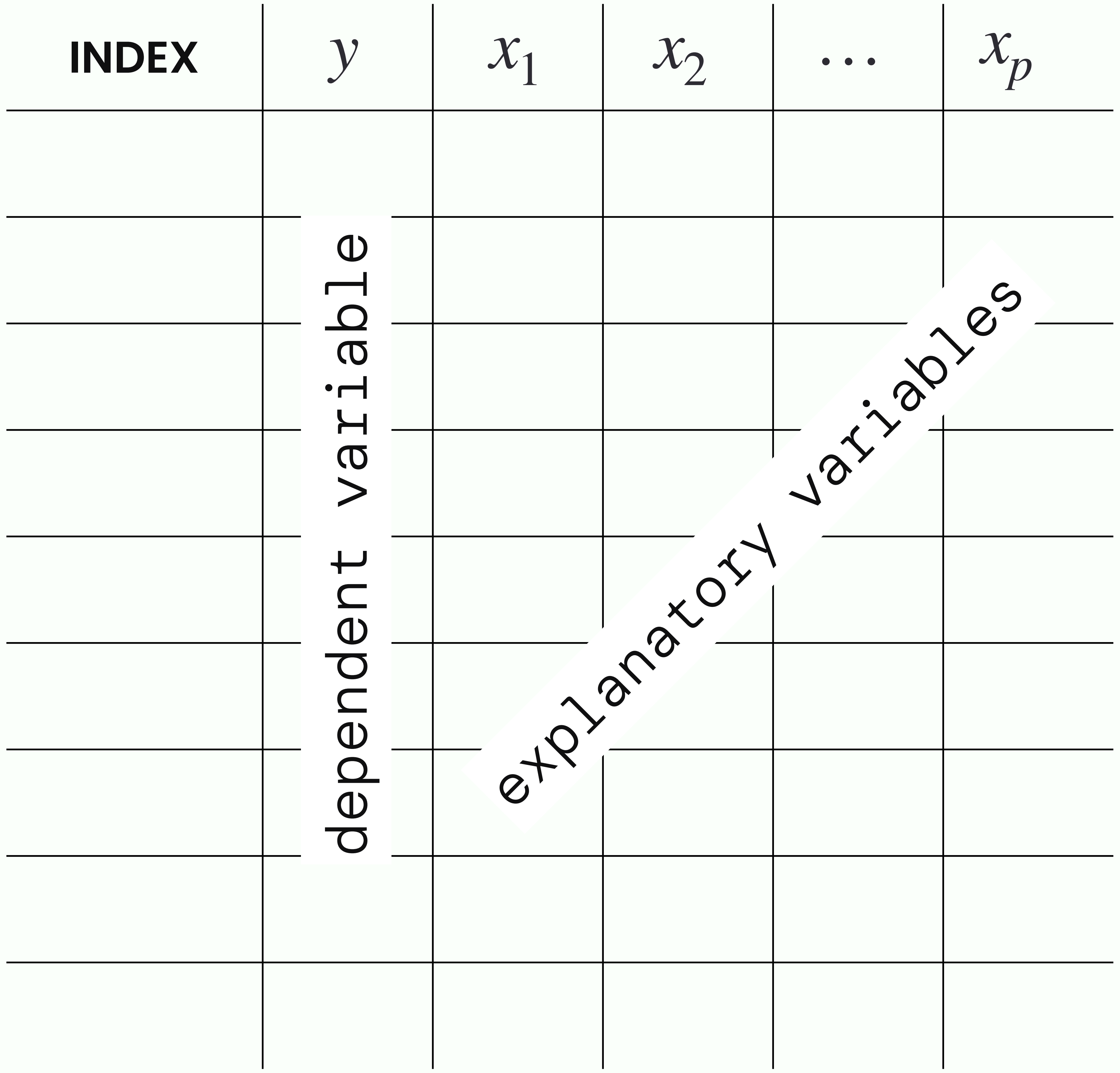
Cross-sectional:

Set of data values collected at a fixed point in time, or where time is of no significance.

Panel data: combines both TS and CS.

DATA STRUCTURE MATTERS

Multiple regression-like setup:

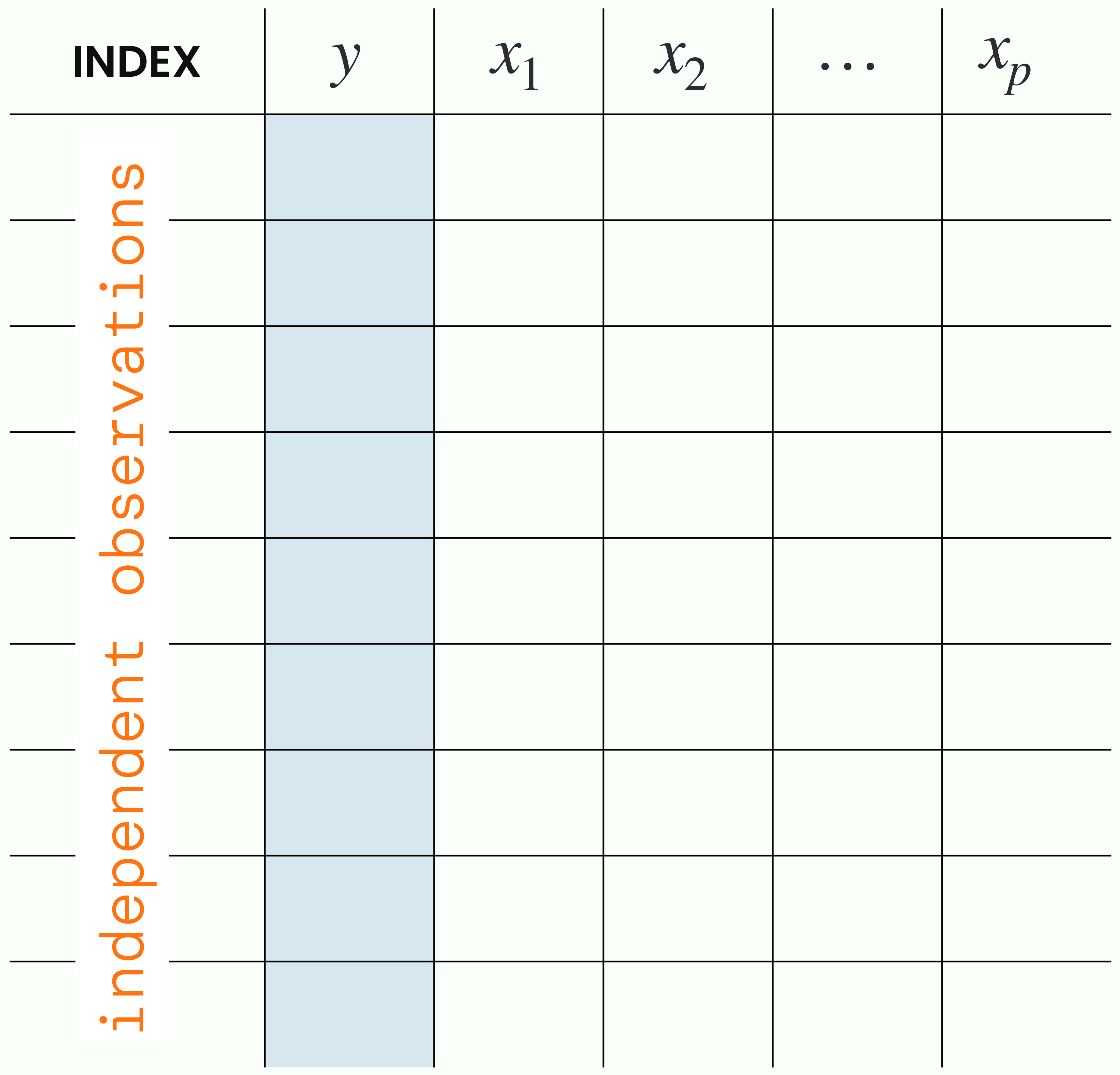


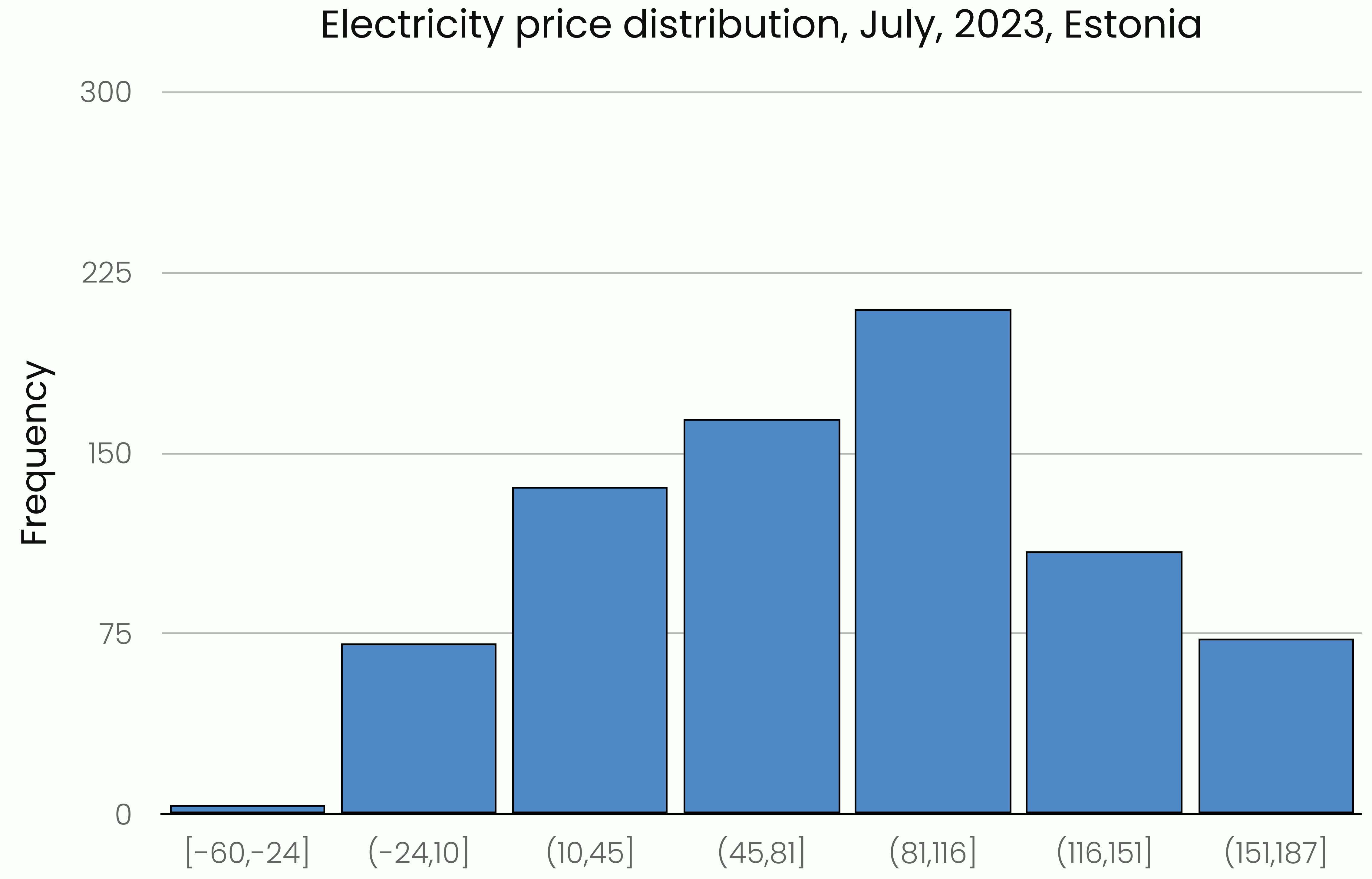
CROSS-SECTIONAL DATA

INDEX	y	x_1	x_2	$\dots$	x_p

independent observations

CROSS-SECTIONAL DATA (2)



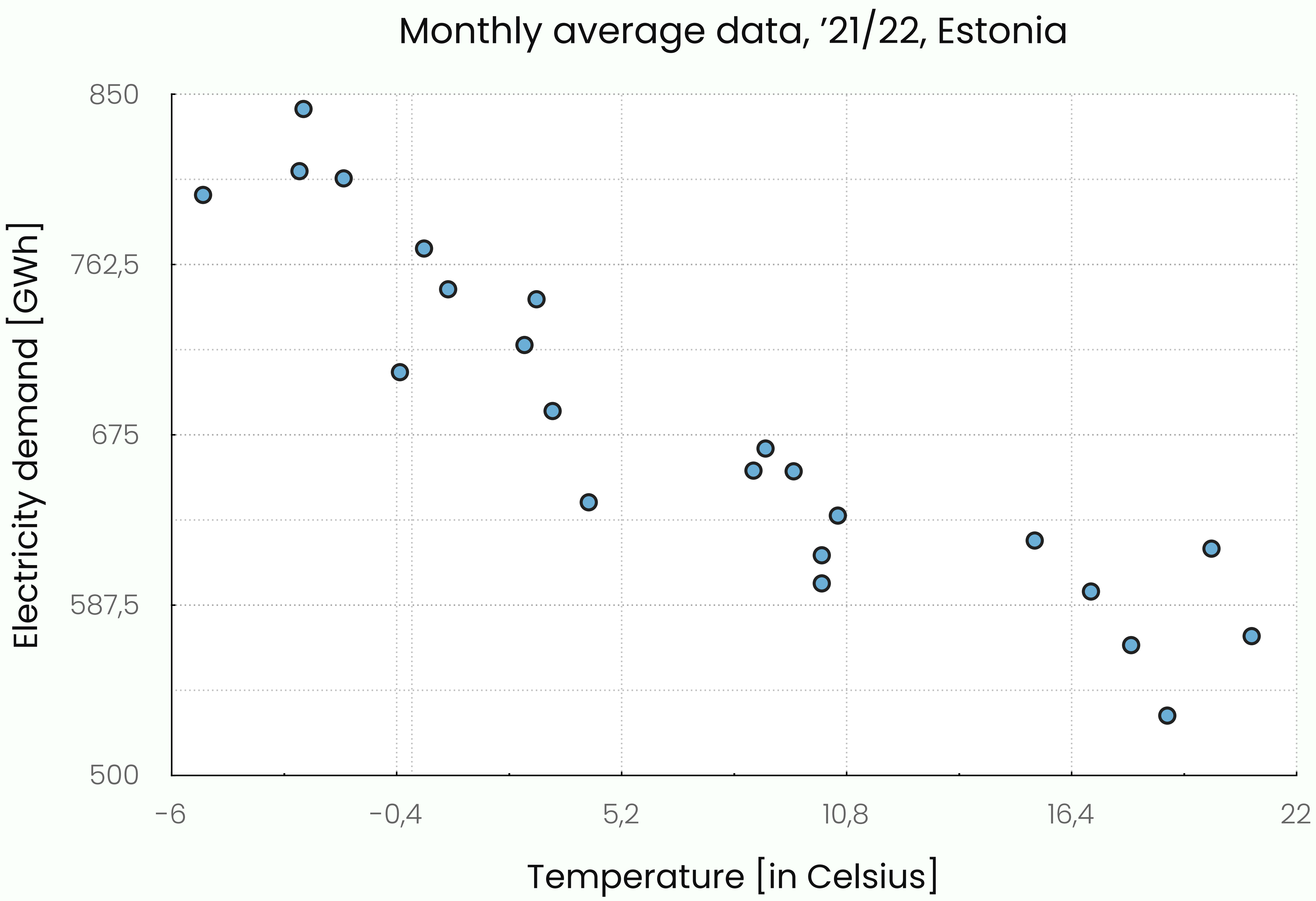


Observations are grouped, but not ordered.

CROSS-SECTIONAL DATA (3)

INDEX	y	x_1	x_2	$\dots$	x_p

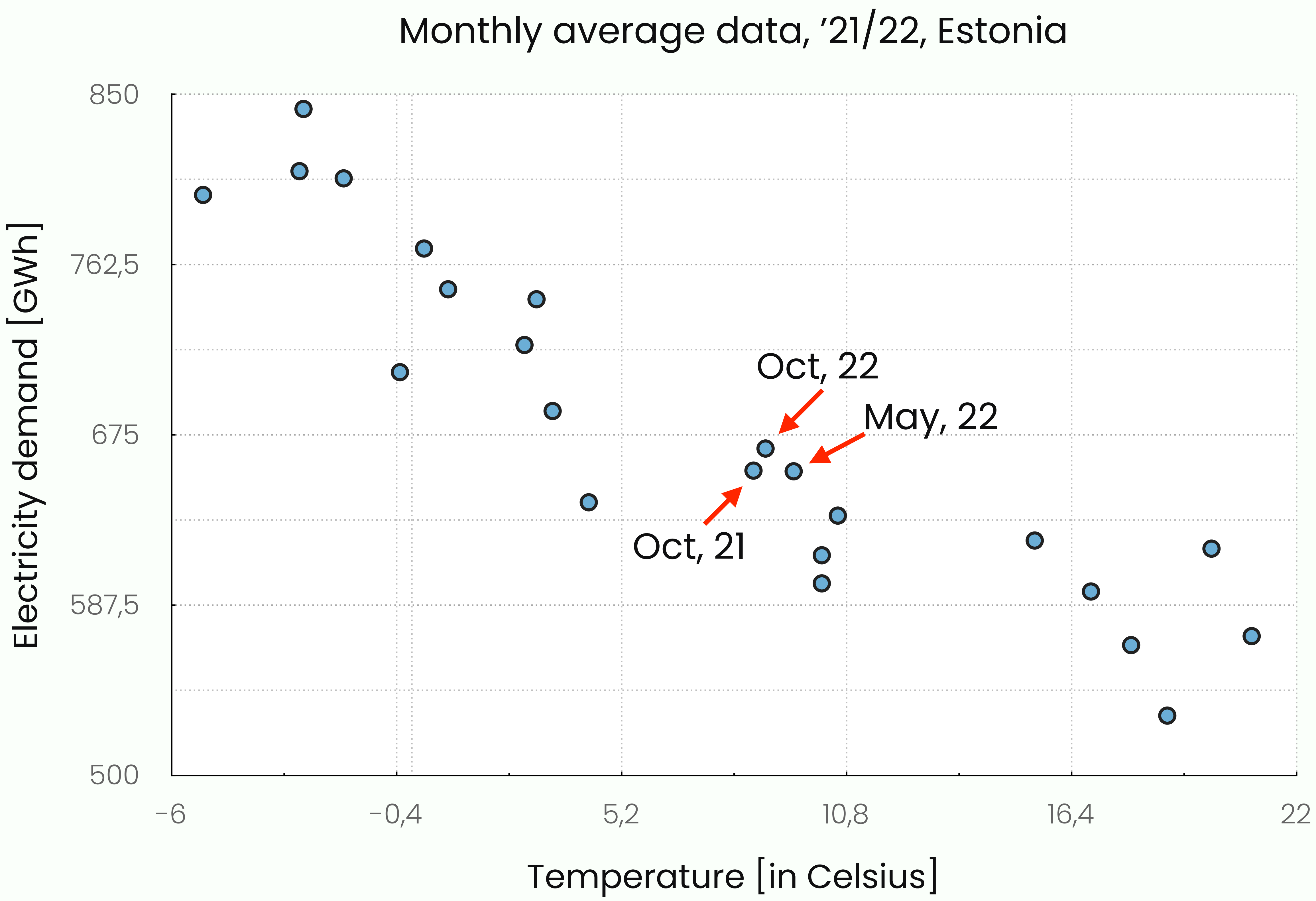
independent observations



CROSS-SECTIONAL DATA (3)

INDEX	y	x_1	x_2	$\dots$	x_p

independent observations



TIME SERIES DATA

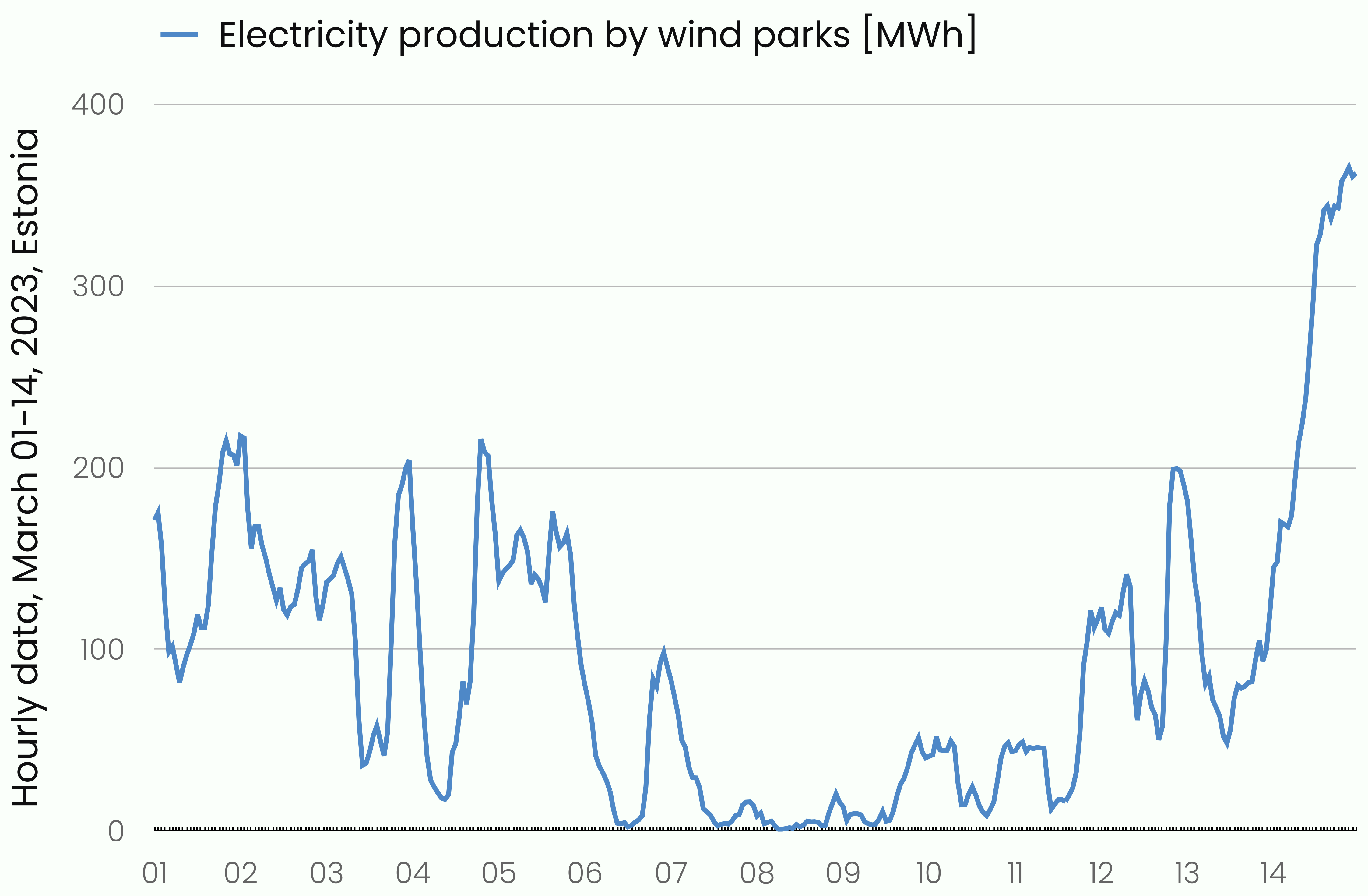
INDEX	y	x_1	x_2	$\ldots$	x_p

TIME DEPENDENT

TIME SERIES DATA (2)

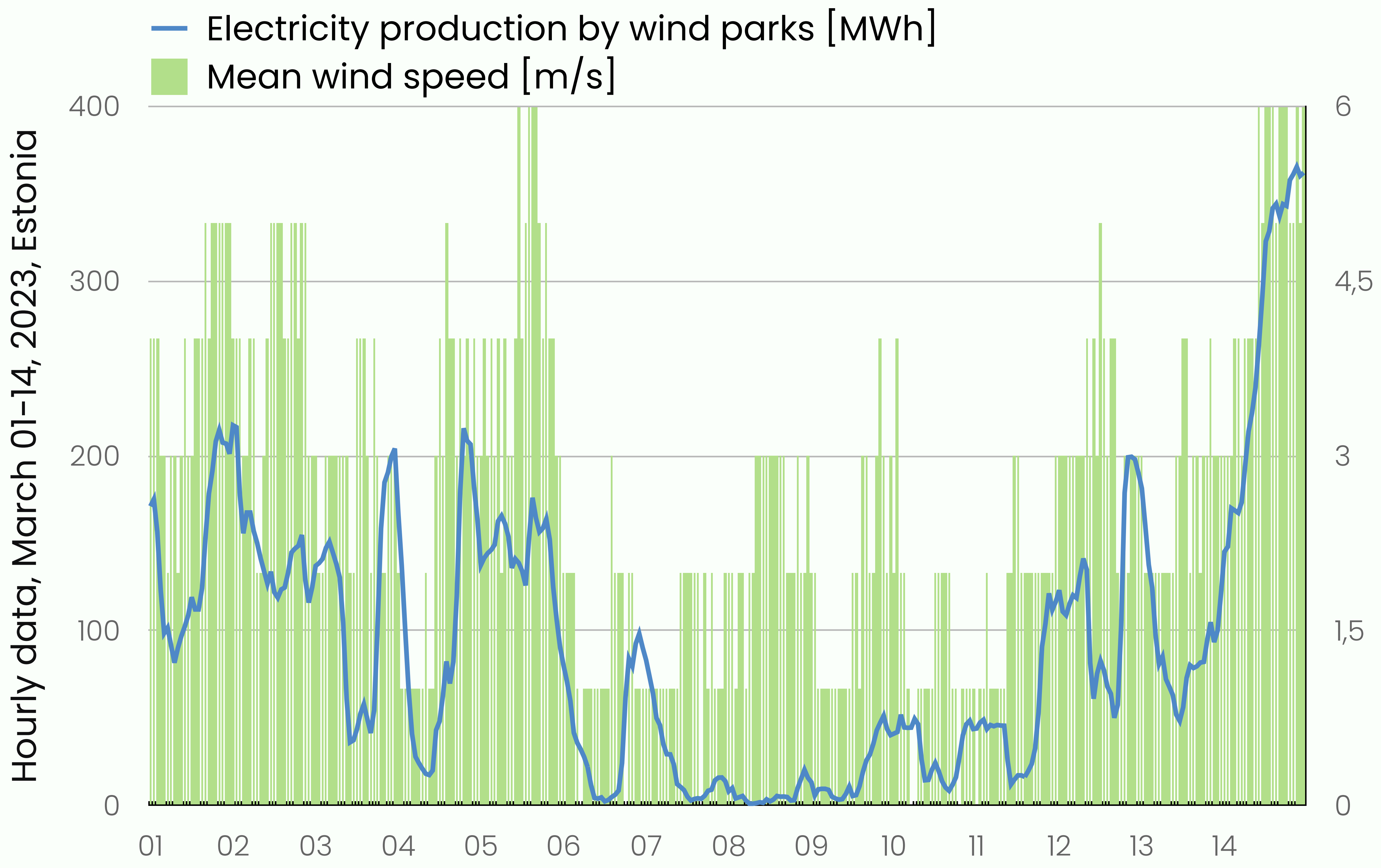
INDEX	y	x_1	x_2	$\dots$	x_p

TIME DEPENDENT



TIME SERIES DATA (3)

INDEX	y	x_1	x_2	$\dots$	x_p
TIME DEPENDENT					

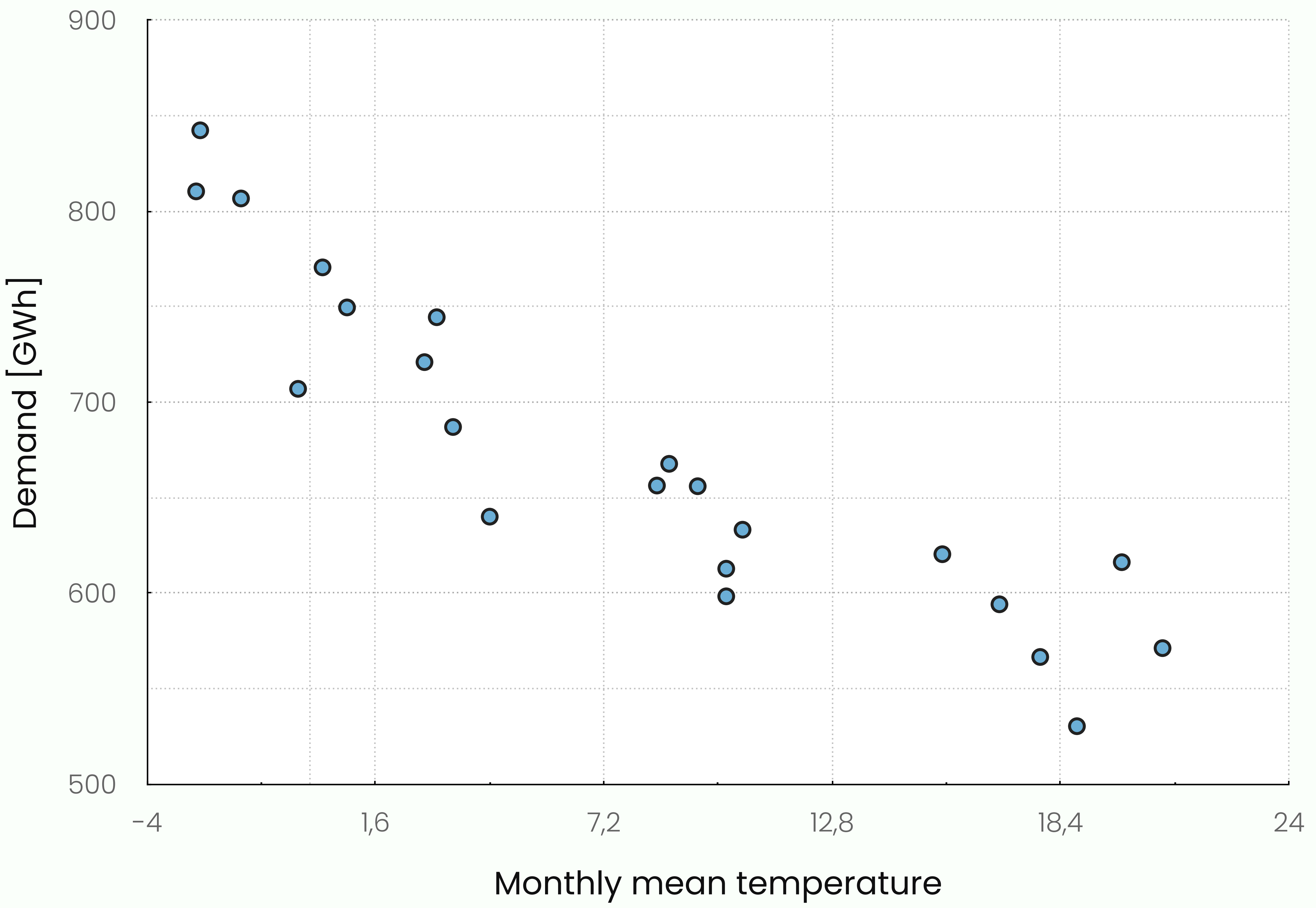


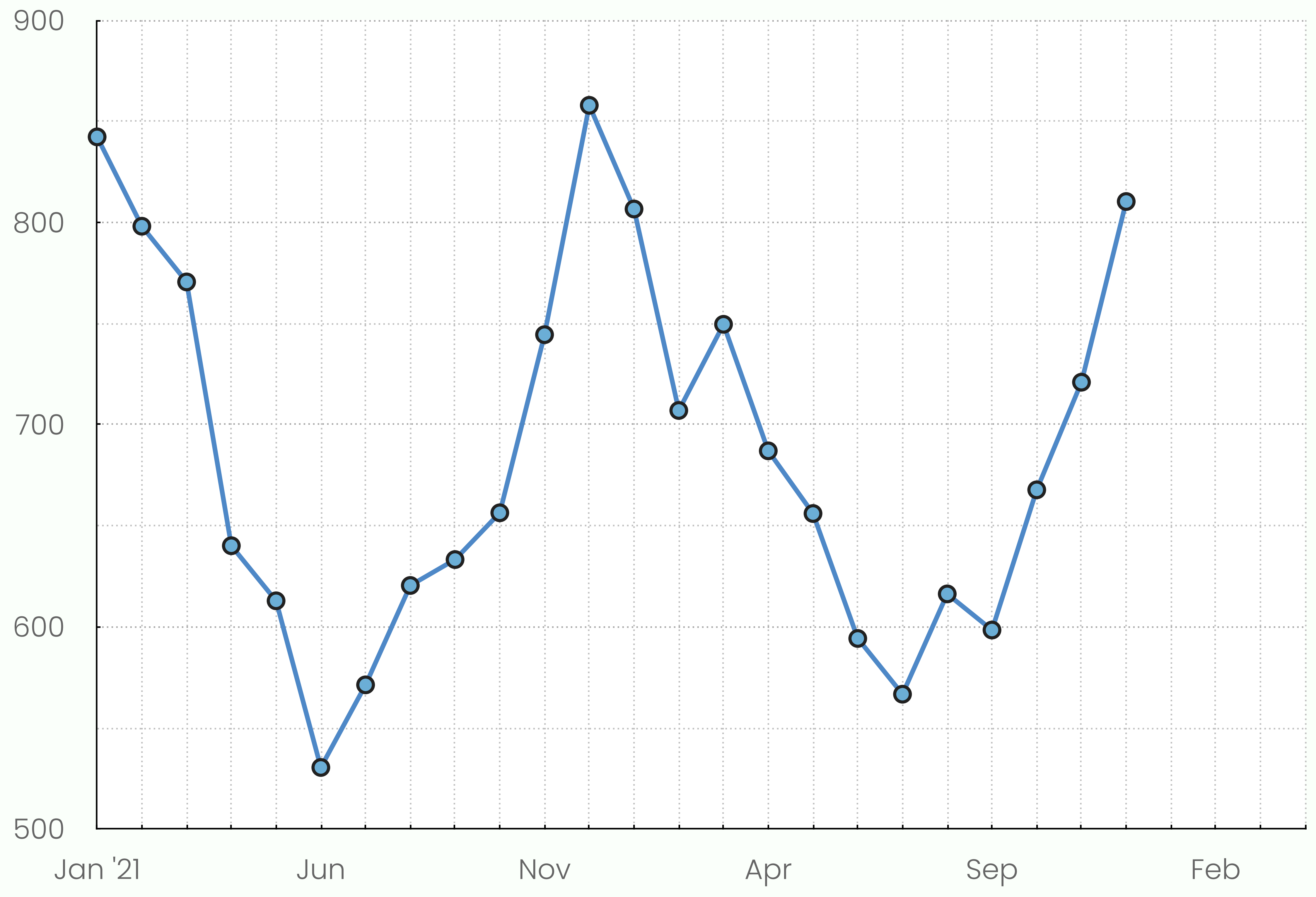
Another perspective

TIME SERIES: ANOTHER LOOK

Setup: Given monthly average electricity demand and temperature data for 21/22 in Estonia.

Goal: We want to understand what will be energy demand for a given month.

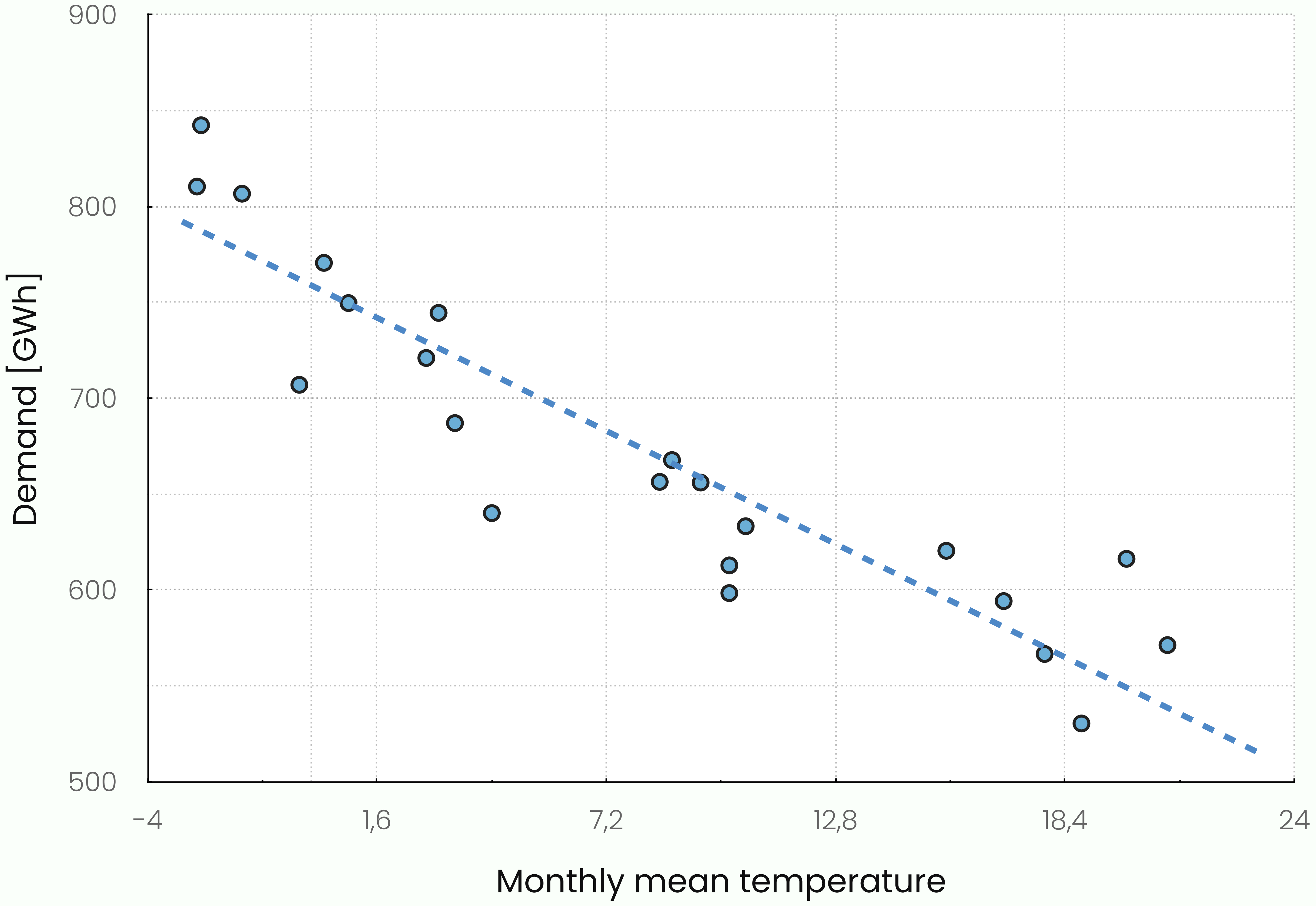


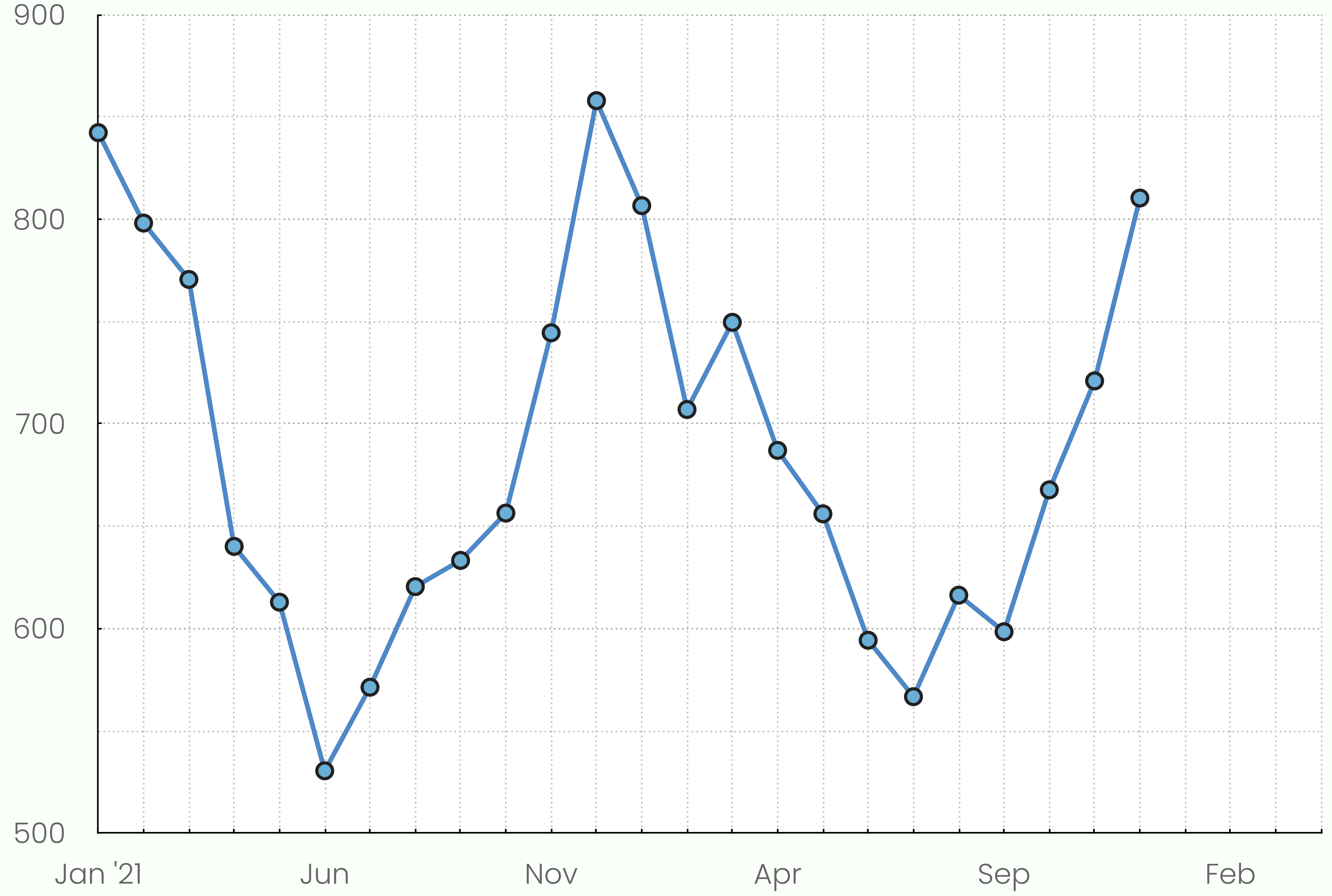


TIME SERIES: ANOTHER LOOK

Setup: Given monthly average electricity demand and temperature data for 21/22 in Estonia.

Goal: We want to understand what will be energy demand for a given month.

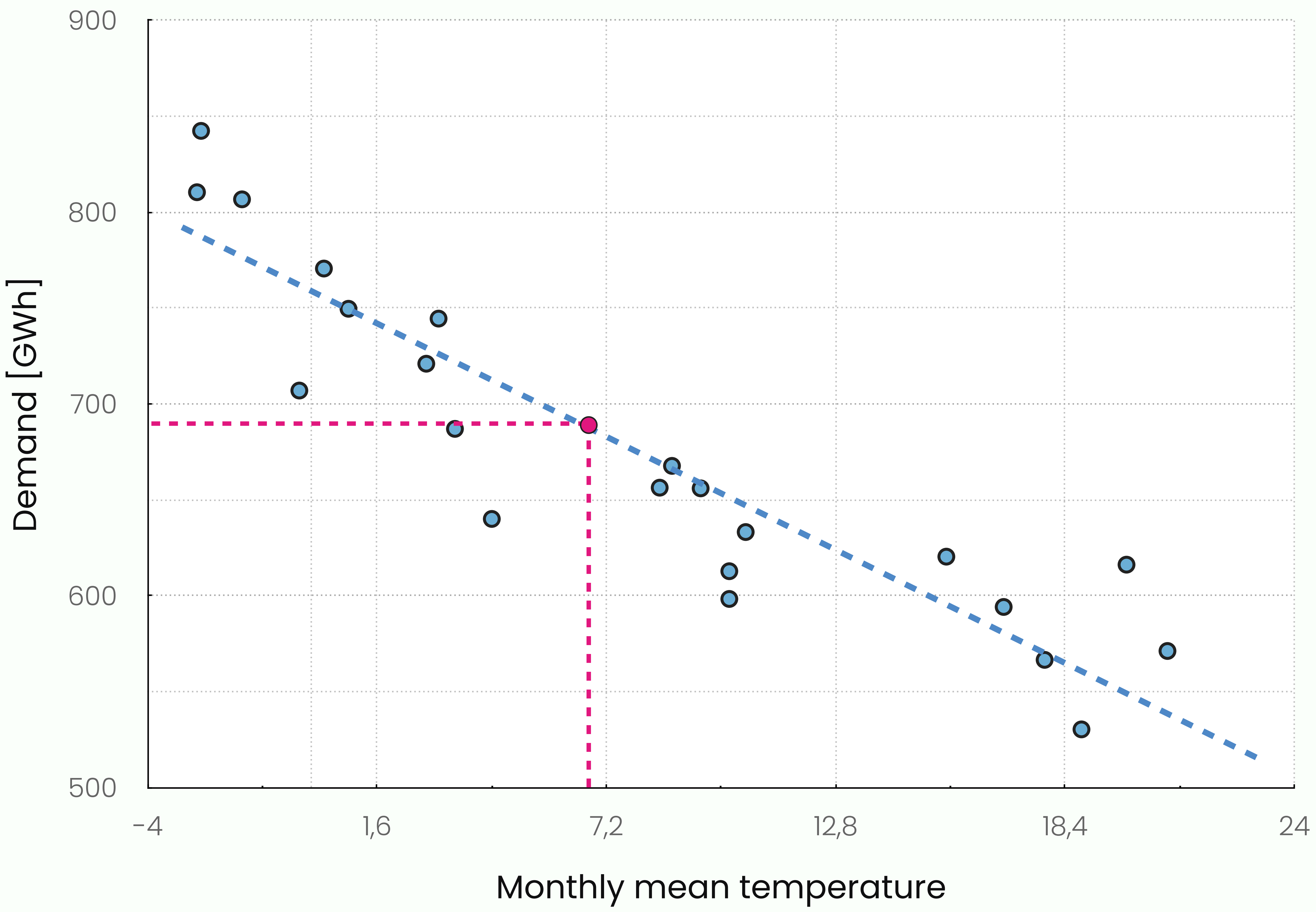


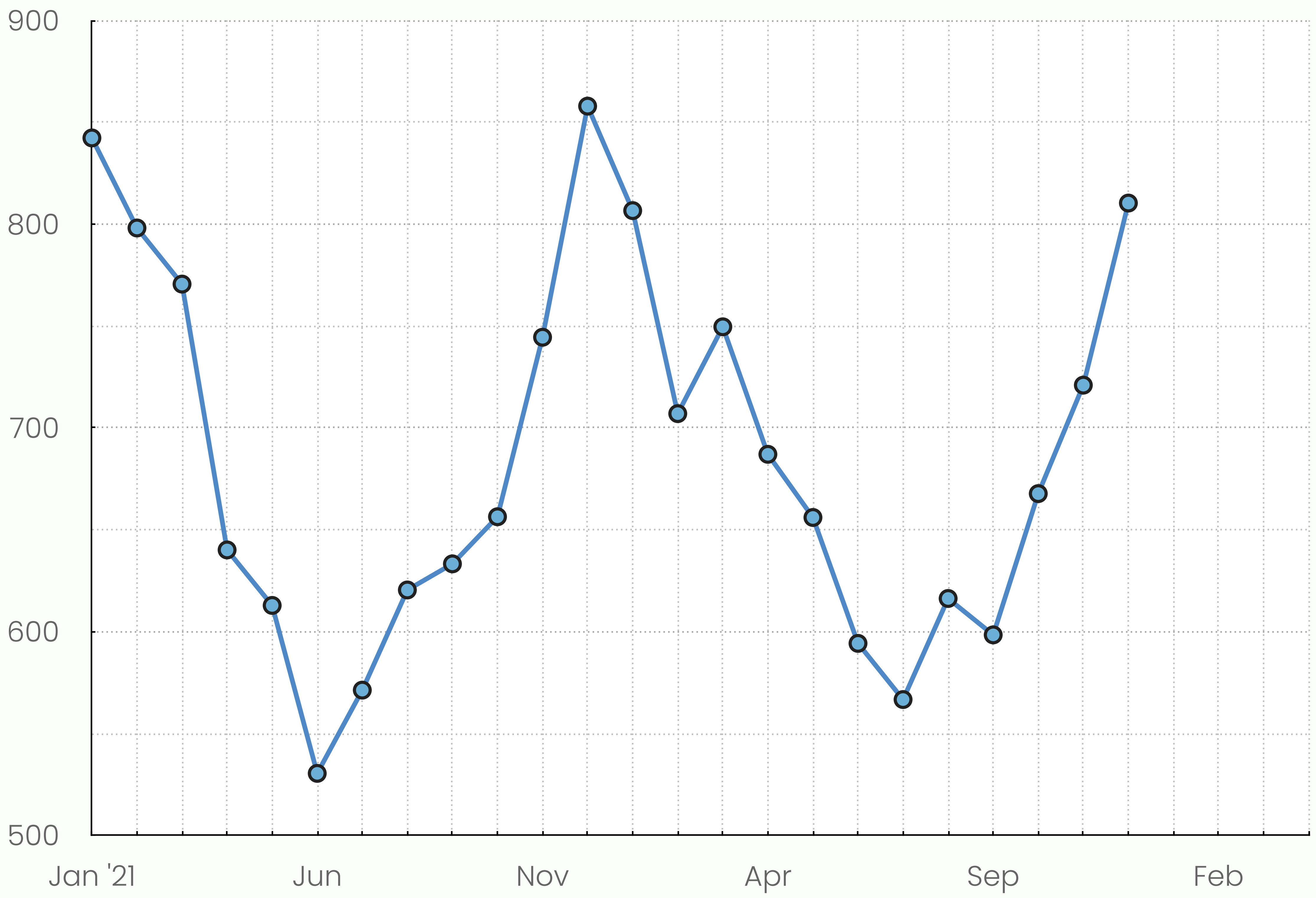


TIME SERIES: ANOTHER LOOK

Setup: Given monthly average electricity demand and temperature data for 21/22 in Estonia.

Goal: We want to understand what will be energy demand for a given month.

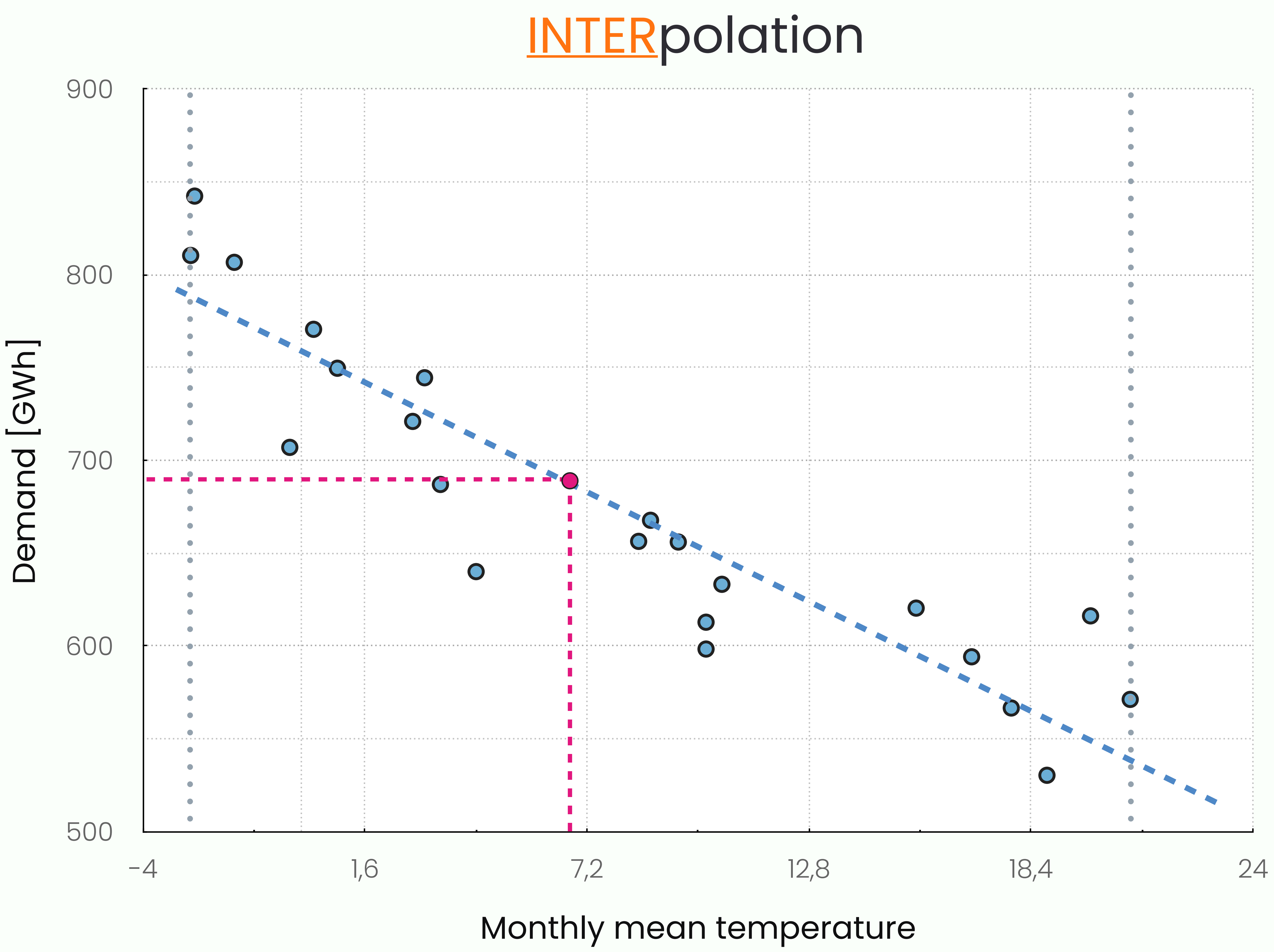


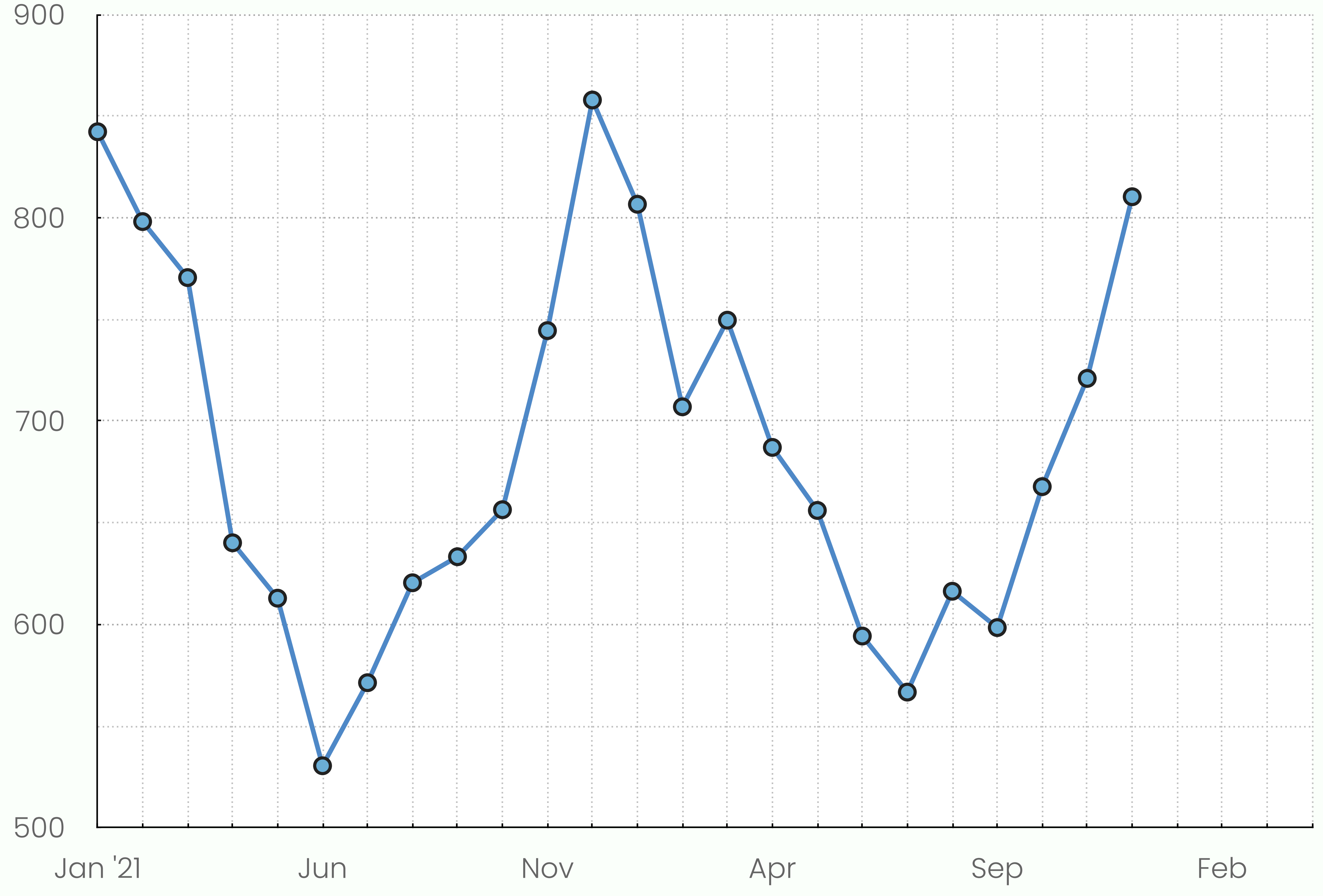


TIME SERIES: ANOTHER LOOK

Setup: Given monthly average electricity demand and temperature data for 21/22 in Estonia.

Goal: We want to understand what will be energy demand for a given month.

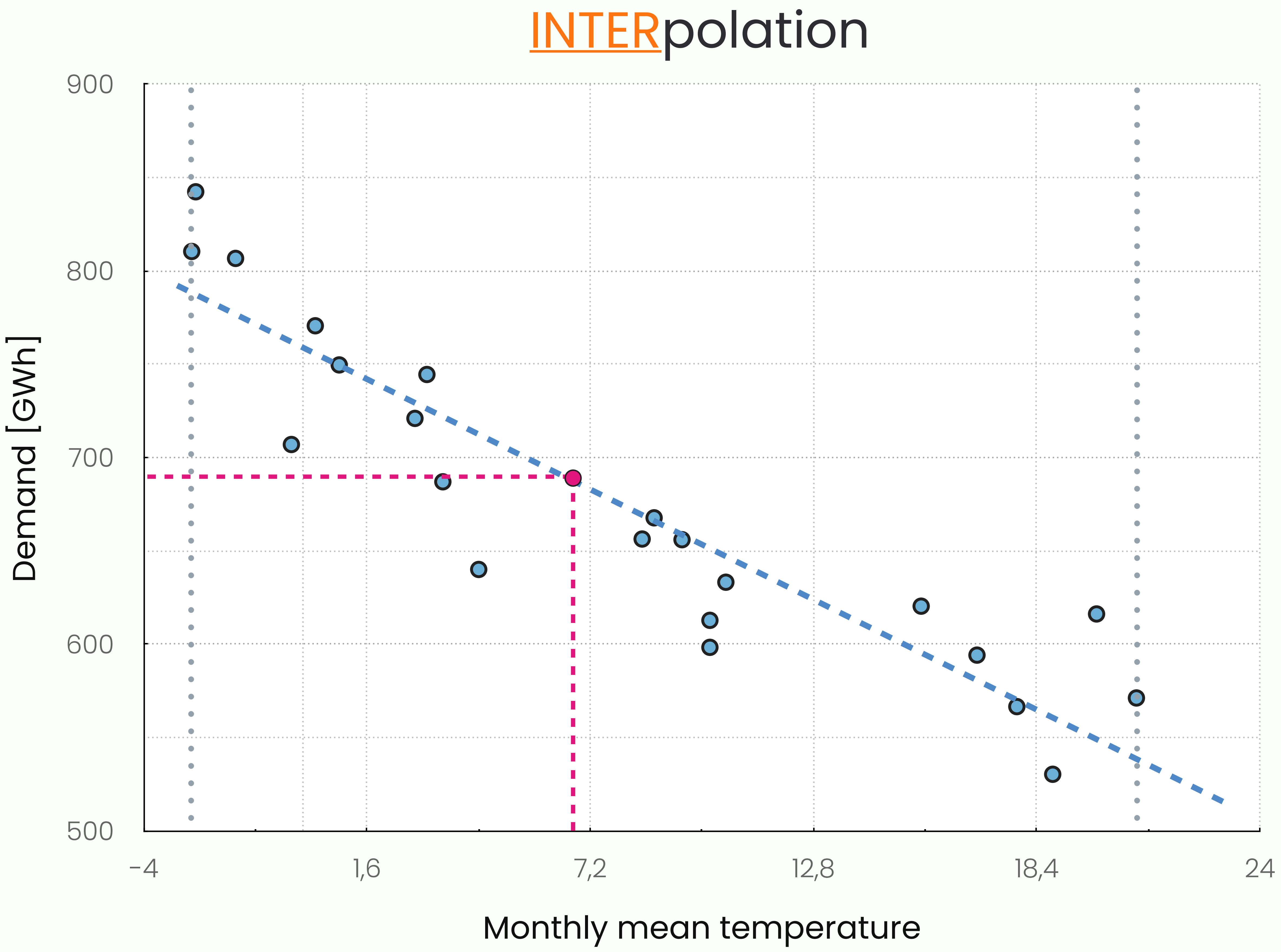


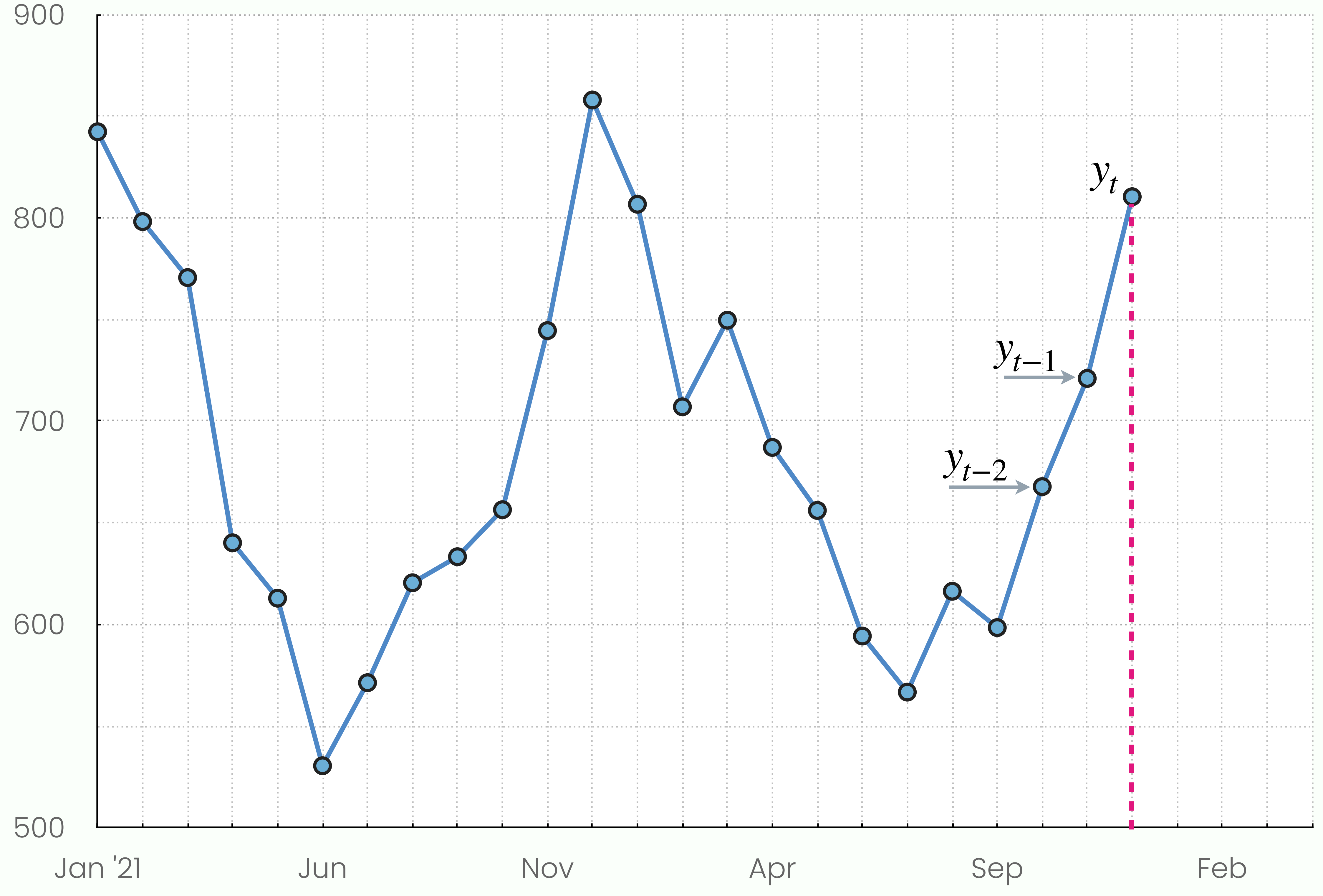


TIME SERIES: ANOTHER LOOK

Setup: Given monthly average electricity demand and temperature data for 21/22 in Estonia.

Goal: We want to understand what will be energy demand for a given month.

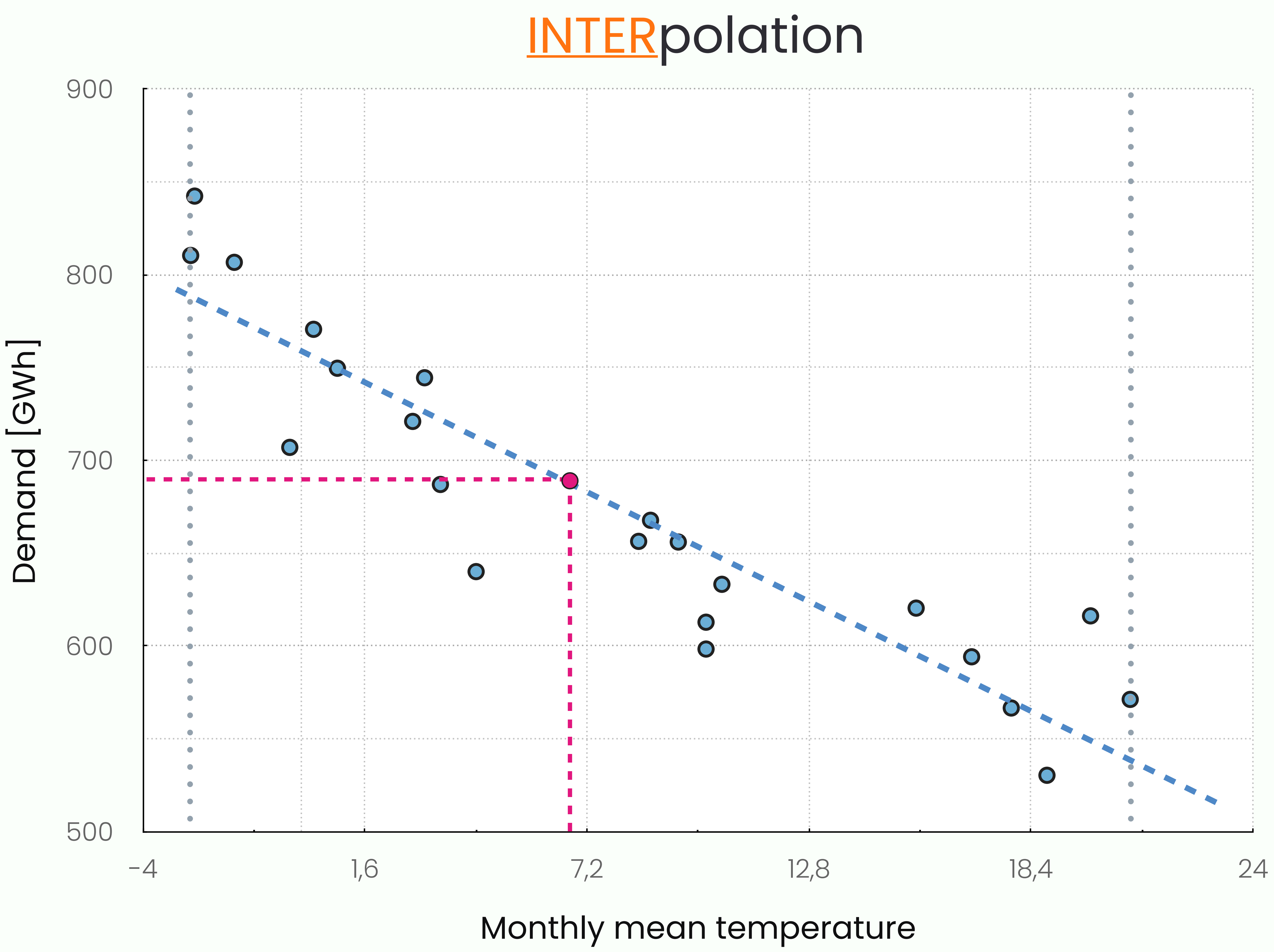


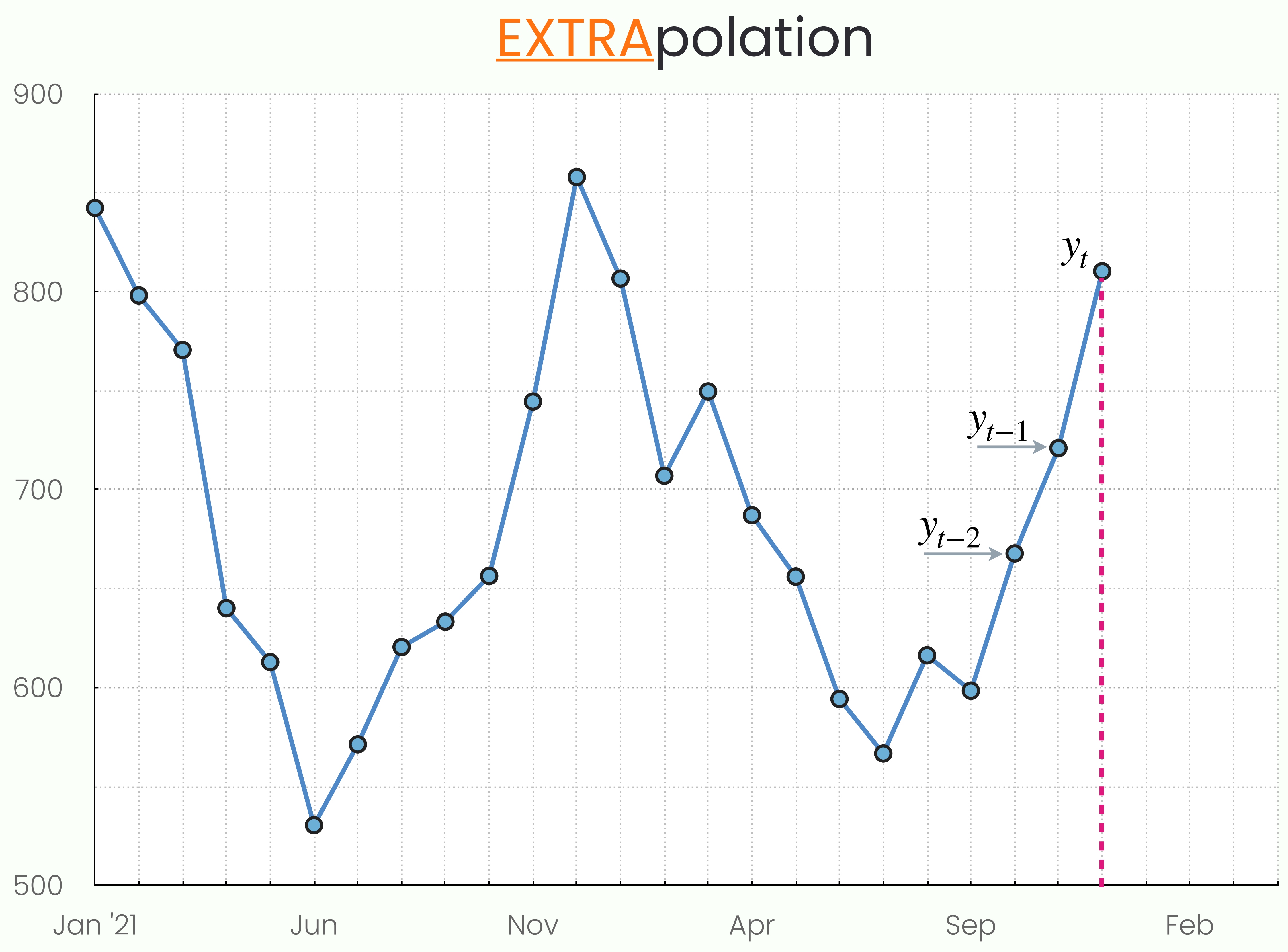


TIME SERIES: ANOTHER LOOK

Setup: Given monthly average electricity demand and temperature data for 21/22 in Estonia.

Goal: We want to understand what will be energy demand for a given month.

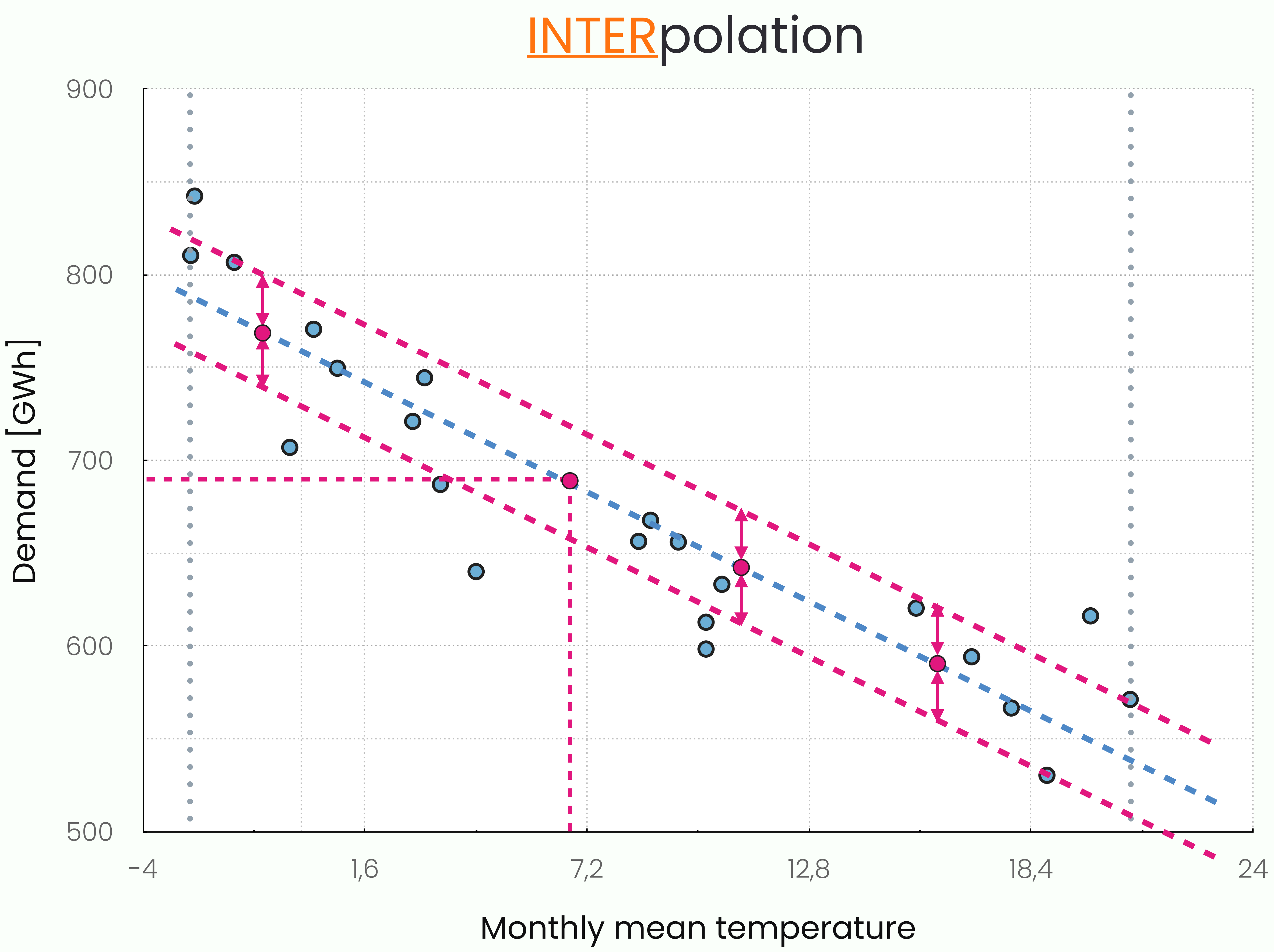


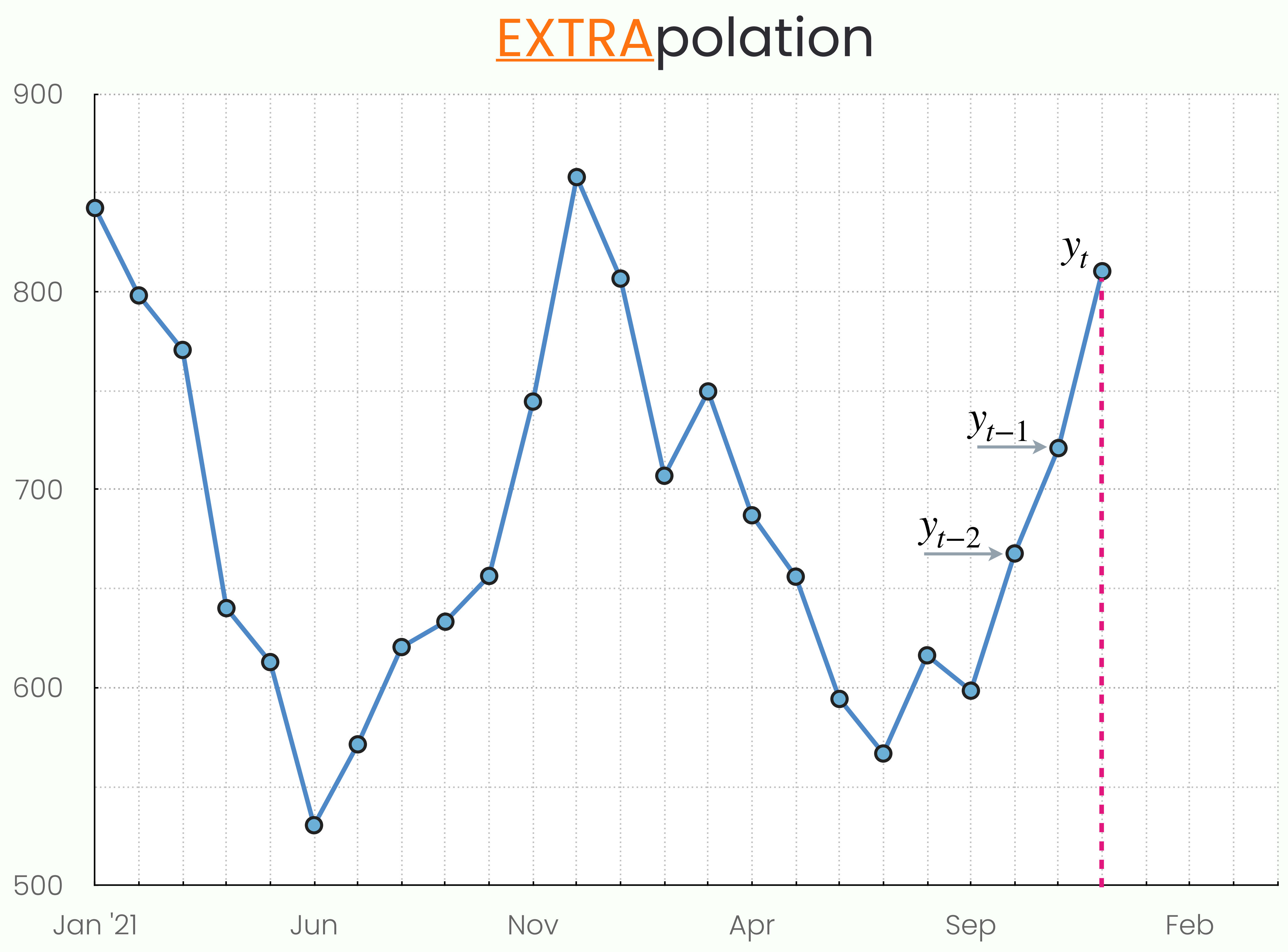


TIME SERIES: ANOTHER LOOK

Setup: Given monthly average electricity demand and temperature data for 21/22 in Estonia.

Goal: We want to understand what will be energy demand for a given month.

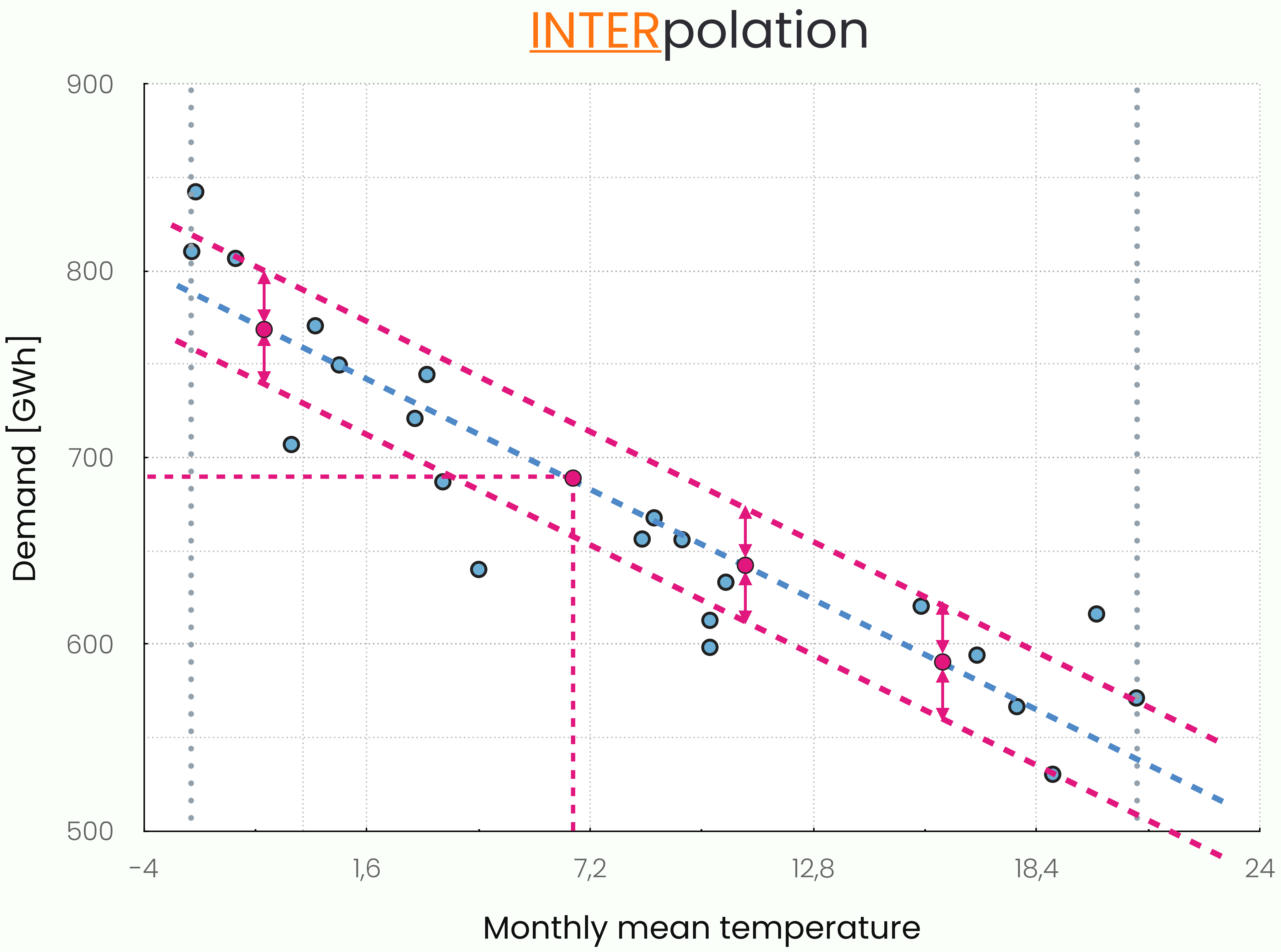


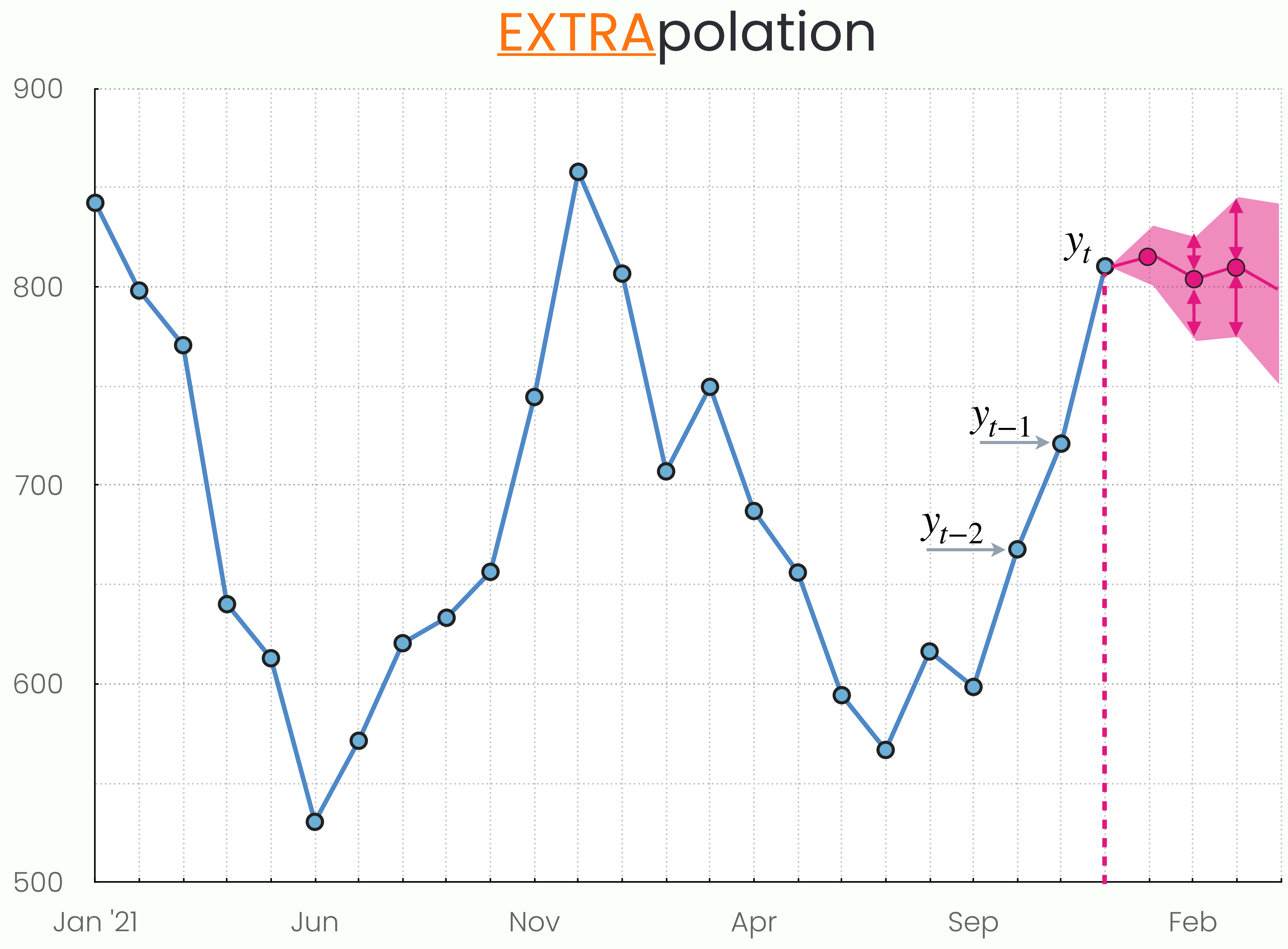


TIME SERIES: ANOTHER LOOK

Setup: Given monthly average electricity demand and temperature data for 21/22 in Estonia.

Goal: We want to understand what will be energy demand for a given month.

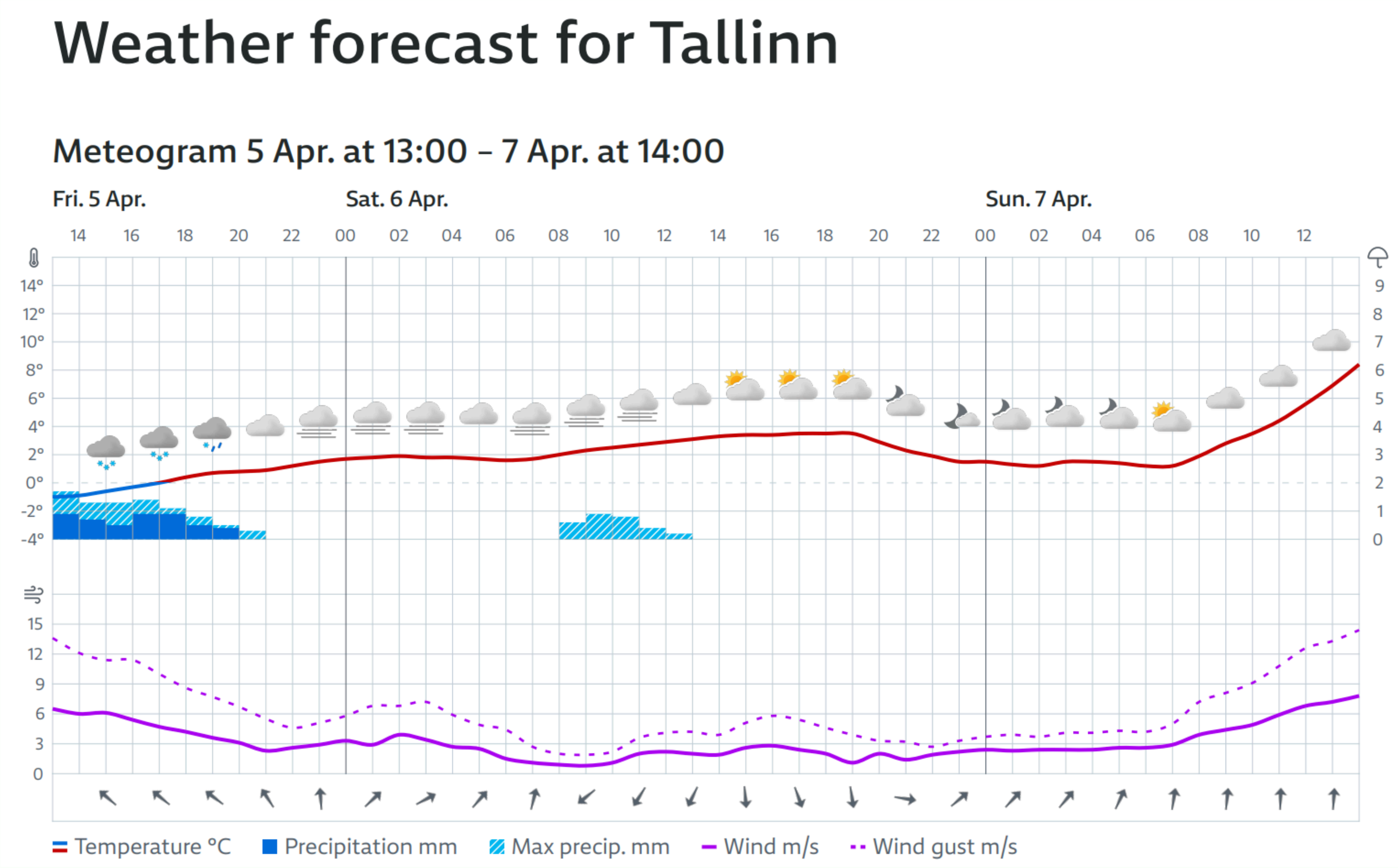




TIME SERIES: EXAMPLES

https://finance.yahoo.com/

TIME SERIES: EXAMPLES (2)

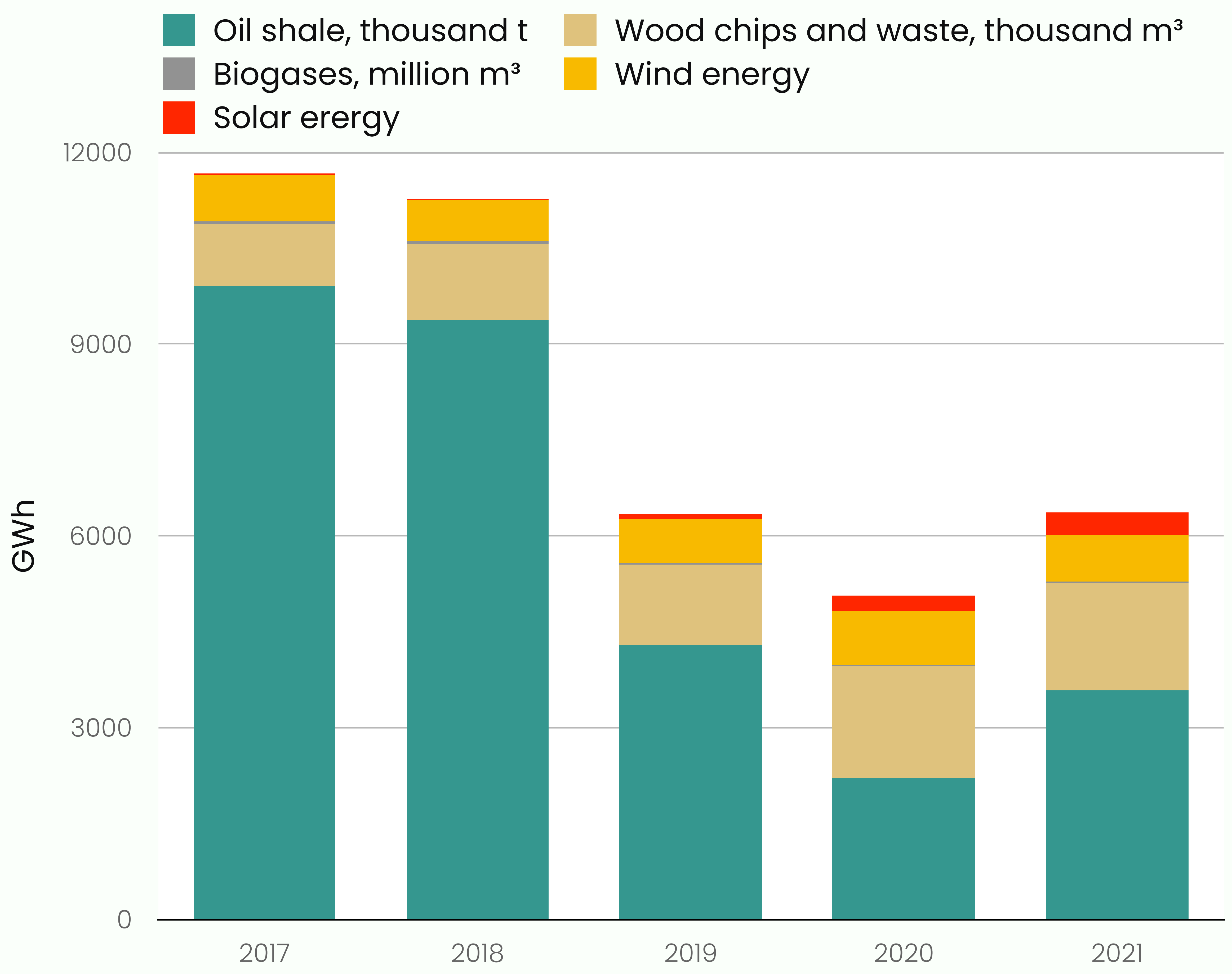


TIME SERIES: EXAMPLES (3)

		2023 Q1	2023 Q2	2023 Q3	2023 Q4	2024 Q1	2024 Q2	2024 Q3	2024 Q4	2025 Q1	2025 Q2
Average monthly gross wages (salaries), euros	Electricity, gas, steam and air conditioning supply	2,435	2,946	2,376	2,548	2,496	3,138	2,540	2,619	2,714	3,164
	Information and communication	3,160	3,257	3,222	3,271	3,393	3,468	3,506	3,571	3,605	3,704
	Financial and insurance activities	2,952	2,953	2,813	2,902	3,114	3,163	3,068	3,164	3,344	3,389
Median (5th decile) of monthly gross wages (salaries), euros	Electricity, gas, steam and air conditioning supply	2,066	2,328	2,100	2,164	2,125	2,614	2,253	2,334	2,331	2,625
	Information and communication	2,661	2,768	2,728	2,812	2,873	2,955	2,996	3,100	3,000	3,178
	Financial and insurance activities	2,395	2,450	2,353	2,400	2,555	2,682	2,594	2,650	2,800	2,900

https://www.stat.ee/en/avasta-statistikat/valdkonnad/tooelu/wages-and-salaries-and-labour-costs

EXAMPLES FROM ENERGY DOMAIN

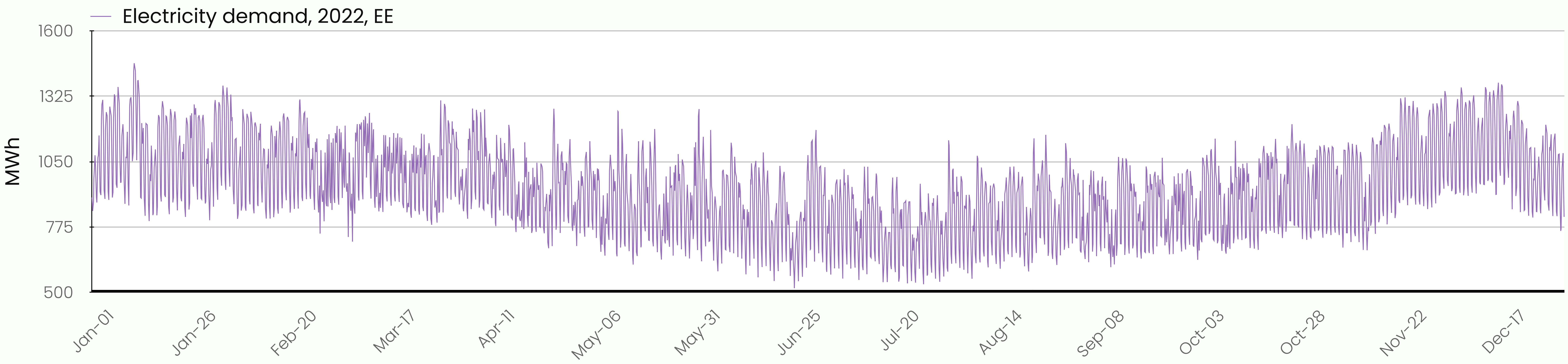


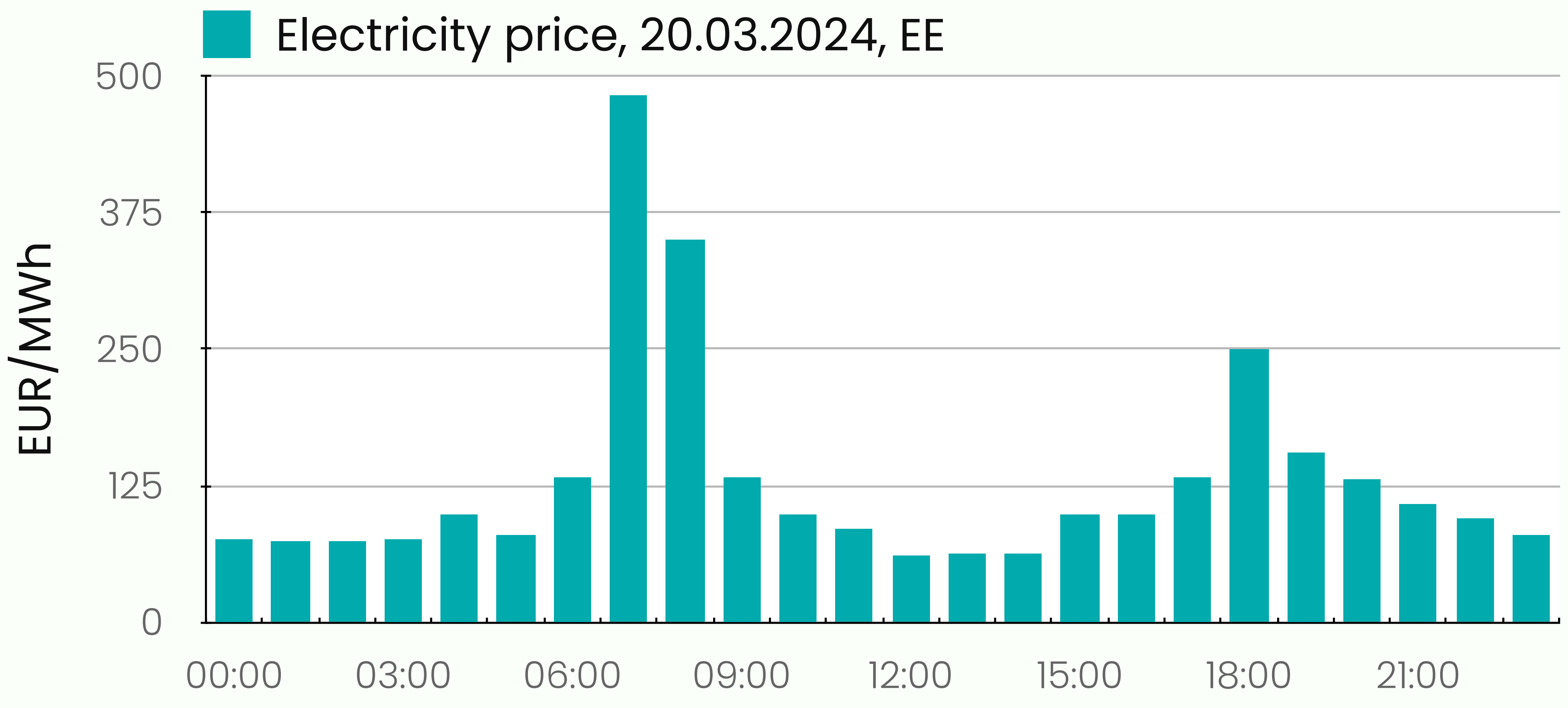
Average cost of fuels and energy consumed by enterprises (EUR)

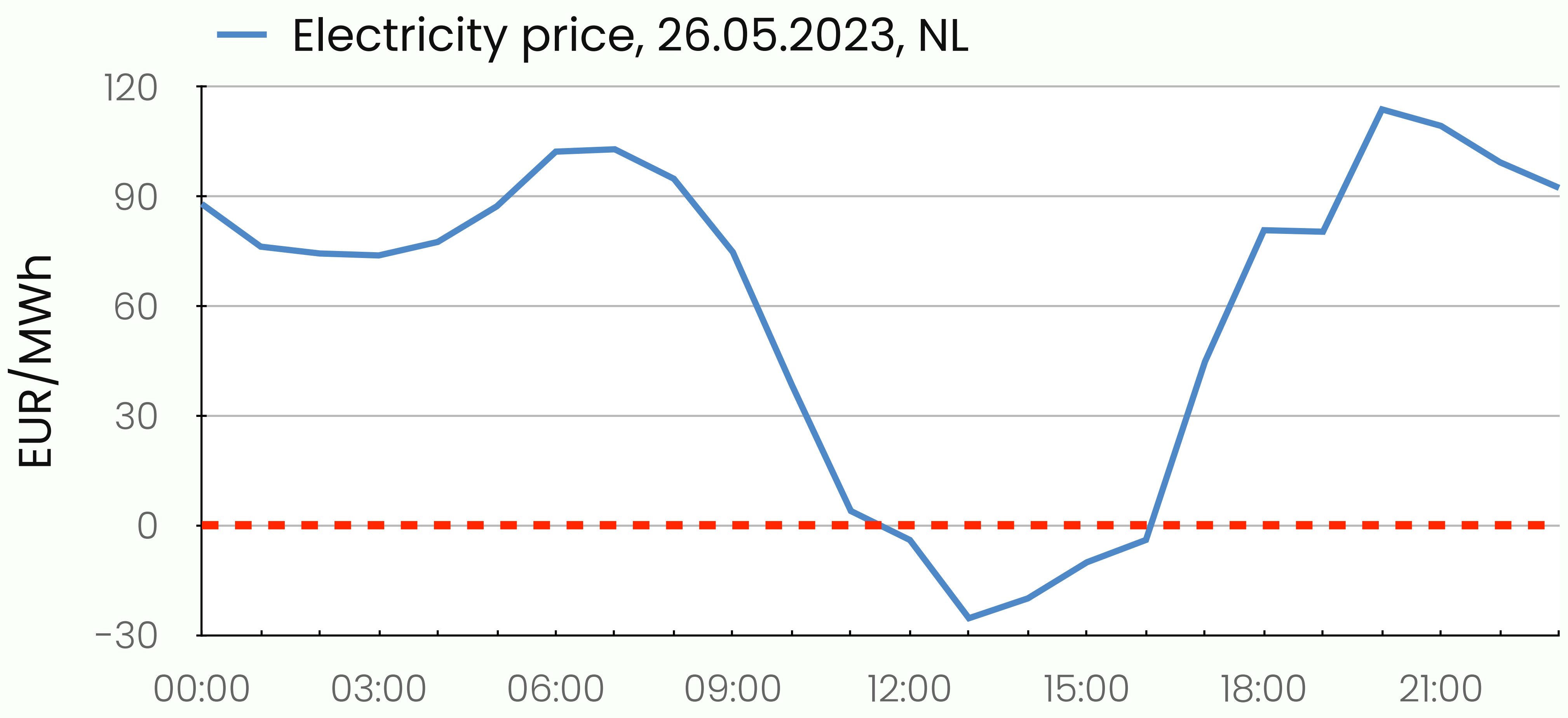
	2017	2018	2019	2020	2021
Oil shale, euros/ton	12.79	13.01	14.26	14.53	17.02
Wood chips, euros/m³	11.23	11.31	13.47	12.34	11.63
Natural gas, euros/1000 m³	289.77	325.45	328.09	262.42	422.69
Light fuel oil, euros/ton	601.28	586.91	572.97	425.02	495.92
Diesel, euros/ton	1032.75	1051.39	1074.62	950.52	1062.79
Electricity, euros/MWh	76.51	76.69	77.12	75.83	91.01
Heat, euros/MWh	57.12	65.01	64.26	57.72	60.73

https://www.stat.ee/en/find-statistics/statistics-theme/energy-and-transport/energy

EXAMPLES FROM ENERGY DOMAIN (2)







TERMINOLOGY

Univariate vs multivariate

A time series is called univariate if it contains data of a single variable, and multivariate otherwise.

Linear vs nonlinear

A time series model is said to be (non)linear if relation between past observations is described by (non)linear function.

Discrete vs continuous

In a continuous time series observations are measured at every instance of time, whereas a discrete time series contains observations measured at discrete points in time.

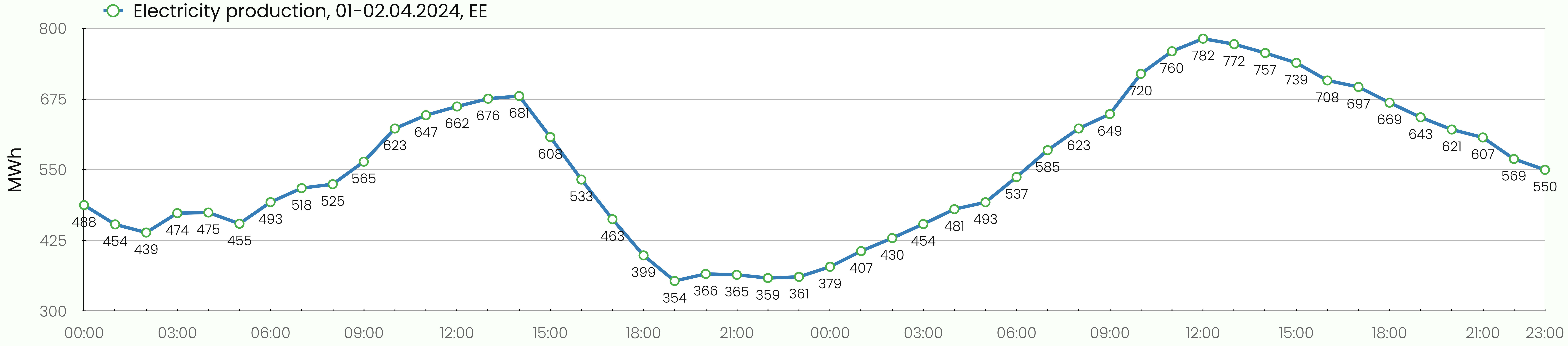
TIME SERIES NOTATIONS

A time series is an ordered sequence of data points, representing the values of a variable at successive time intervals.

Below notations are equivalent:

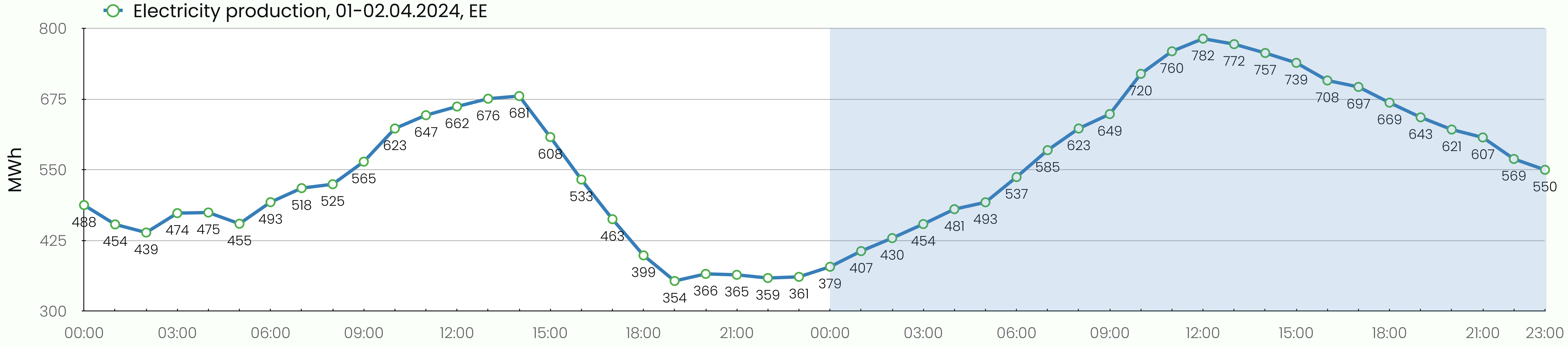
- $y = (y_1, y_2, \ldots, y_T)$
- $(y_t : t = 1,\ldots, T)$
- $y_{t|t-1} = y_t \mid \{y_1, y_2, \ldots, y_{t-1}\}$
- $\{y_t\}$

LAGGED OR *SHIFTED* TIME SERIES



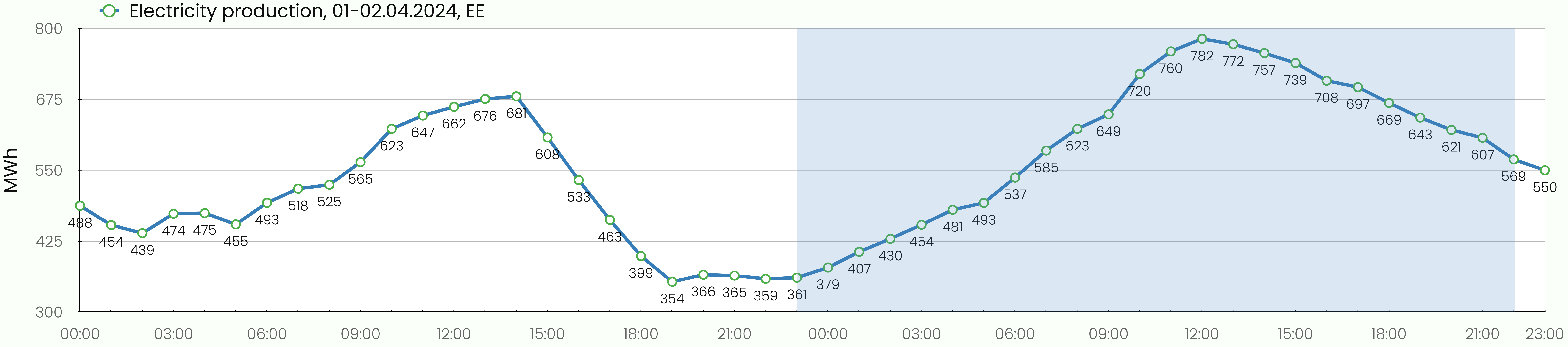
...	21	22	23	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	time

LAGGED OR *SHIFTED* TIME SERIES



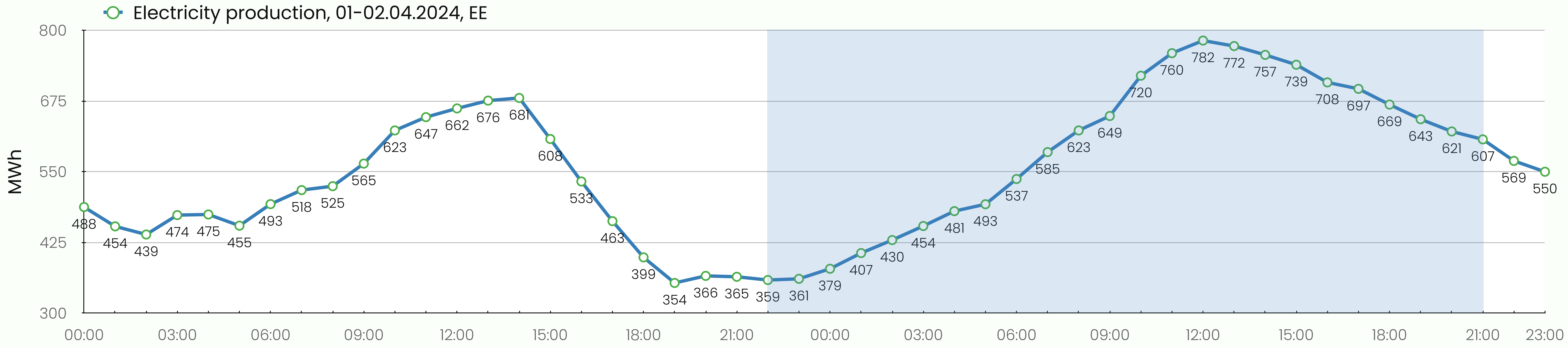
…	21	22	23	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	time
				379	407	430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621	607	569	550	y_t

LAGGED OR *SHIFTED* TIME SERIES



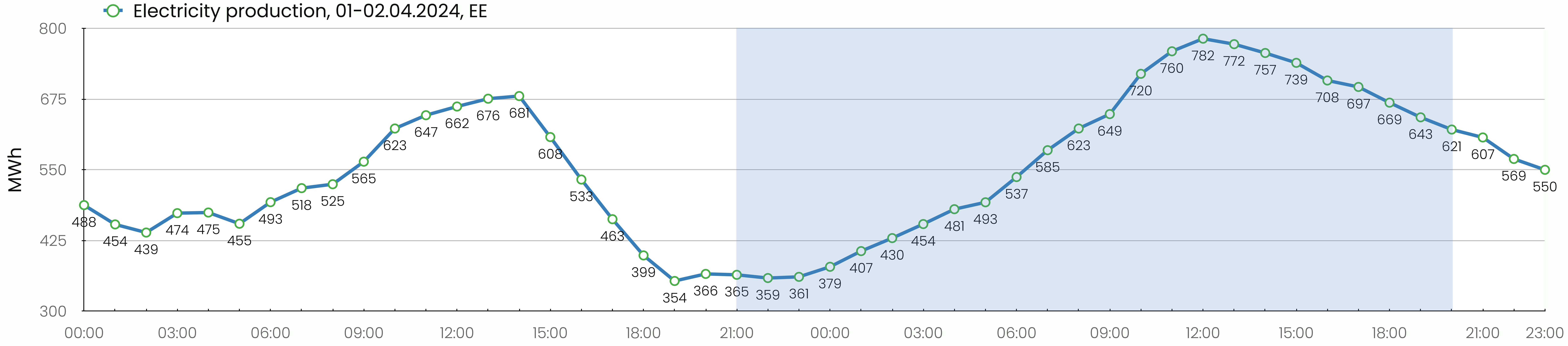
…	21	22	23	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	time
				379	407	430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621	607	569	550	y_t
			361	379	407	430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621	607	569		y_{t-1}

LAGGED OR *SHIFTED* TIME SERIES



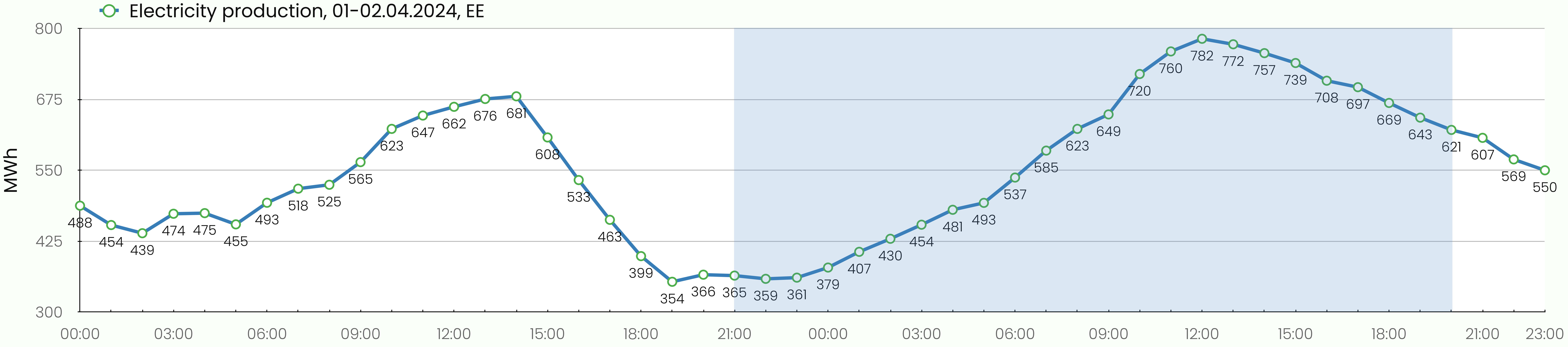
...	21	22	23	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	time
				379	407	430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621	607	569	550	y_t
			361	379	407	430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621	607	569		y_{t-1}
		359	361	379	407	430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621	607			y_{t-2}

LAGGED OR *SHIFTED* TIME SERIES



...	21	22	23	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	time
				379	407	430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621	607	569	550	y_t
			361	379	407	430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621	607	569		y_{t-1}
		359	361	379	407	430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621	607			y_{t-2}
	365	359	361	379	407	430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621				y_{t-3}

LAGGED OR *SHIFTED* TIME SERIES



…	21	22	23	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	time
				379	407	430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621	607	569	550	y_t
			361	379	407	430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621	607	569		y_{t-1}
		359	361	379	407	430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621	607			y_{t-2}
	365	359	361	379	407	430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621				y_{t-3}

What do we need to keep in mind?

PRACTICAL OBSERVATIONS

Say we want to study two time series y_t and y_{t-1}.

Case 1 (with history) we need to match dimensions, i.e., for y_{t-1}

drop the element from y_t at $t = 23$ on April 2 and

add another element at $t = 23$ on April 1.

DROP:

...	23	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	time
		379	407	430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621	607	569	550	y_t
	361	379	407	430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621	607	569		y_{t-1}

SHIFT:

24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	index
379	407	430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621	607	569	550	y_t
361	379	407	430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621	607	569	y_{t-1}

PRACTICAL OBSERVATIONS (2)

Say we want to study two time series y_t and y_{t-1}.

Case 2 (no history) we need to match dimensions:

for y_{t-1} drop the element from y_t at $t = 23$ on April 2 and

for y_t add drop the element at $t = 0$ on April 2.

DROP:

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	time
	407	430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621	607	569	550	y_t
379	407	430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621	607	569		y_{t-1}

SHIFT:

23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	index
407	430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621	607	569	550	y_t
379	407	430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621	607	569	y_{t-1}

PRACTICAL OBSERVATIONS (3)

Say we want to study two time series y_t and y_{t-2} with no history.

ORIG:

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	time
379	407	430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621	607	569	550	y_t

DROP:

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	time
		430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621	607	569	550	y_t
379	407	430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621	607			y_{t-2}

SHIFT:

22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	index
430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621	607	569	550	y_t
379	407	430	454	481	493	537	585	623	649	720	760	782	772	757	739	708	697	669	643	621	607	y_{t-2}

BACKSHIFT NOTATION

The backward shift operator B can be used to denote time series lags:

$$By_t = y_{t-1}.$$

B shifts the data back one period.

BACKSHIFT NOTATION

The backward shift operator B can be used to denote time series lags:

$$By_t = y_{t-1}.$$

B shifts the data back one period.

Two applications of B to y_t shift the data back two periods:

$$B(By_t) = B^2y_t = y_{t-2}.$$

BACKSHIFT NOTATION

The backward shift operator B can be used to denote time series lags:

$$By_t = y_{t-1}.$$

B shifts the data back one period.

Two applications of B to y_t shift the data back two periods:

$$B(By_t) = B^2y_t = y_{t-2}.$$

For seasonal data, we denote the lag by m:

$$B^my_t = y_{t-m}.$$

BACKSHIFT NOTATION

The backward shift operator B can be used to denote time series lags:

$$By_t = y_{t-1}.$$

B shifts the data back one period.

Two applications of B to y_t shift the data back two periods:

$$B(By_t) = B^2 y_t = y_{t-2}.$$

For seasonal data, we denote the lag by m:

$$B^m y_t = y_{t-m}.$$

For example, the same hour yesterday $(m = 24)$ is denoted as $B^{24} y_t = y_{t-24}$.

BACKSHIFT NOTATION: USE CASES

Useful to denote time series transformations.

A first-order difference can be written as:

$$y'_t = y_t - y_{t-1} = y_t - By_t = (1 - B)y_t.$$

BACKSHIFT NOTATION: USE CASES

Useful to denote time series transformations.

A first-order difference can be written as:

$$y'_t = y_t - y_{t-1} = y_t - By_t = (1 - B)y_t.$$

Similarly, the second-order difference can be written as:

$$\begin{aligned} y''_t &= y'_t - y'_{t-1} \\ &= y_t - y_{t-1} - (y_{t-1} - y_{t-2}) \\ &= y_t - 2y_{t-1} + y_{t-2} \\ &= \left(1 - 2B + B^2\right) y_t \\ &= (1 - B)^2 y_t\,. \end{aligned}$$

BACKSHIFT NOTATION: USE CASES (2)

An nth-order difference can denoted as:

$$(1 - B)^n y_t.$$

BACKSHIFT NOTATION: USE CASES (2)

An nth-order difference can denoted as:

$$(1 - B)^n y_t.$$

Given time series y_t.

We want to apply the seasonal difference followed a first difference. This can be written as:

$$\begin{aligned}(1 - B)\left(1 - B^m\right) y_t &= \left(1 - B - B^m + B^{m+1}\right) y_t \\ &= y_t - y_{t-1} - y_{t-m} + y_{t-m-1}\end{aligned}.$$

TASKS FOR TIME SERIES ANALYSIS

- Visualisation
- Anomaly Detection
- Components Analysis
- Time Series Decomposition
- Descriptive Analysis
- Modelling
- Forecasting
- Validation

TASKS FOR TIME SERIES ANALYSIS

- Visualisation
- Anomaly Detection
- Components Analysis
- Time Series Decomposition
- Descriptive Analysis
- Modelling
- Forecasting
- Validation

Component analysis

TIME SERIES COMPONENTS

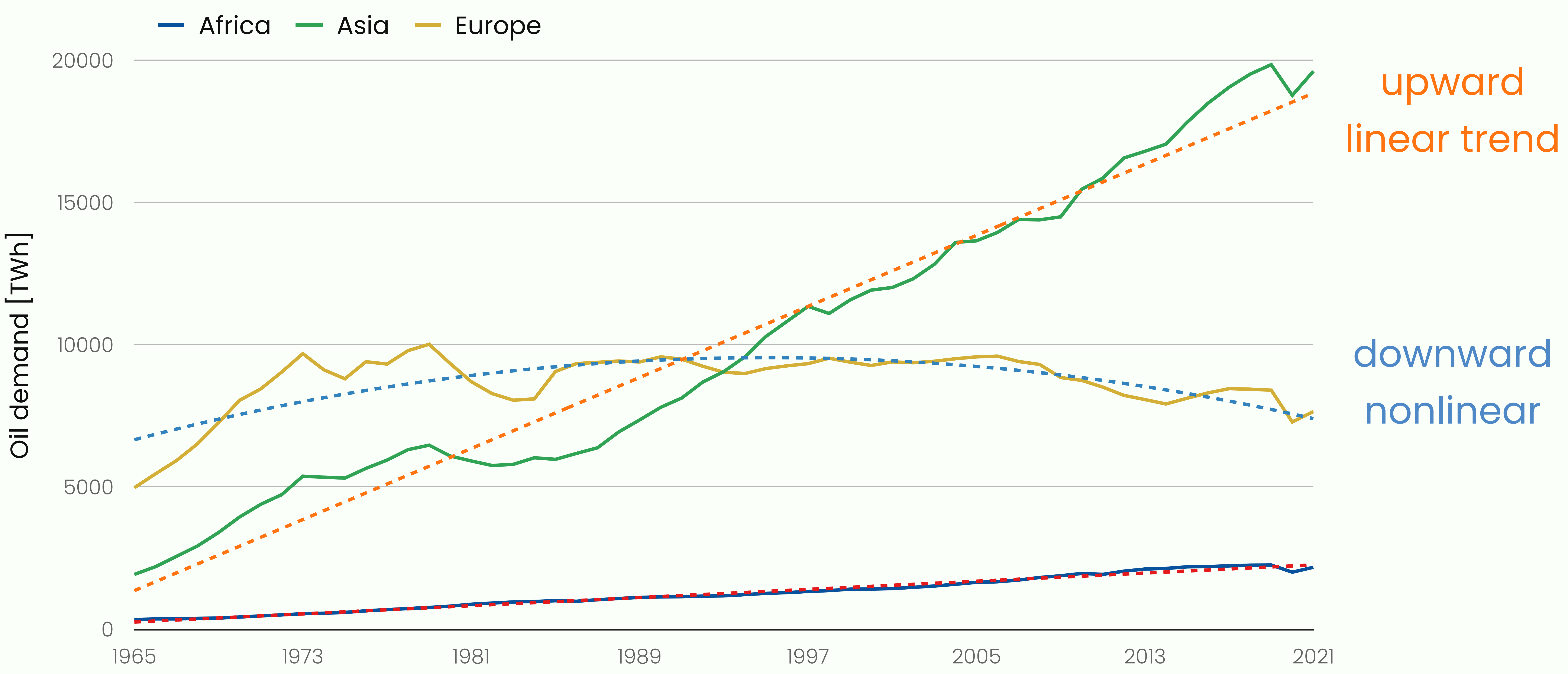
TIME SERIES = SIGNAL + NOISE

TIME SERIES	=	SIGNAL	+	NOISE
		Trend / Seasonal / Cycle		Noise / Random / Irregular

TREND

Trends indicates a long-term increase, decrease, or stagnation in the data.

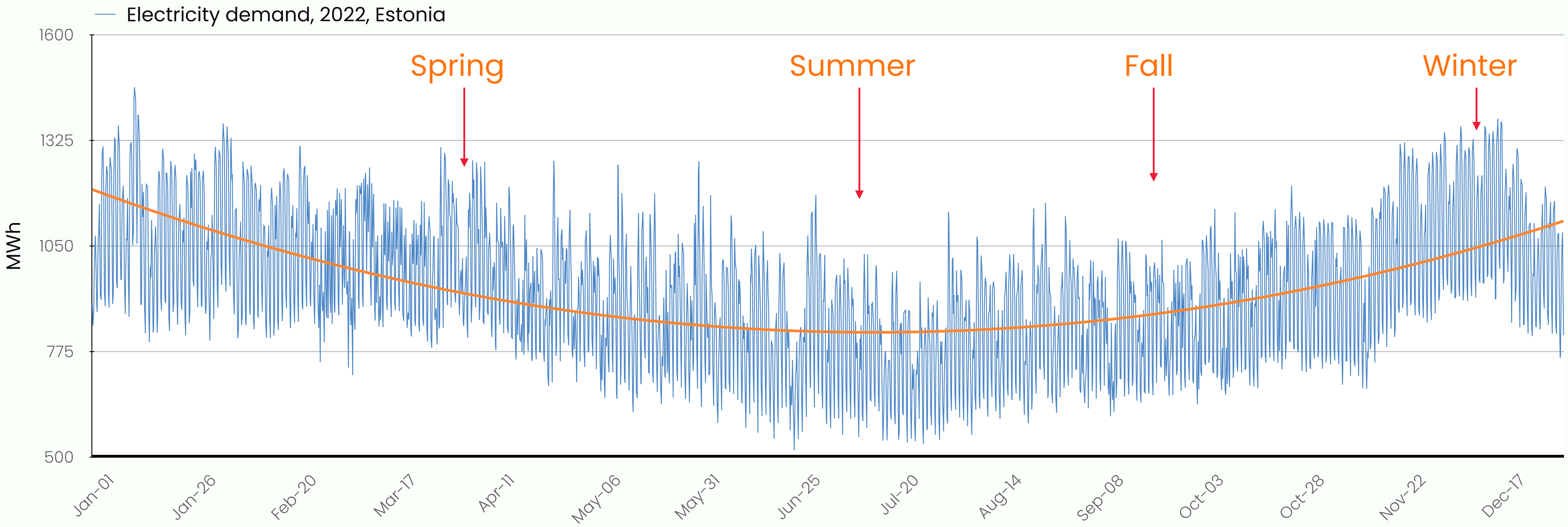
Trends can be linear and nonlinear.



[Accessed September 10, 2025], https://ourworldindata.org/grapher/oil-consumption-by-region-terawatt-hours-twh

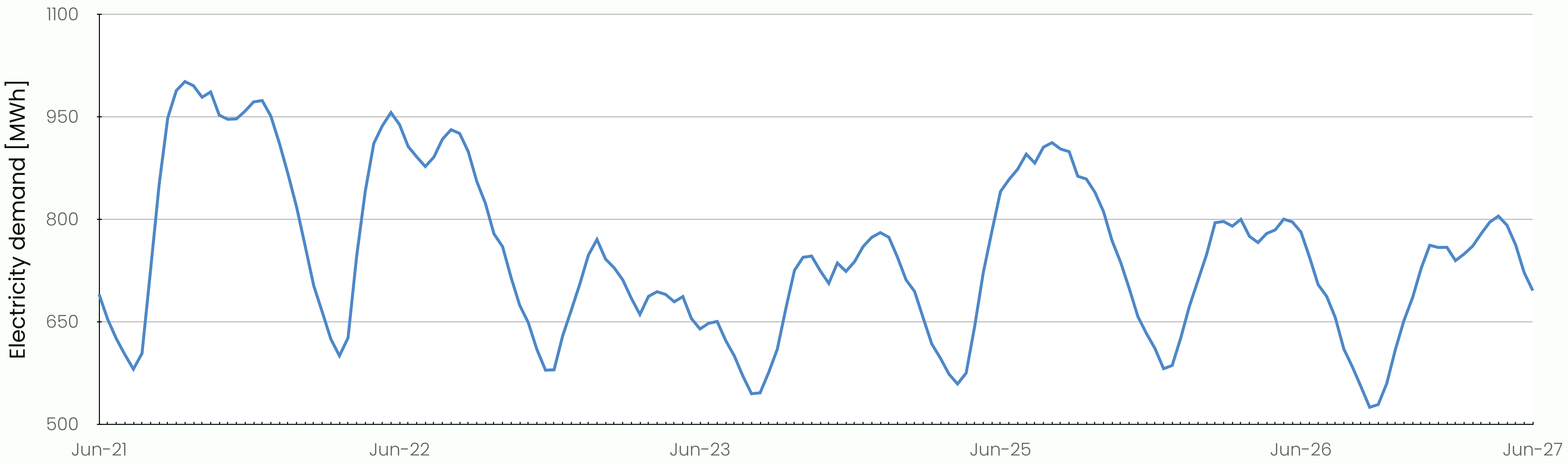
SEASONAL

Seasonality occurs when a time series is affected by seasonal factors (eg weather). Seasonality is of a fixed and known frequency.



SEASONAL (2)

Seasonal patterns are not limited by the traditional seasons (spring, summer, fall, or winter), but may also be hourly, daily, weekly, or monthly.

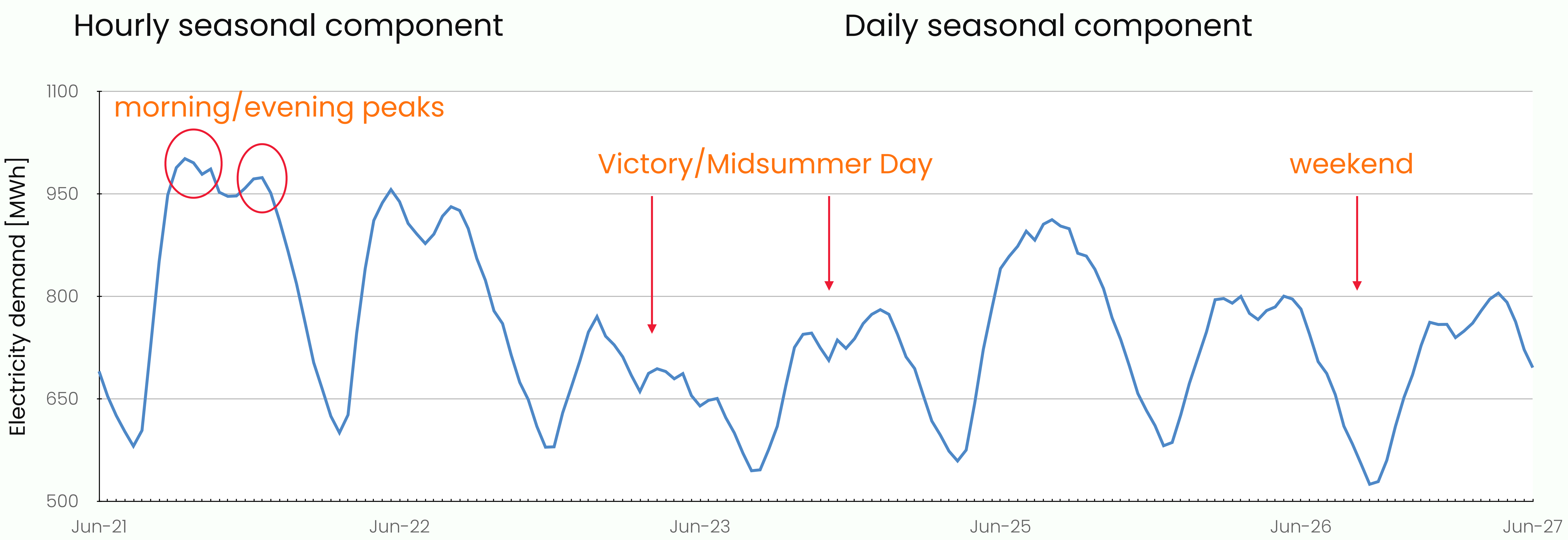


SEASONAL (2)

Seasonal patterns are not limited by the traditional seasons (spring, summer, fall, or winter), but may also be hourly, daily, weekly, or monthly.

SEASONAL (2)

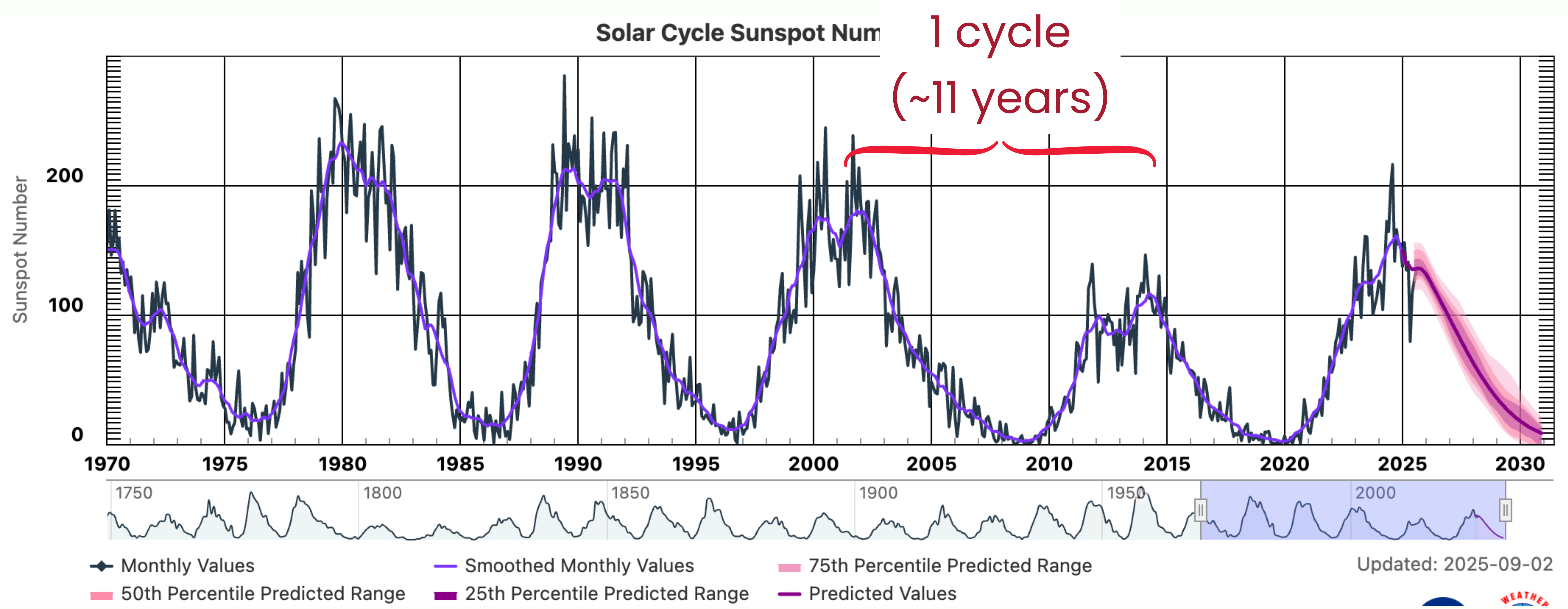
Seasonal patterns are not limited by the traditional seasons (spring, summer, fall, or winter), but may also be hourly, daily, weekly, or monthly.



CYCLE

A **cycle** occurs when the data exhibit rises and falls.

Regularly occur, but may vary in length.

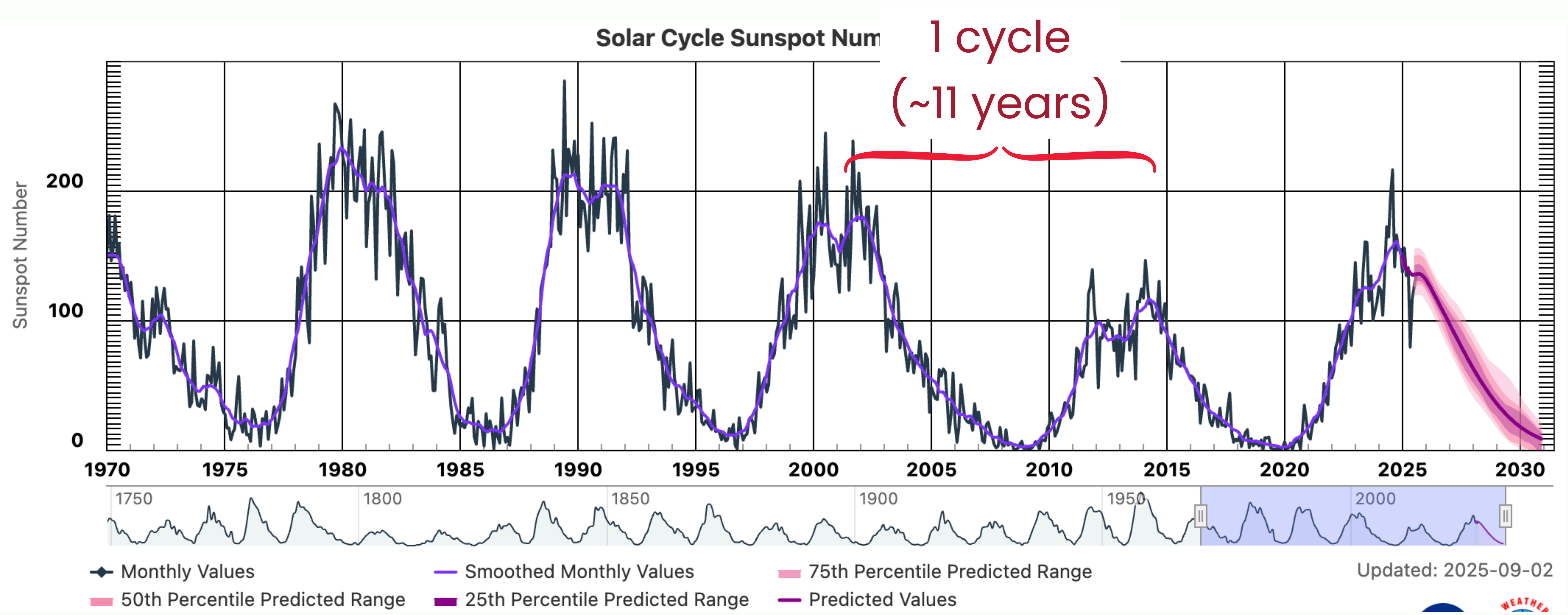


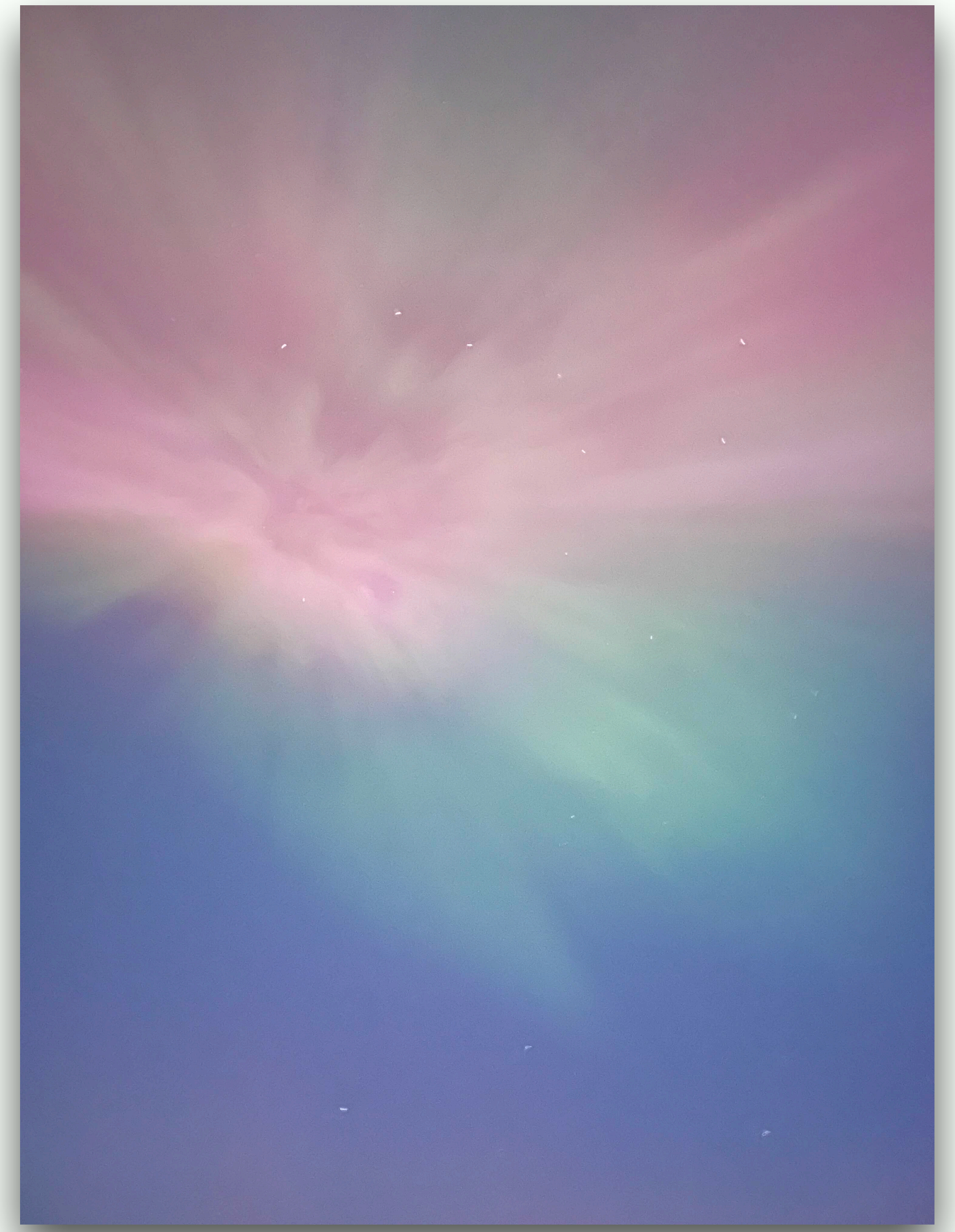
[Accessed September 10, 2025], https://www.swpc.noaa.gov/products/solar-cycle-progression

CYCLE

A **cycle** occurs when the data exhibit rises and falls.

Regularly occur, but may vary in length.





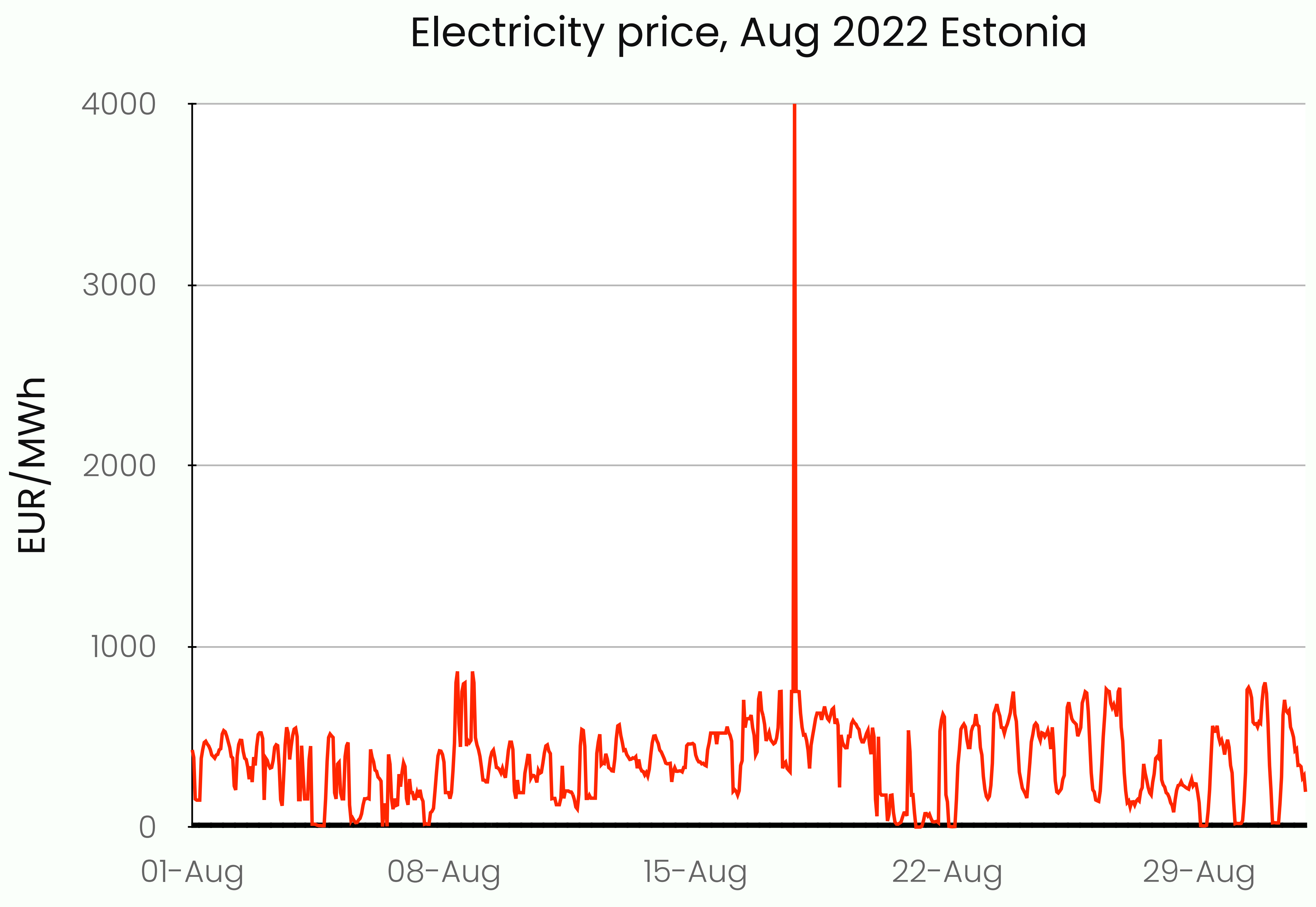
SEASONAL VS CYCLIC

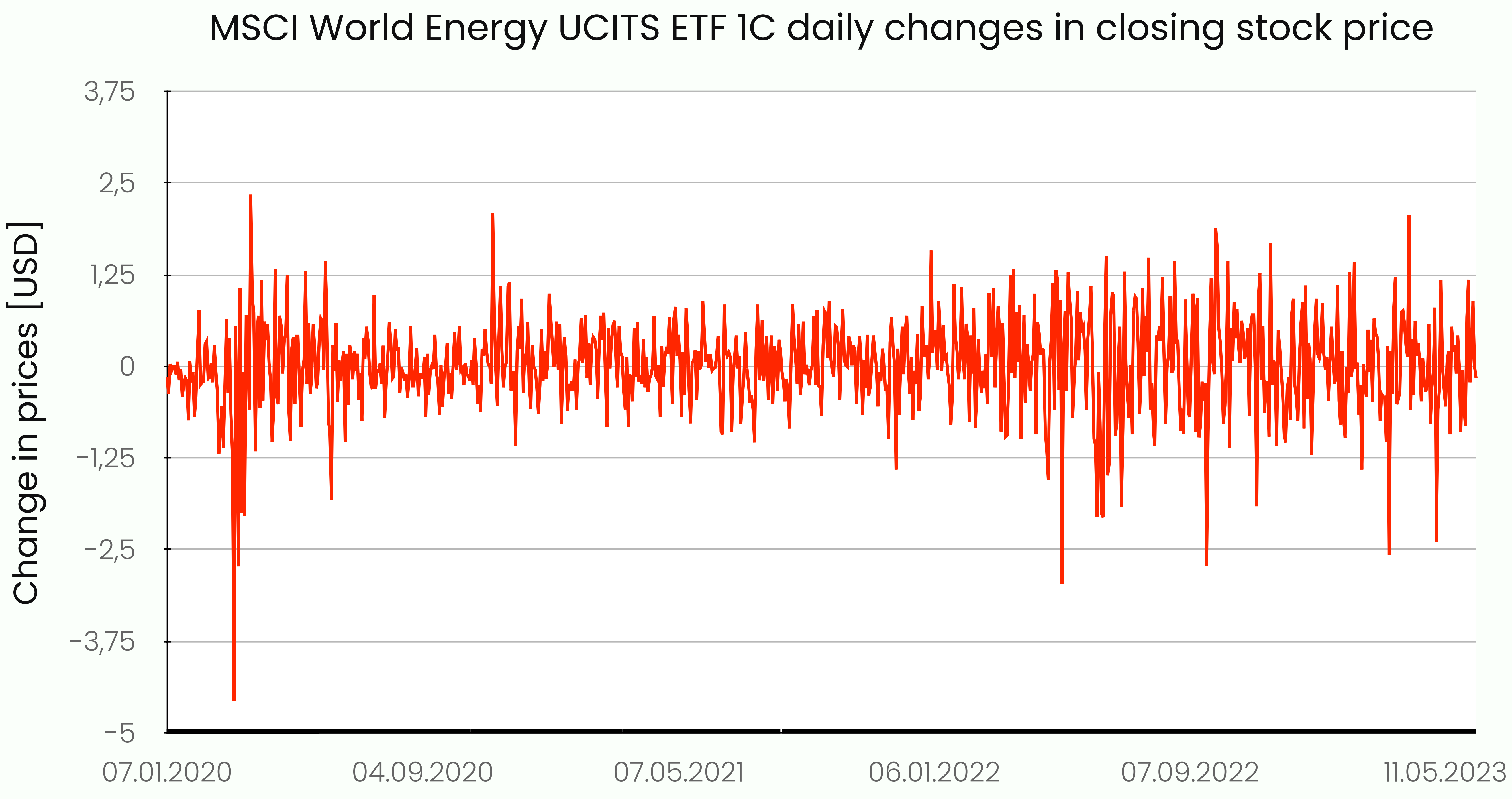
Cyclic and seasonal behaviour similar but have certain distinctions:

- If the fluctuations are not of a fixed frequency then they are cyclic.
- If the frequency is unchanging and associated with some aspect of the calendar, then the pattern is seasonal.

RANDOM/IRREGULAR

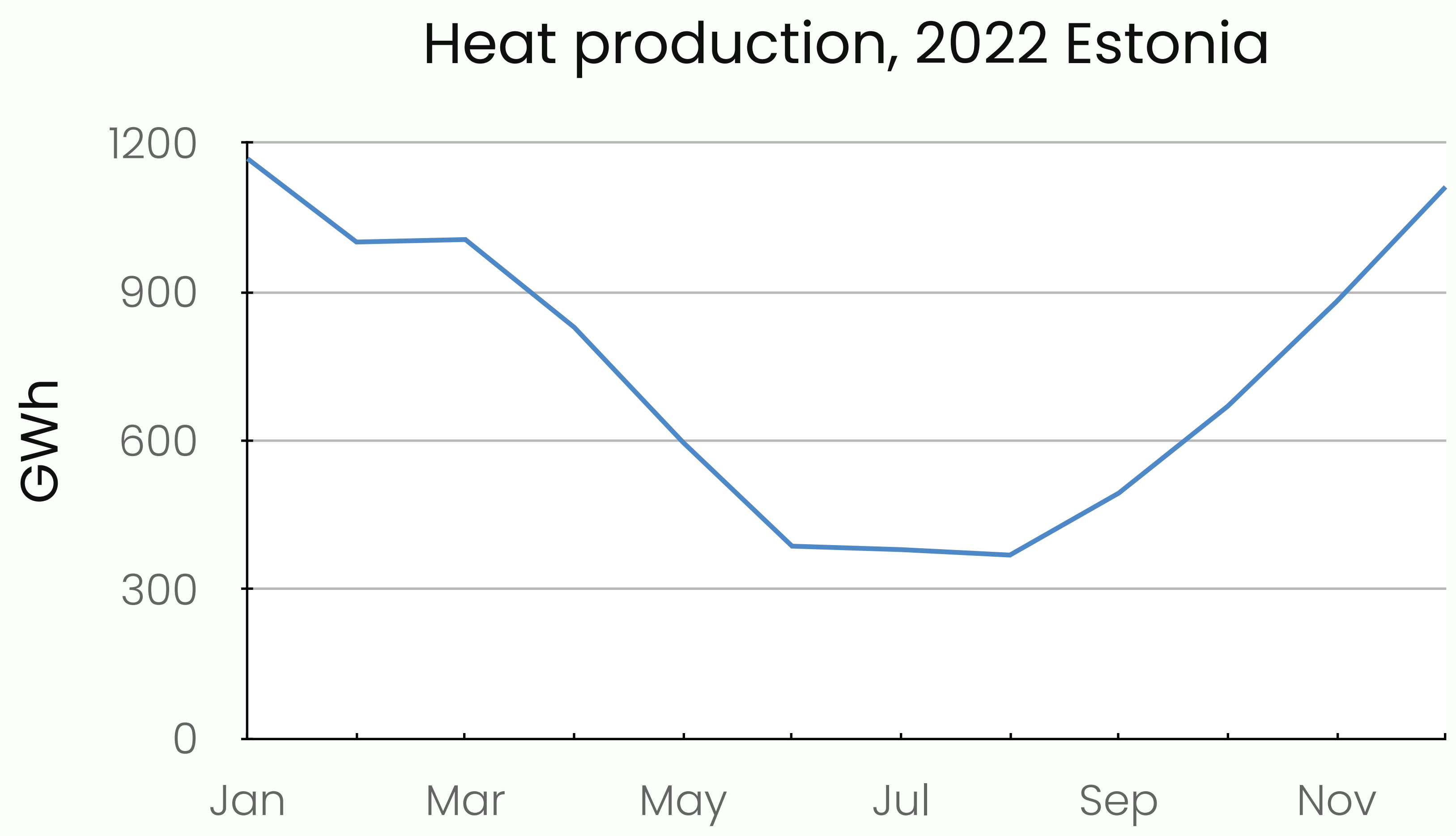
Unexpected events (also referred to as noise, irregularities, or random variations) may occur due to unpredictable reasons (eg war or earthquake).





[Accessed September 10, 2025], https://etf.dws.com/en/IE00BM67HM91-msci-world-energy-ucits-etf-1c/

EXAMPLES

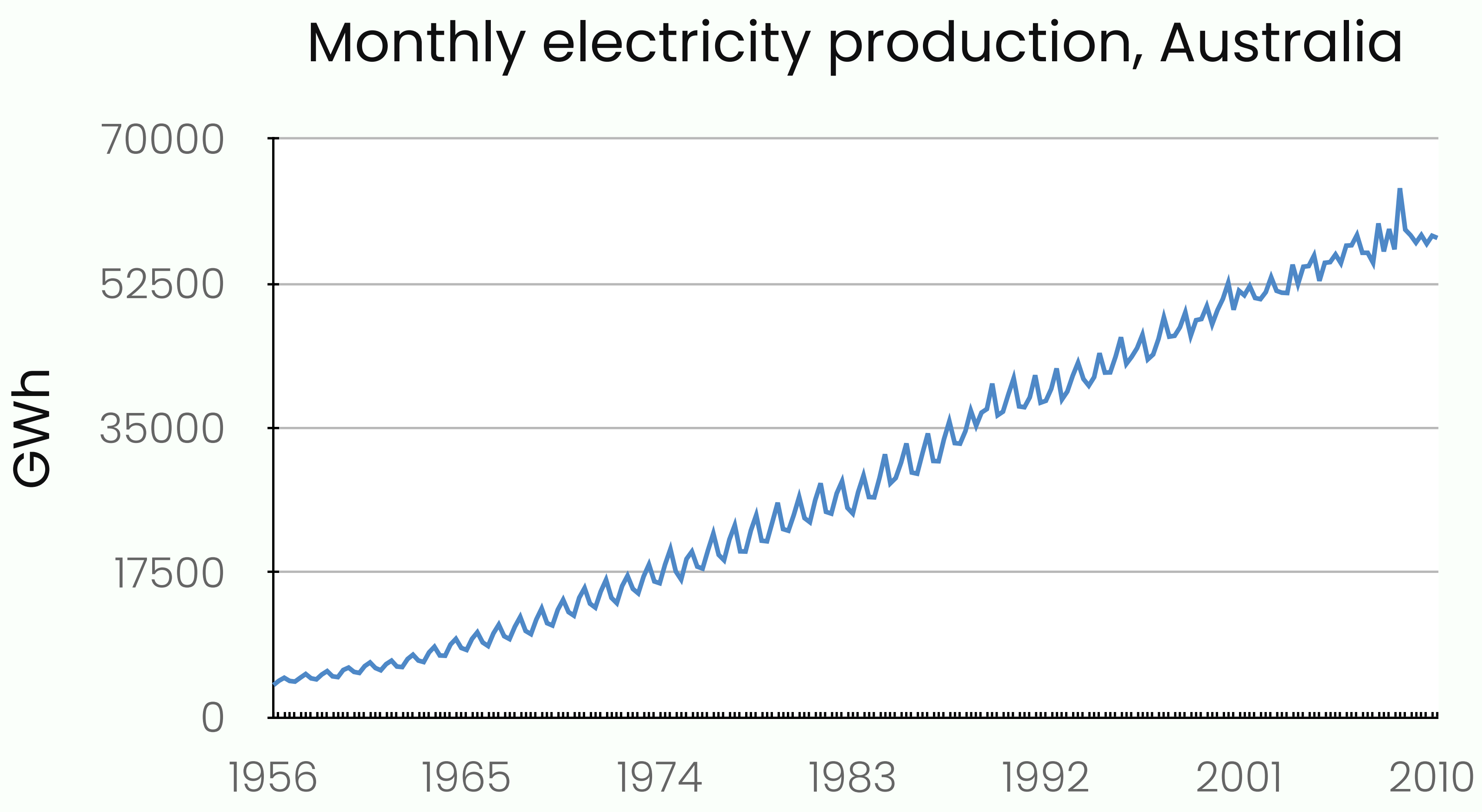


Trend

Seasonal

Cyclical

Random

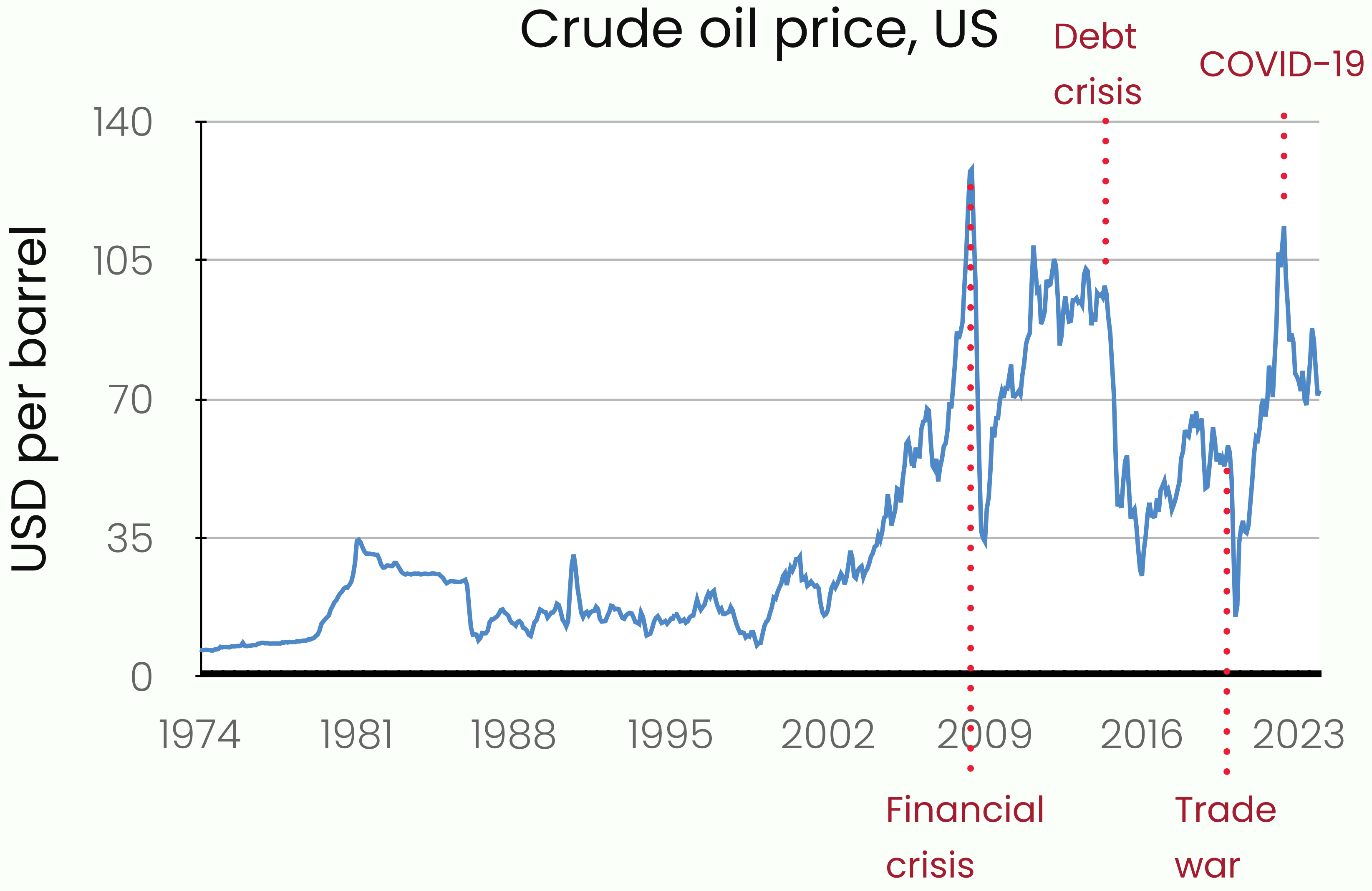


Trend

Seasonal

Cyclical

Random

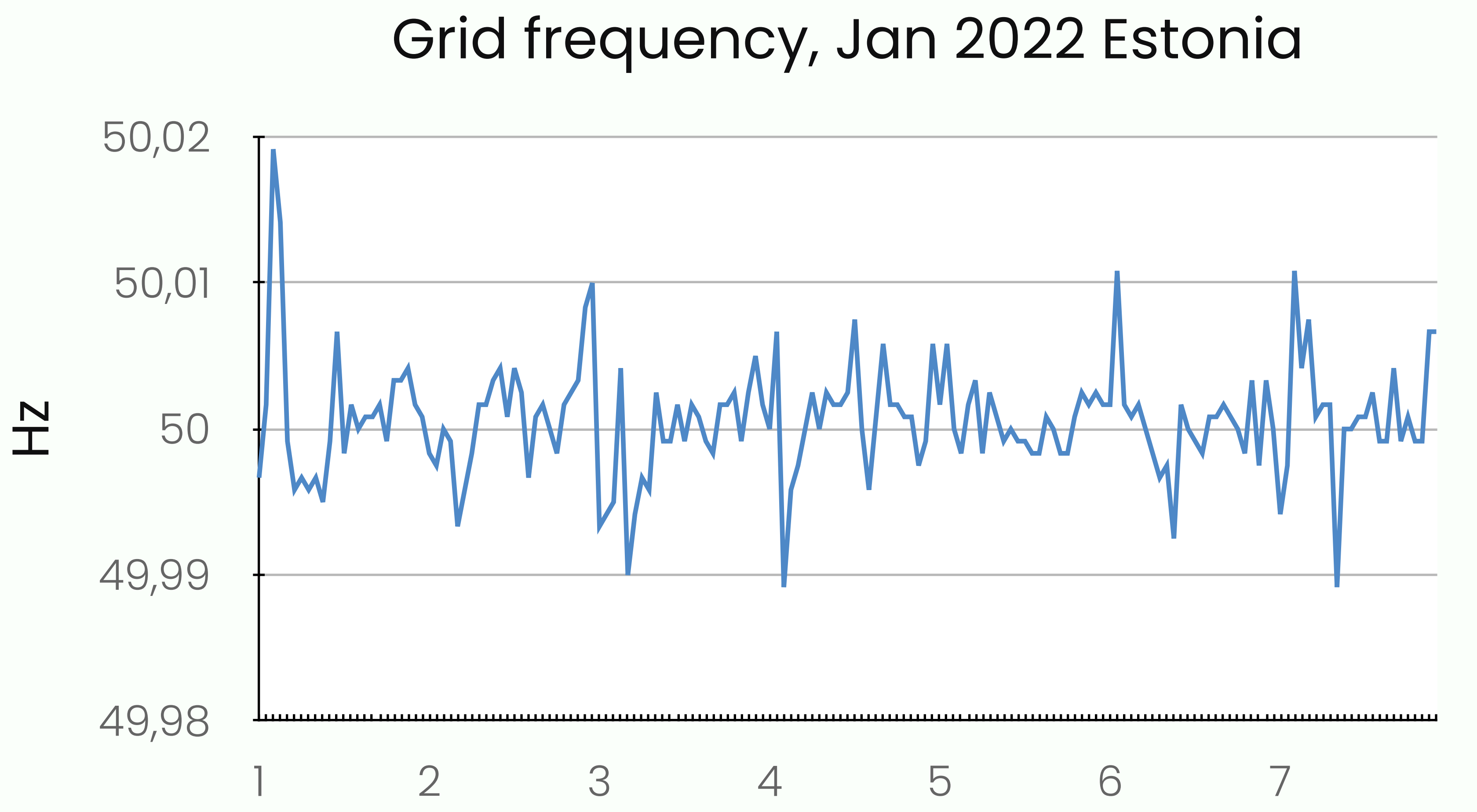


Trend

Seasonal

Cyclical

Random



Trend

Seasonal

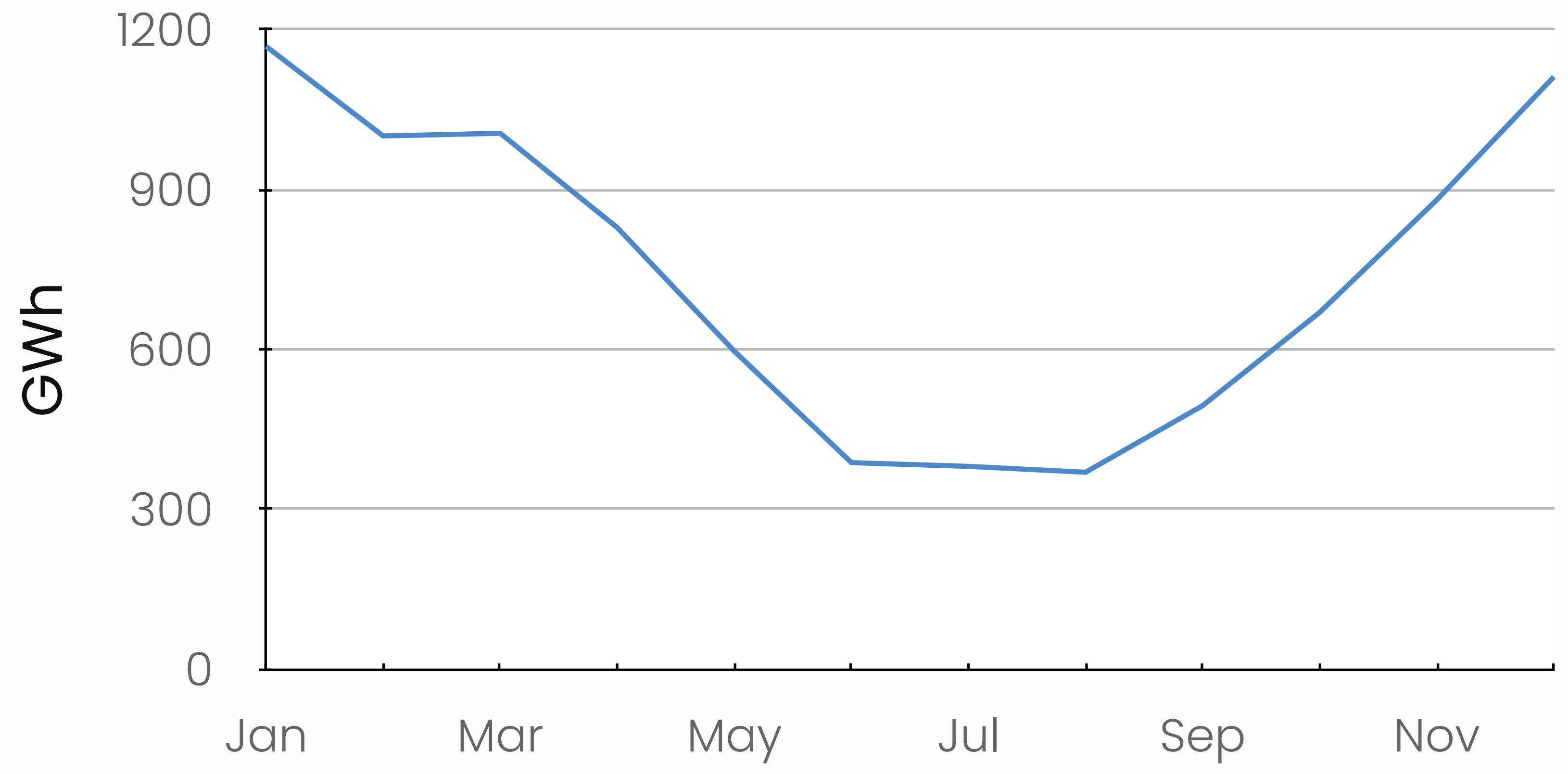
Cyclical

Random

[Accessed September 10, 2025], https://github.com/robjhyndman/fpp2-package
https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=pet&s=f000000__3&f=m

EXAMPLES

Heat production, 2022 Estonia



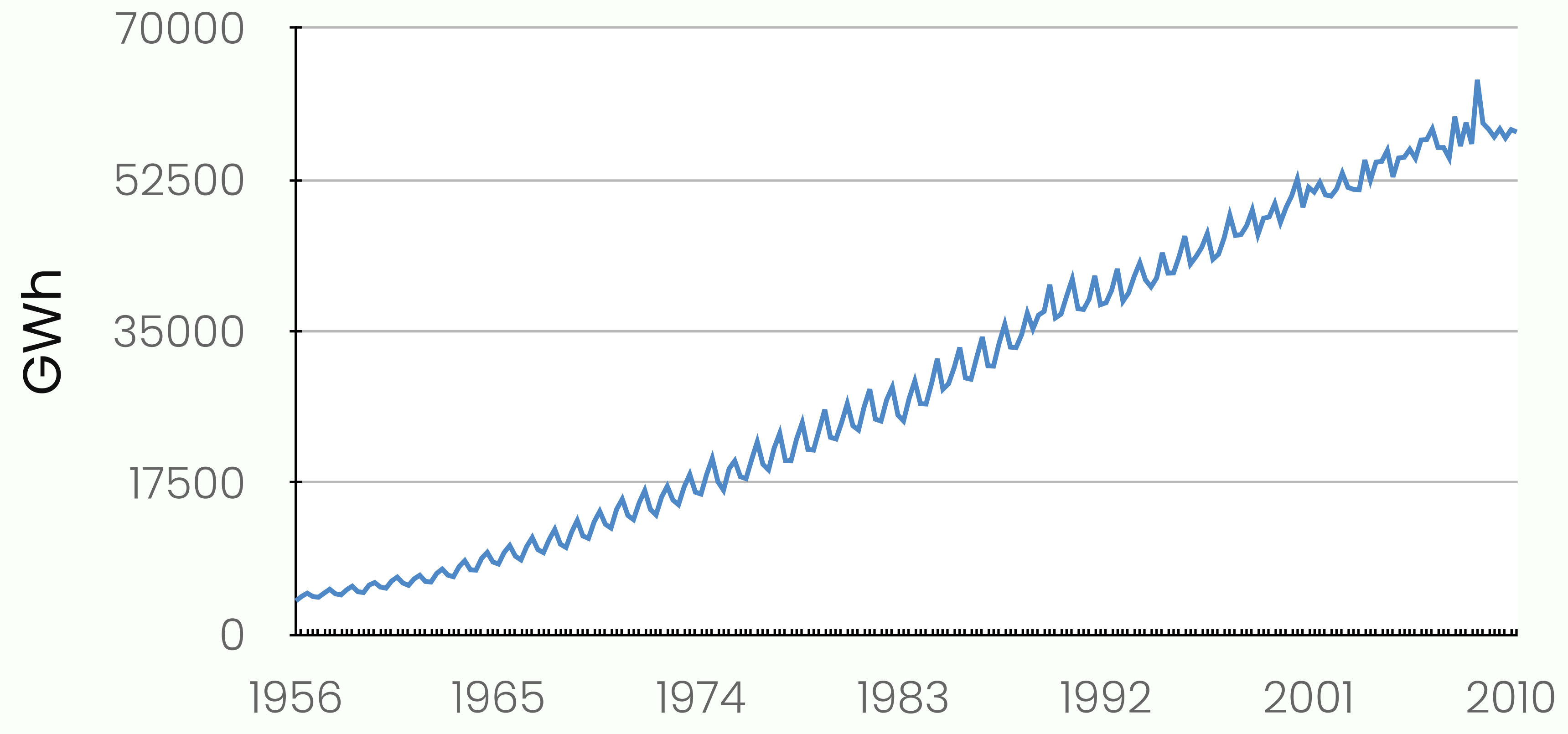
Trend ✗

Seasonal ✓

Cyclical ✗

Random ✗

Monthly electricity production, Australia



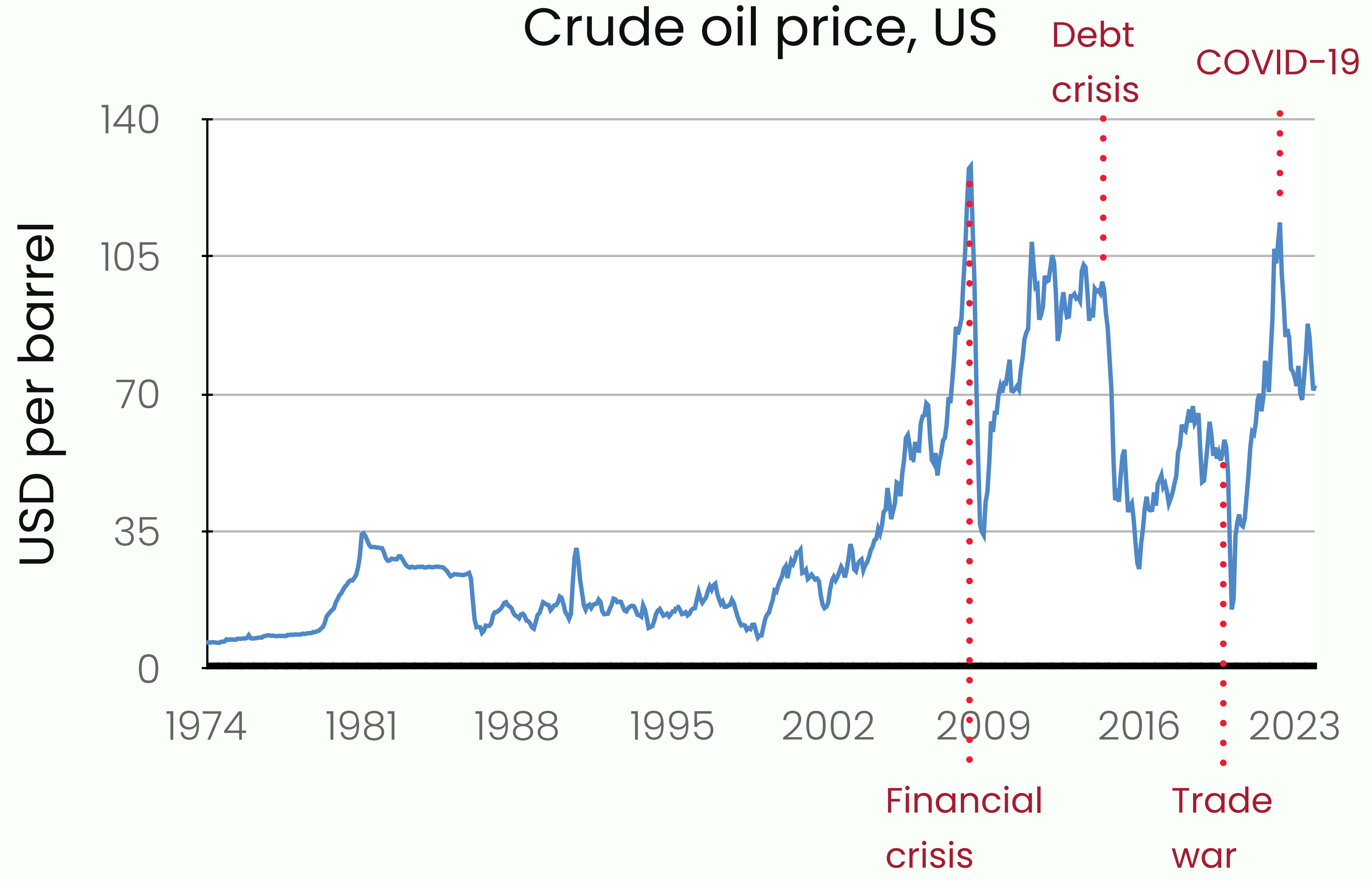
Trend

Seasonal

Cyclical

Random

Crude oil price, US



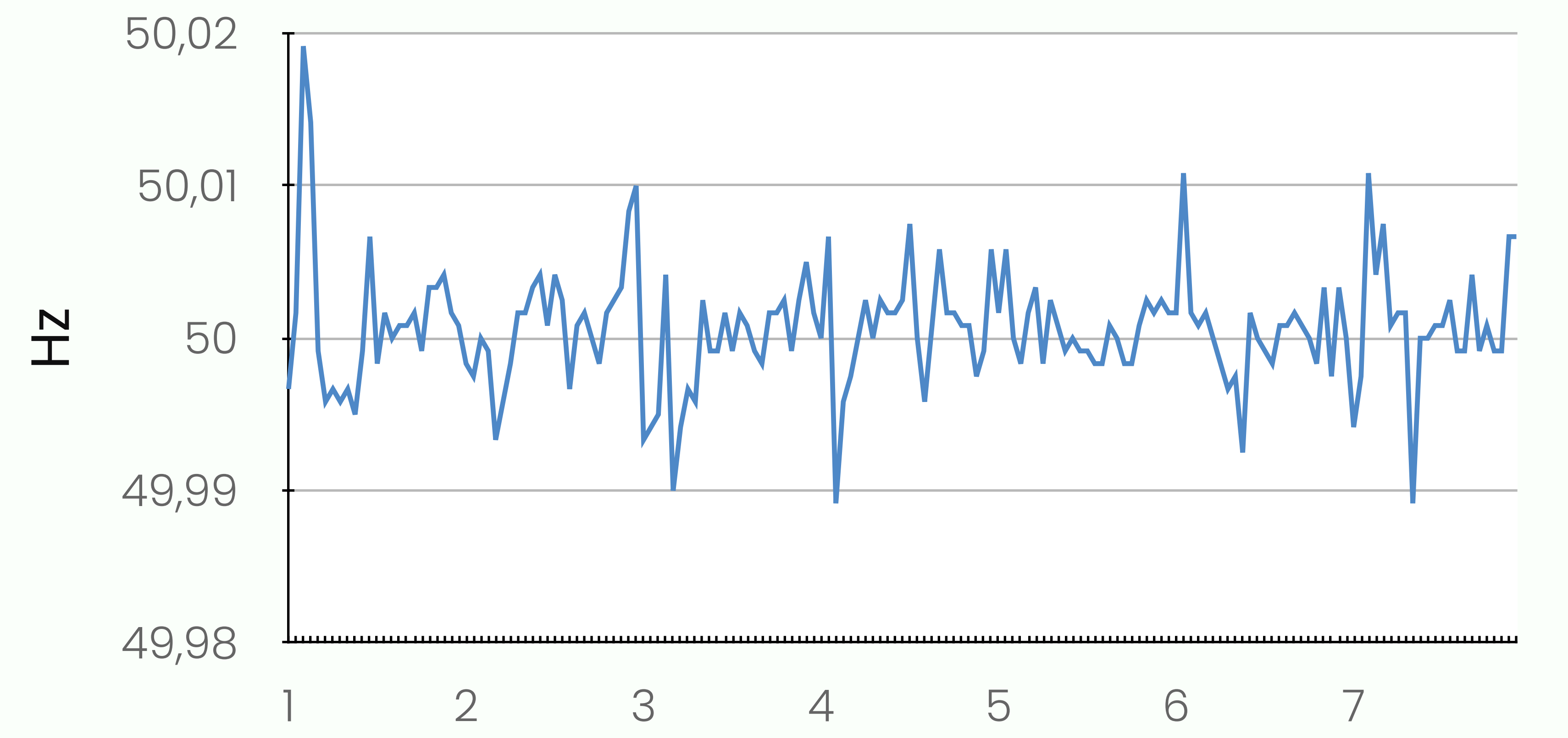
Trend

Seasonal

Cyclical

Random

Grid frequency, Jan 2022 Estonia



Trend

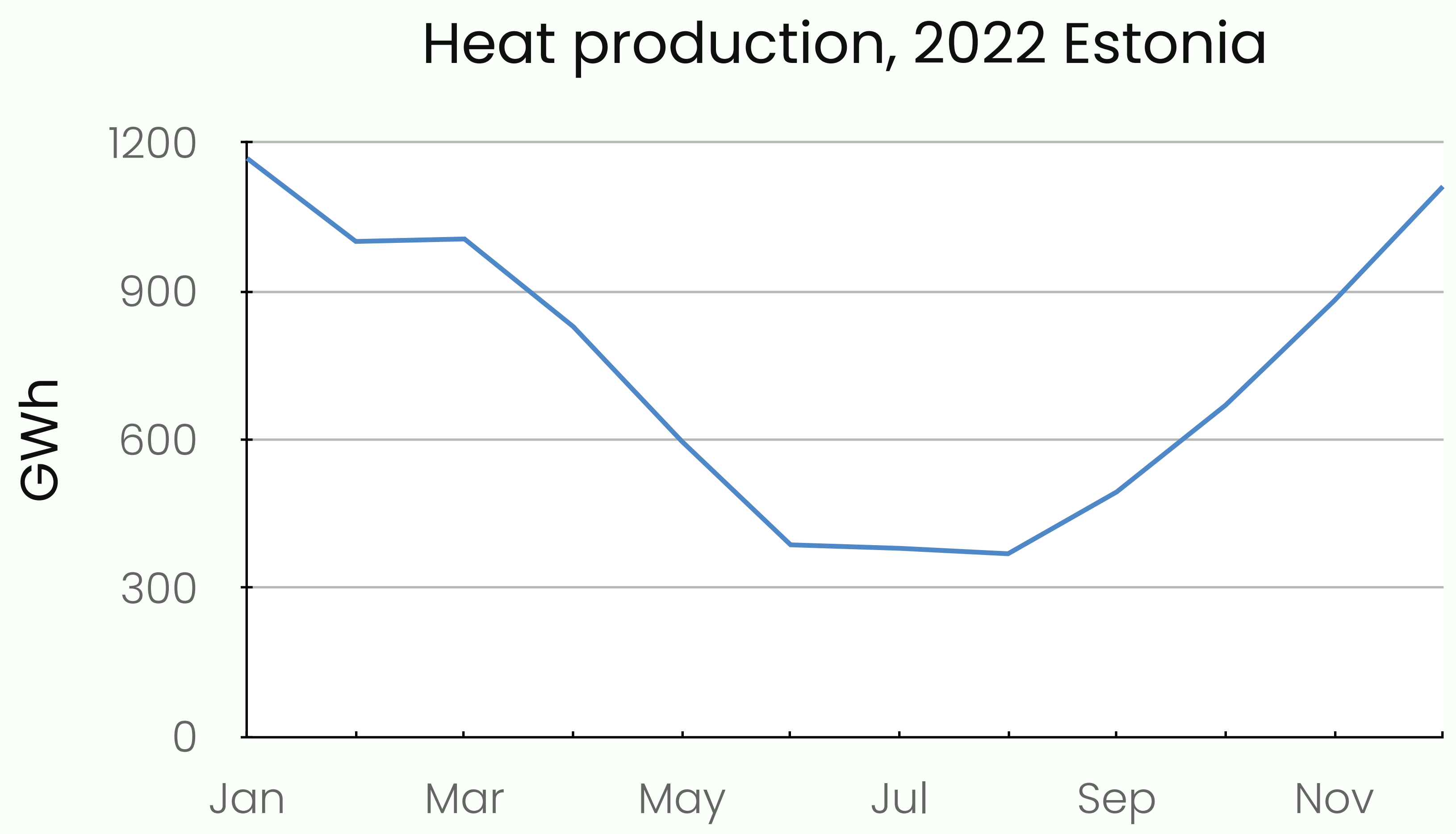
Seasonal

Cyclical

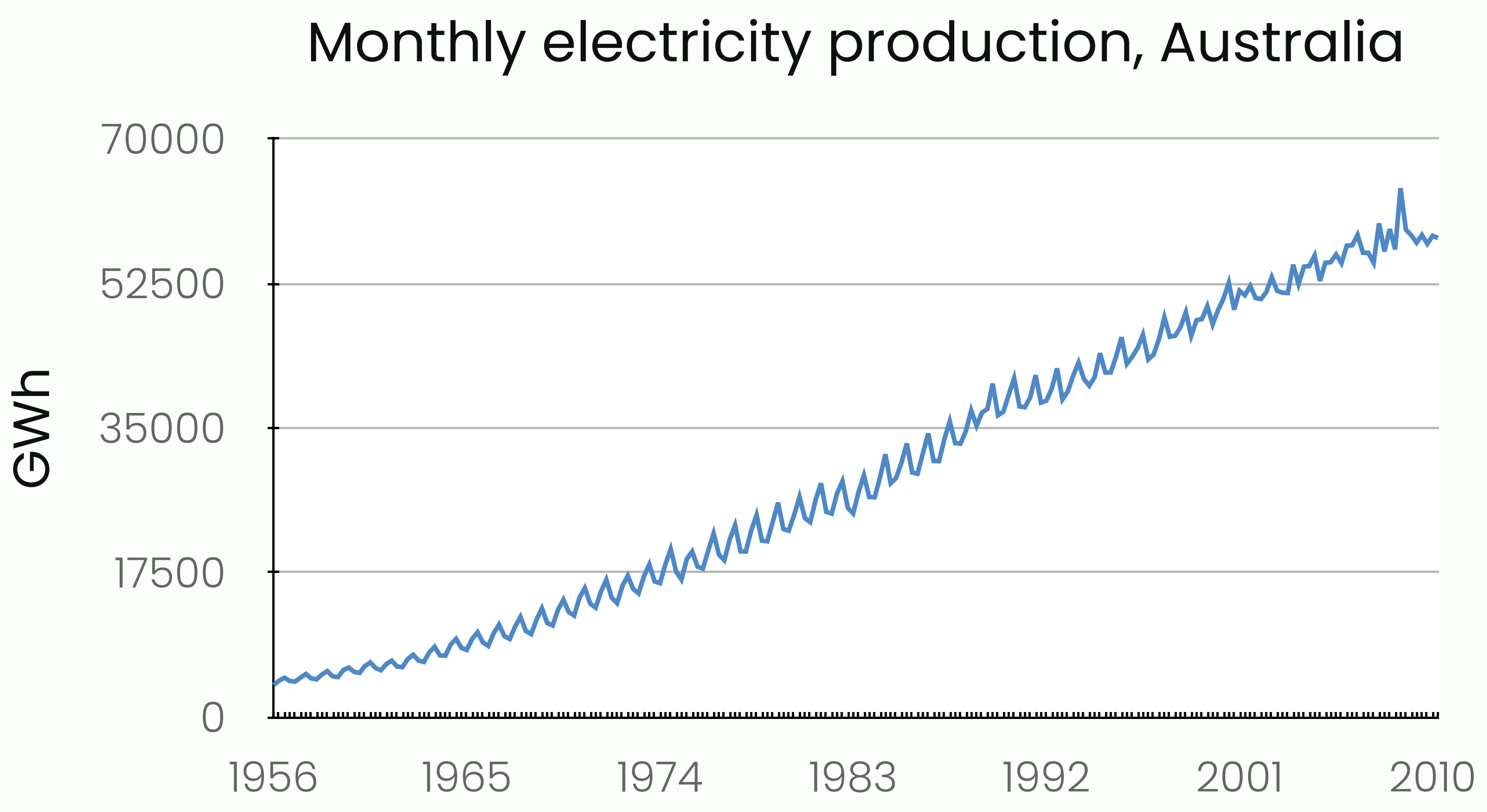
Random

[Accessed September 10, 2025], https://github.com/robjhyndman/fpp2-package
https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=pet&s=f000000__3&f=m

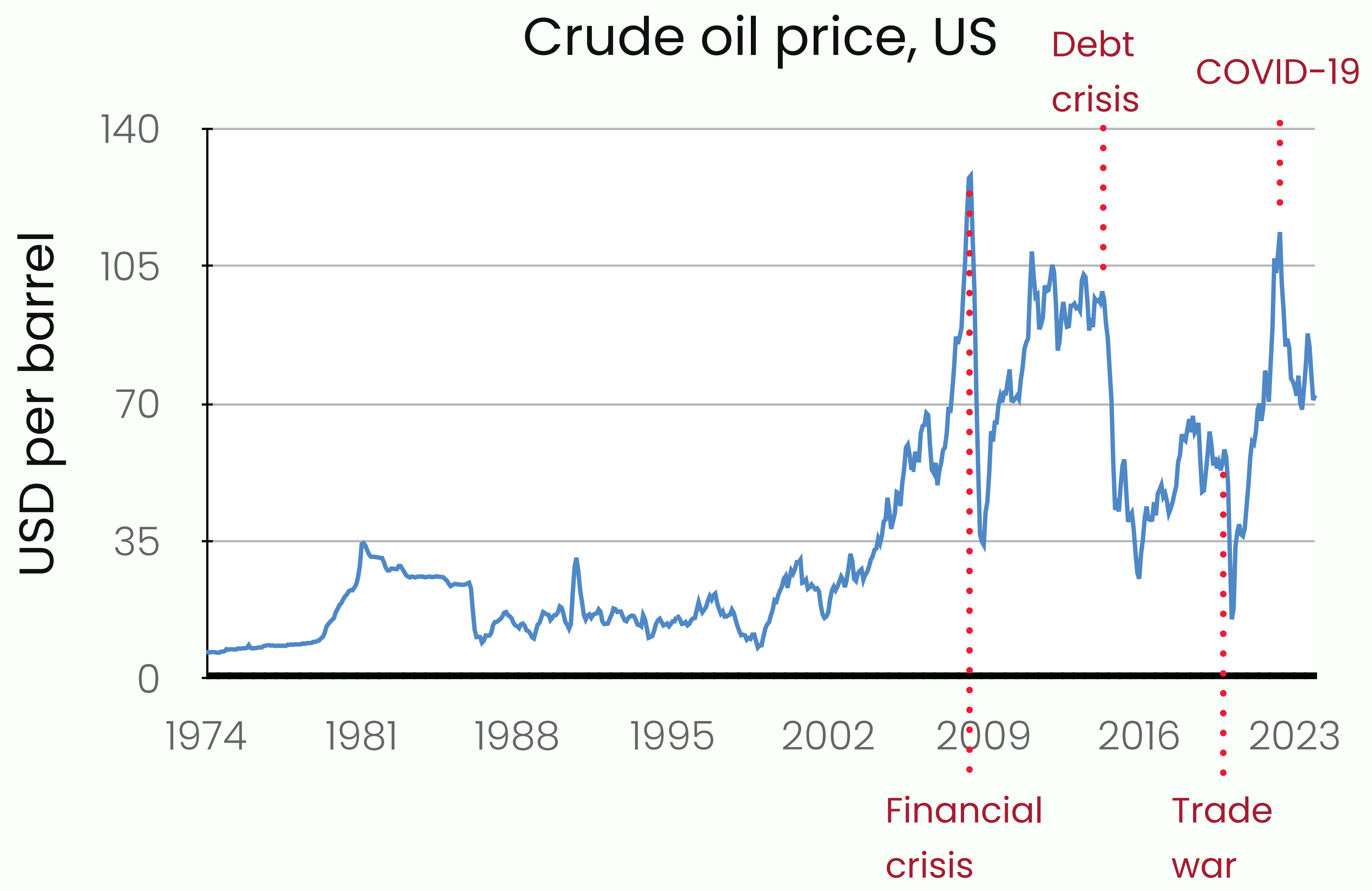
EXAMPLES



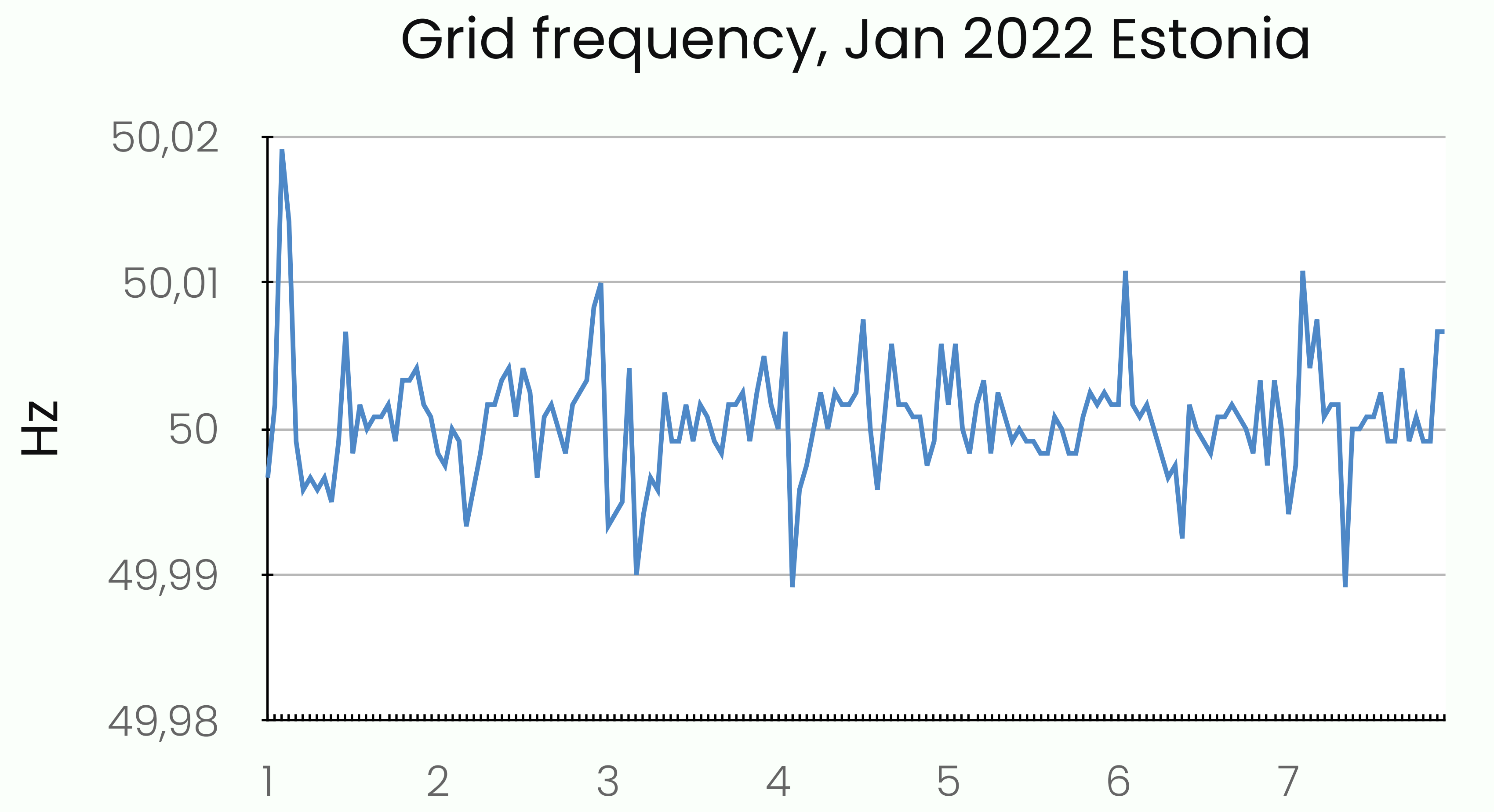
- Trend ✗
- Seasonal ✓
- Cyclical ✗
- Random ✗



- Trend ✓
- Seasonal ✓
- Cyclical ✗
- Random ✗



- Trend
- Seasonal
- Cyclical
- Random

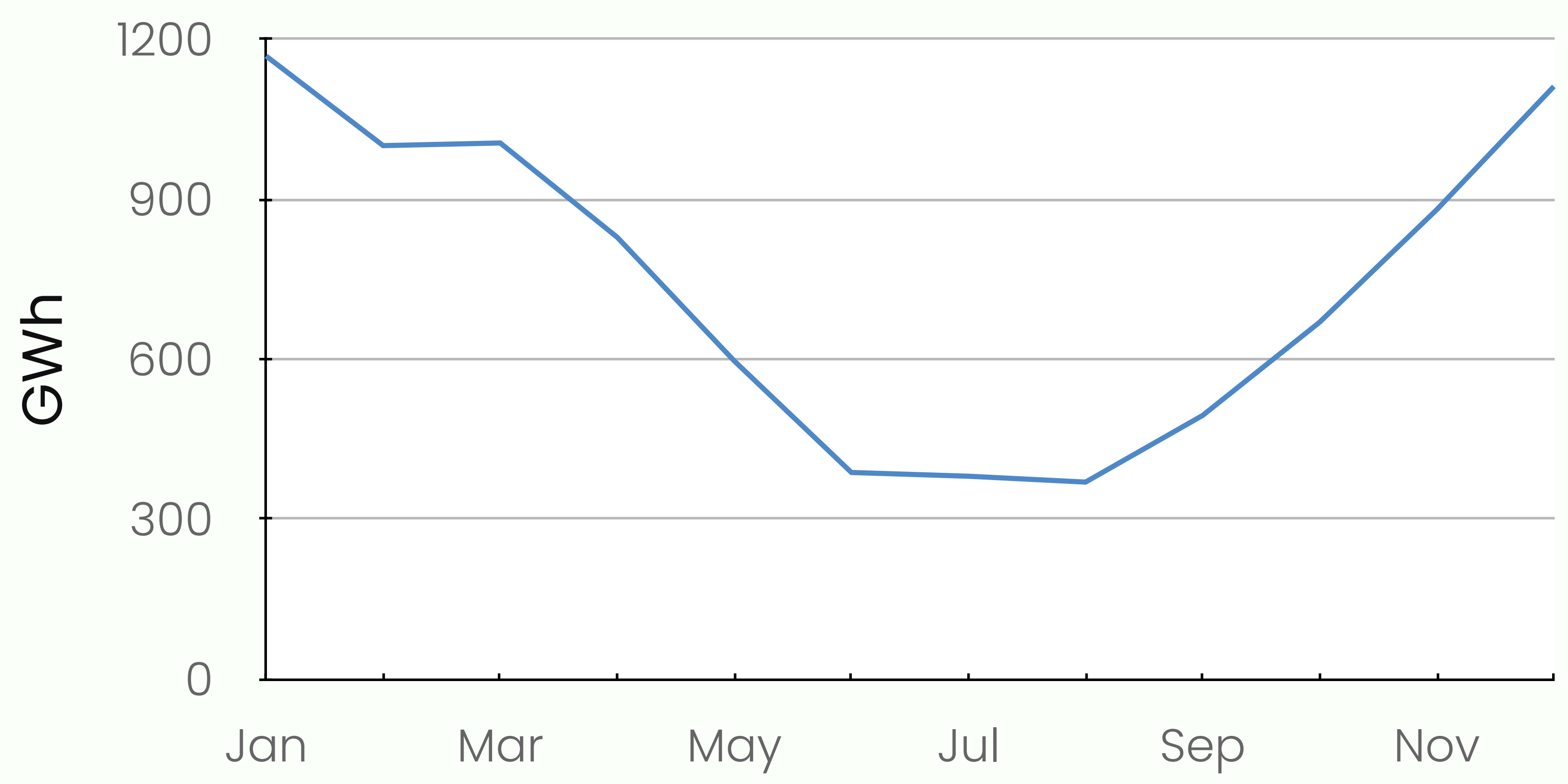


- Trend
- Seasonal
- Cyclical
- Random

[Accessed September 10, 2025], https://github.com/robjhyndman/fpp2-package
https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=pet&s=f000000__3&f=m

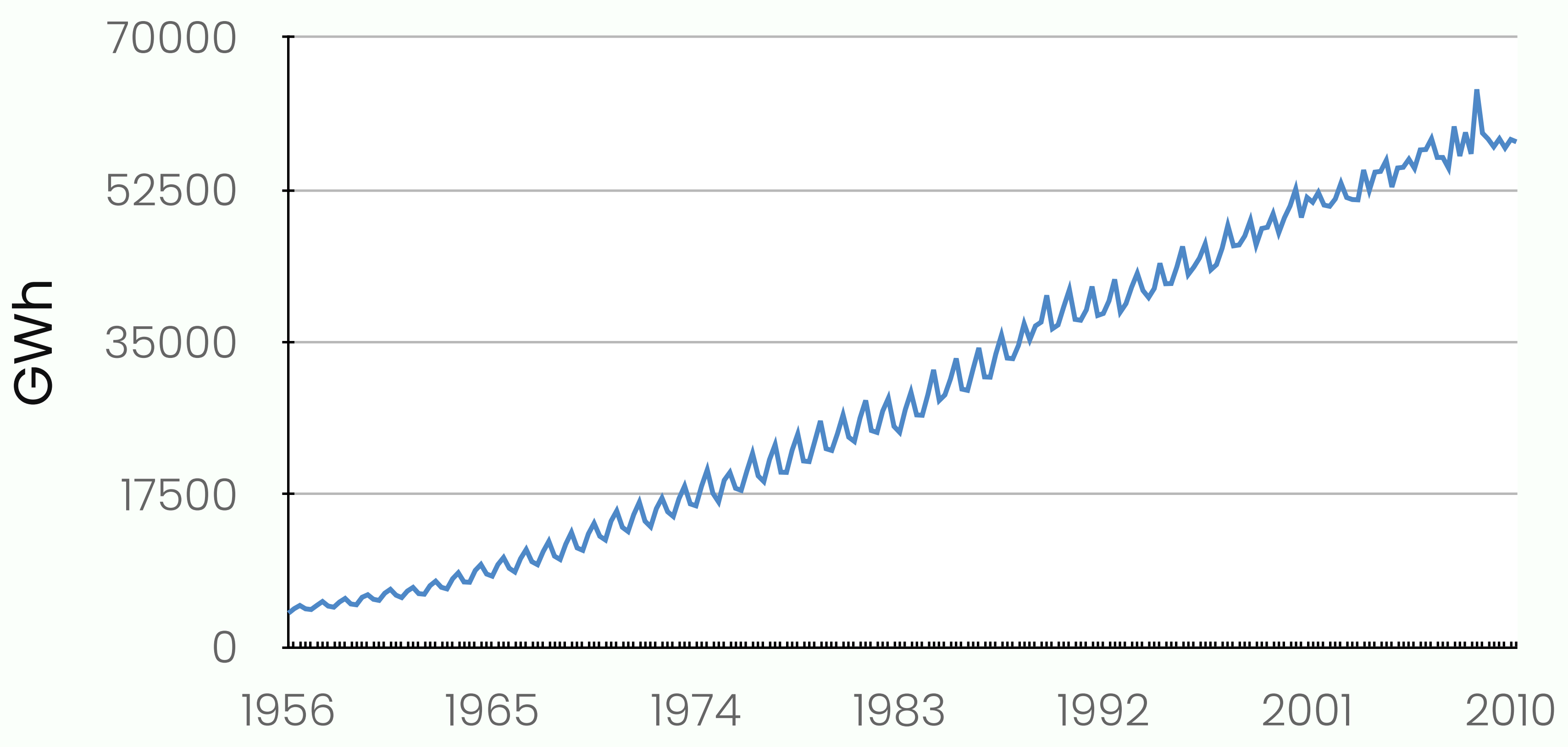
EXAMPLES

Heat production, 2022 Estonia



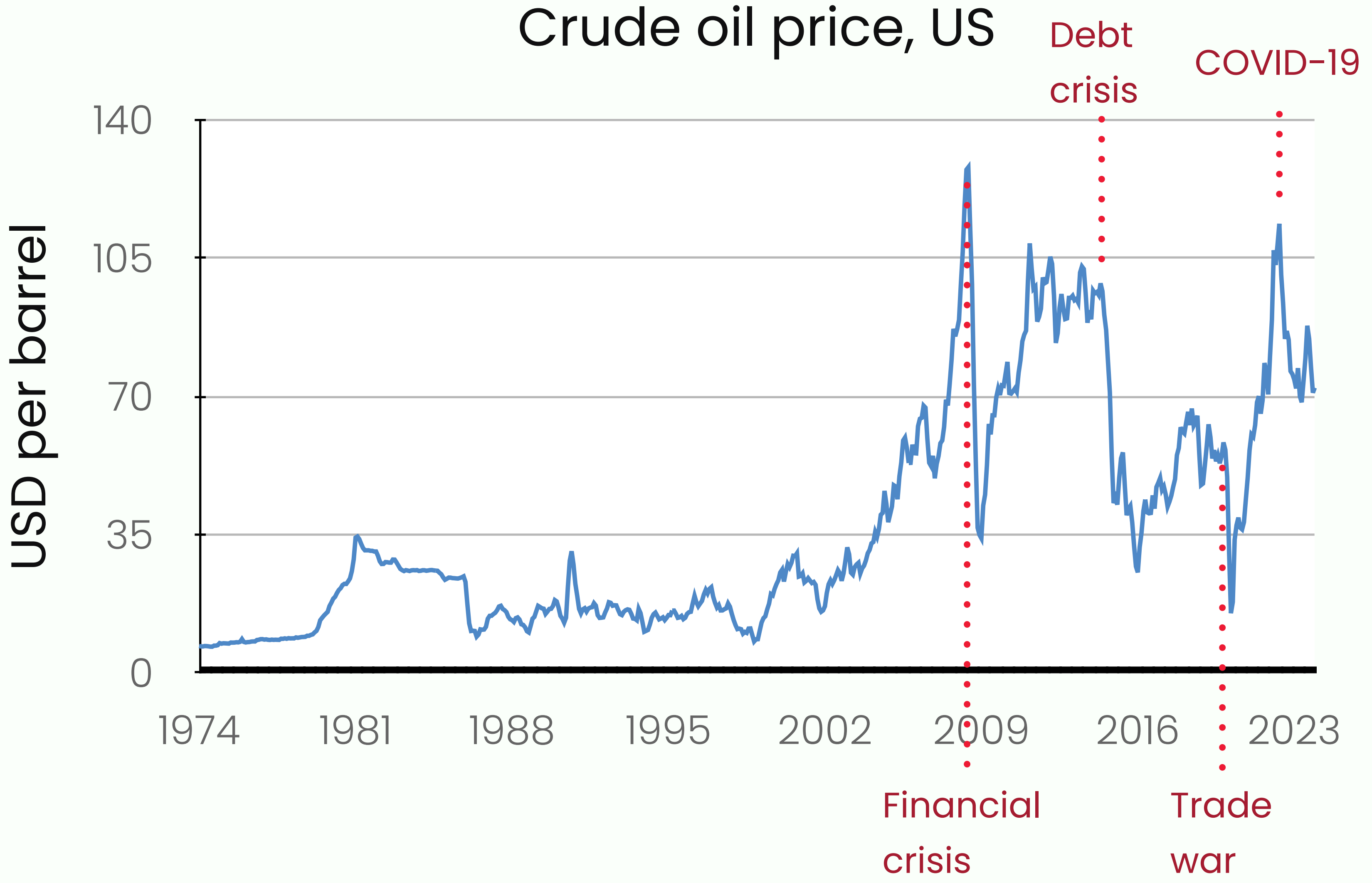
- **Trend** ✗
- **Seasonal** ✓
- **Cyclical** ✗
- **Random** ✗

Monthly electricity production, Australia



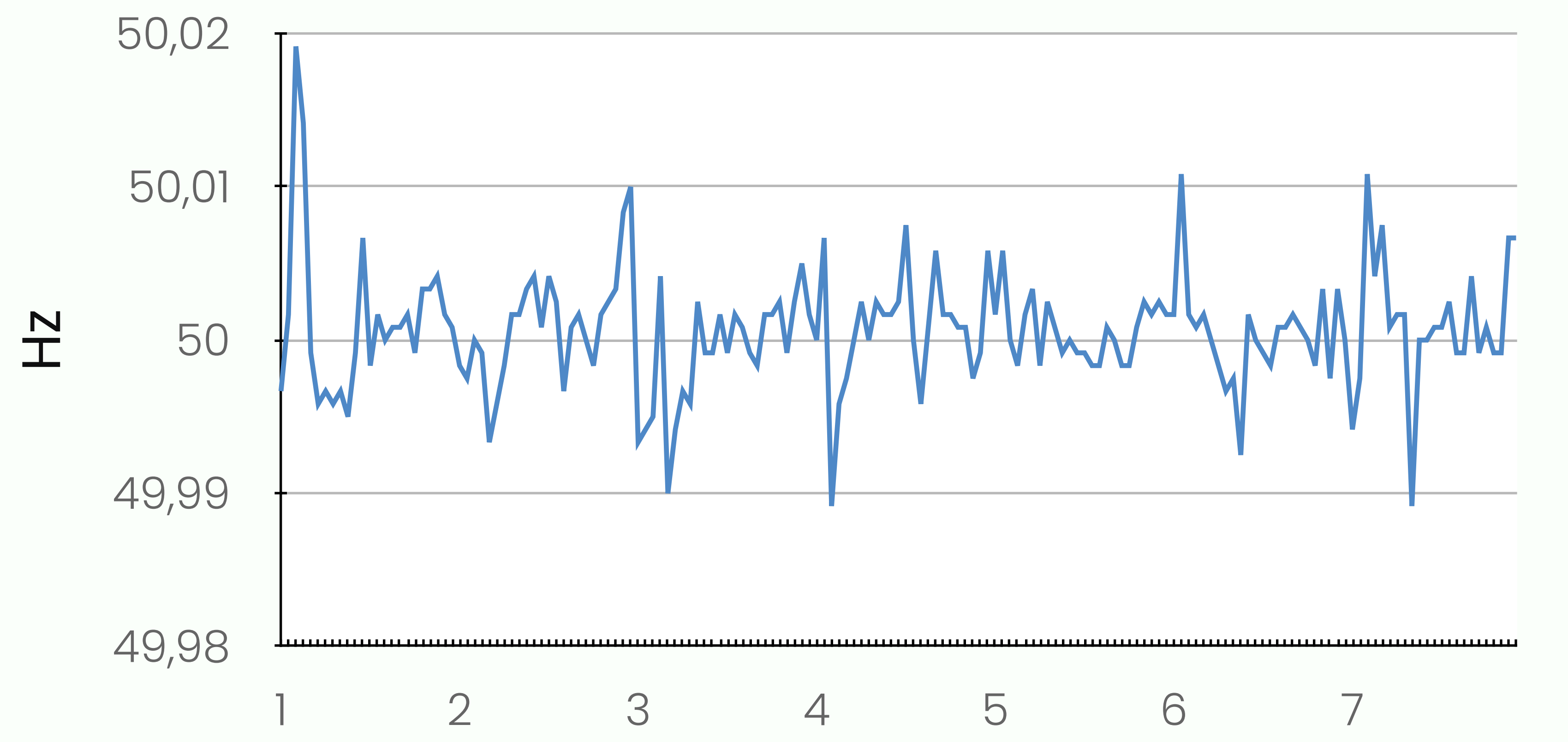
- **Trend** ✓
- **Seasonal** ✓
- **Cyclical** ✗
- **Random** ✗

Crude oil price, US



- **Trend** ✓
- **Seasonal** ✗
- **Cyclical** ✓
- **Random** ✓

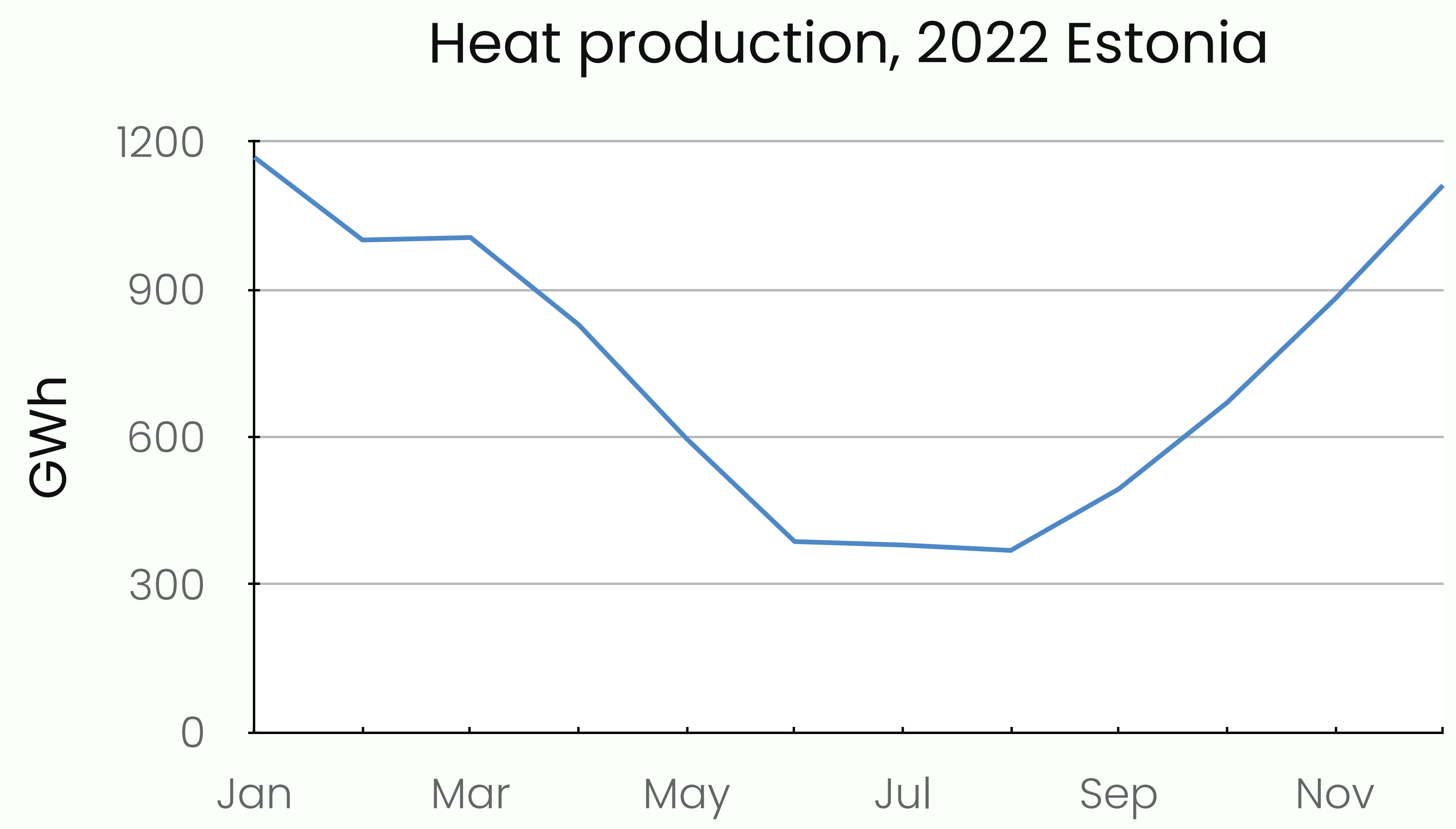
Grid frequency, Jan 2022 Estonia



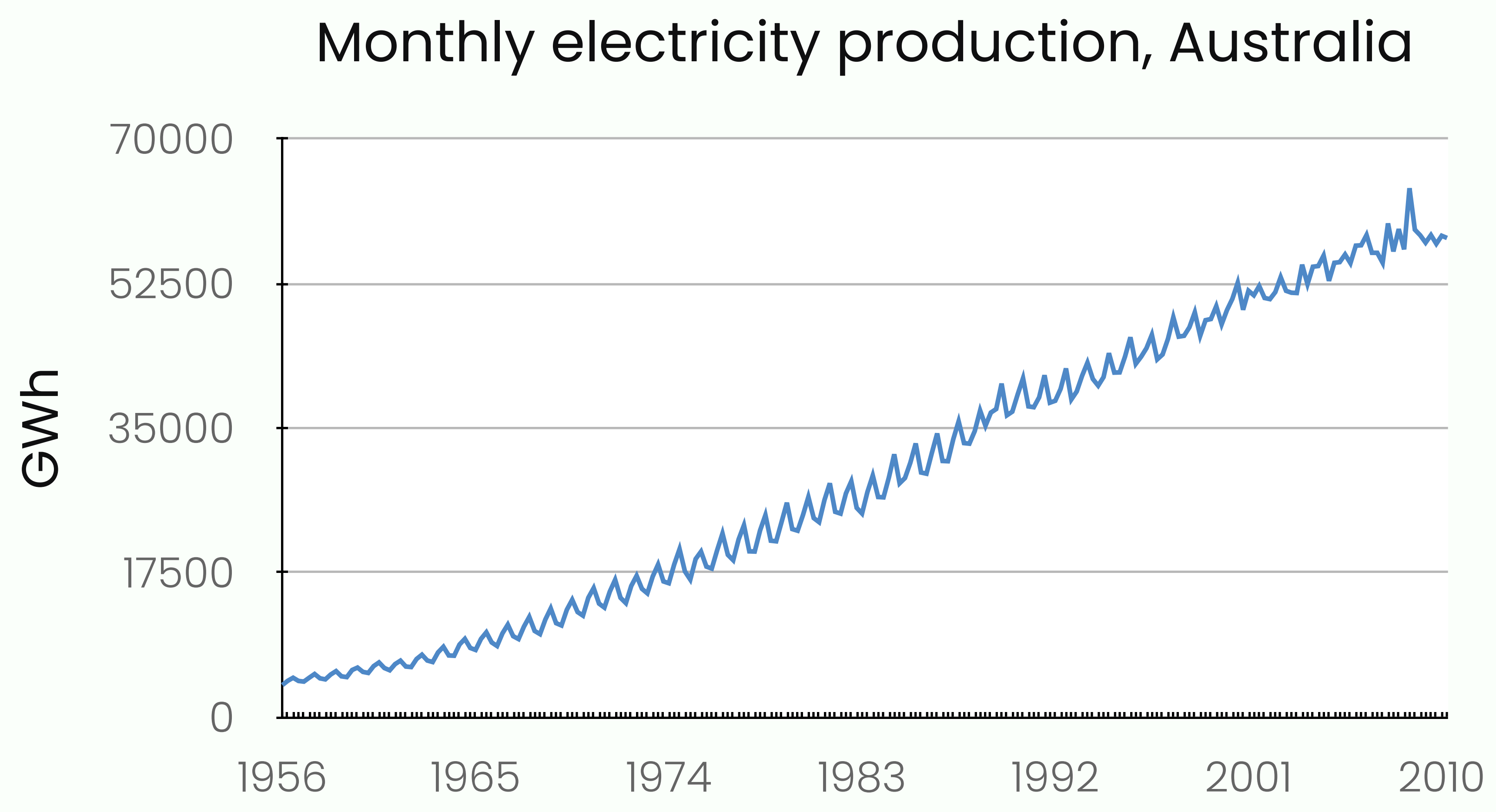
- **Trend**
- **Seasonal**
- **Cyclical**
- **Random**

[Accessed September 10, 2025], https://github.com/robjhyndman/fpp2-package
https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=pet&s=f000000__3&f=m

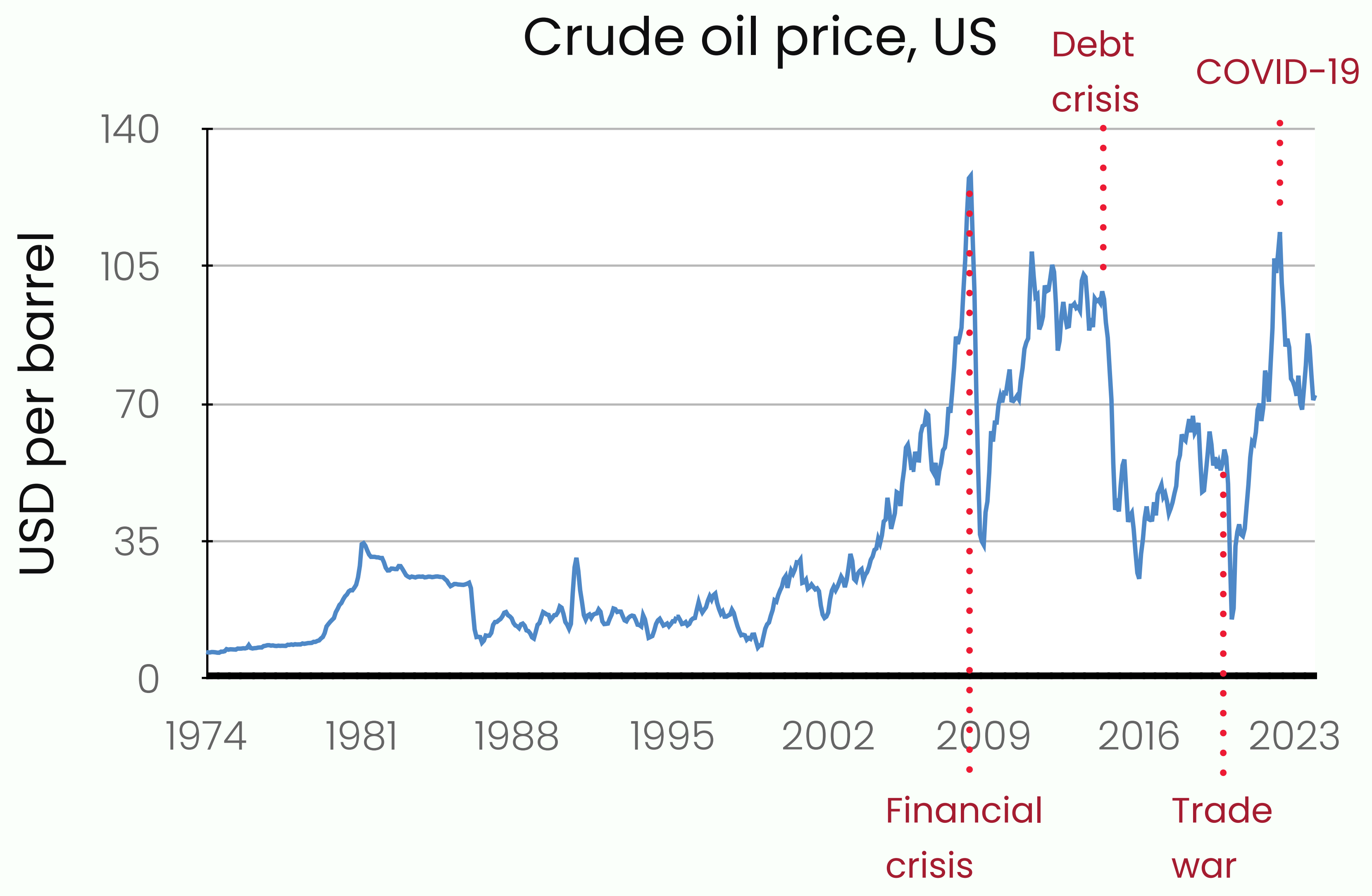
EXAMPLES



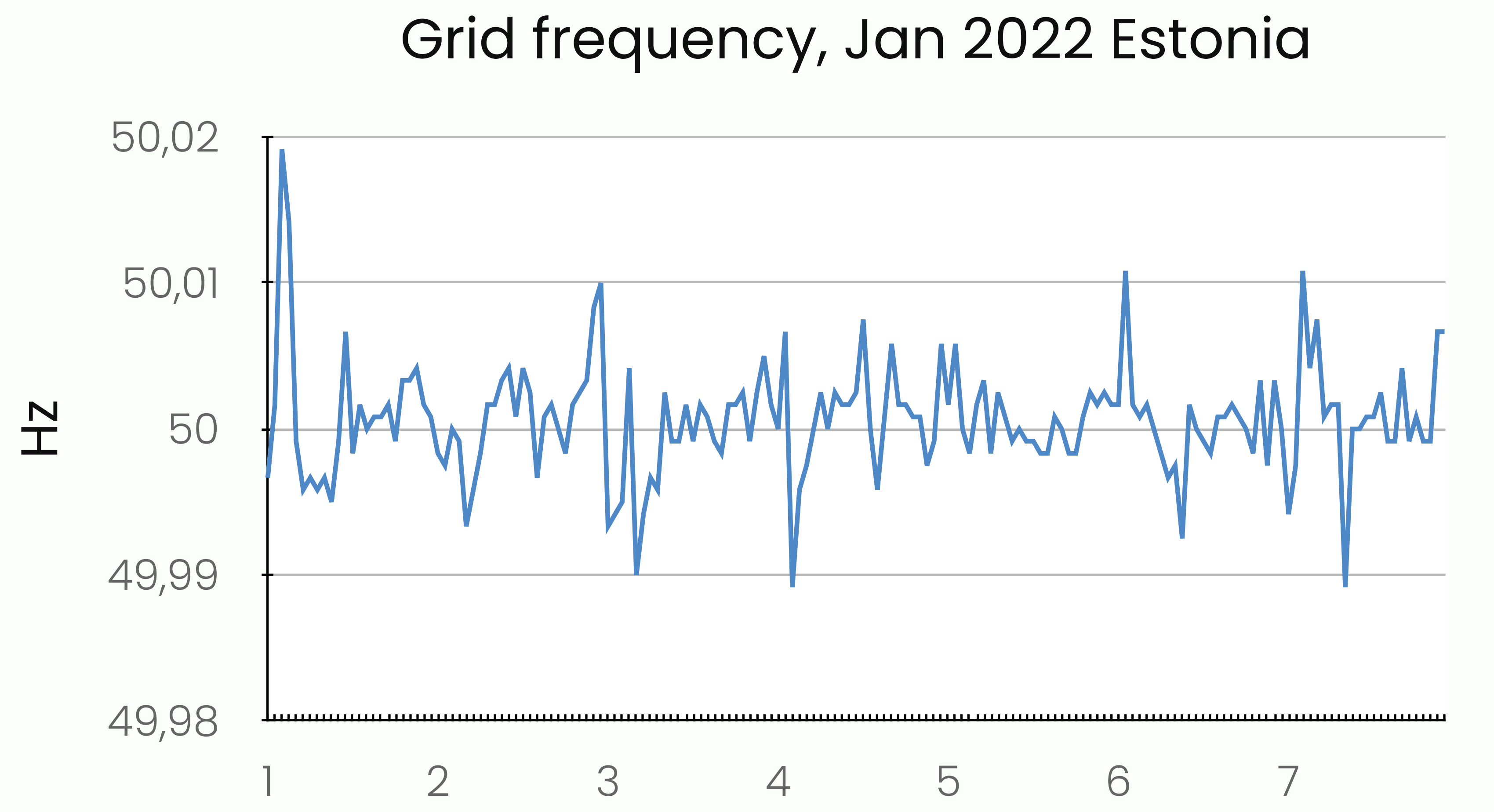
- Trend ✗
- Seasonal ✓
- Cyclical ✗
- Random ✗



- Trend ✓
- Seasonal ✓
- Cyclical ✗
- Random ✗



- Trend ✓
- Seasonal ✗
- Cyclical ✓
- Random ✓



- Trend ✗
- Seasonal ✗
- Cyclical ✗
- Random ✓

[Accessed September 10, 2025], https://github.com/robjhyndman/fpp2-package
https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=pet&s=f000000__3&f=m

Time series decomposition

TIME SERIES DECOMPOSITION

Time-series decomposition is an approach that involves explicitly representing data by separating it into distinct components, including *seasonal*, *trend*, *cycle*, and *remainder* elements.

TIME SERIES = TREND + SEASONAL

+ CYCLE + REMAINDER

DECOMPOSITION METHODS

- Classical decomposition
- X11
- x-13ARIMA-SEATS (Seasonal Extraction in ARIMA Time Series)
- STL (Seasonal and Trend decomposition using LOESS (locally estimated scatterplot smoothing))

National Statistics Offices:

  - US Census Bureau and EuroStat use X-13ARIMA-SEATS
  - Statistics Canada and ONS (UK) use X-12-ARIMA

TIME SERIES DECOMPOSITION (2)

The classical decomposition method is a relatively simple procedure that has two forms: an additive and a multiplicative.

- **Additive**:

$$y_t = T_t + S_t + R_t$$

- **Multiplicative**:

$$y_t = T_t \times S_t \times R_t$$

y_t is the time series data

S_t is the seasonal component

T_t is the trend (-cycle) component

R_t is the remainder (irregular) component

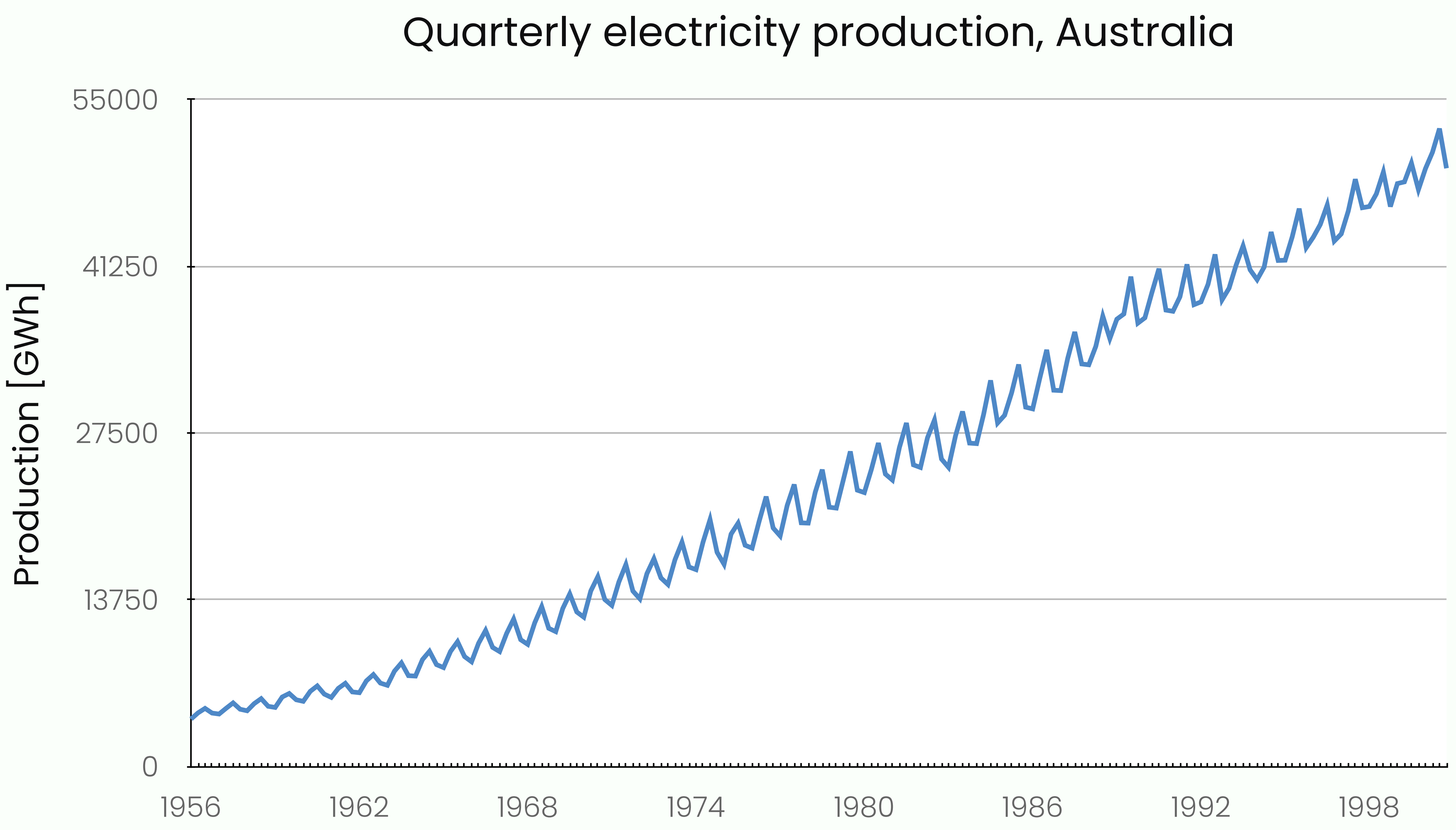
TIME SERIES DECOMPOSITION (2)

Additive:

$$y_t = T_t + S_t + R_t$$

Multiplicative:

$$y_t = T_t \times S_t \times R_t$$



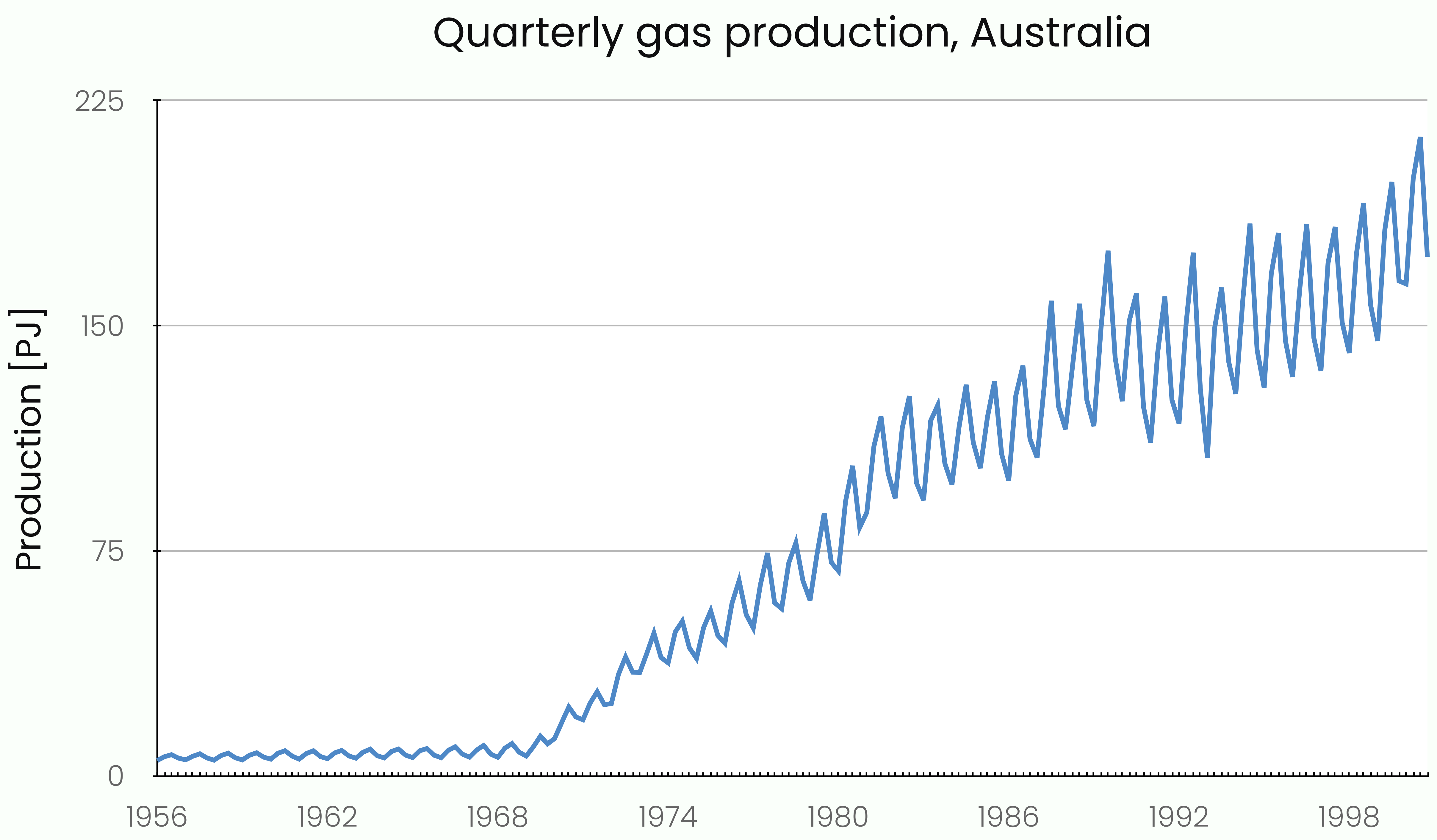
TIME SERIES DECOMPOSITION (2)

Additive:

$$y_t = T_t + S_t + R_t$$

Multiplicative:

$$y_t = T_t \times S_t \times R_t$$



[Accessed September 10, 2025], https://www.key2stats.com/data-set/view/777

TIME SERIES DECOMPOSITION (3)

Additive model is appropriate if magnitude of seasonal component does not vary with time; otherwise, multiplicative model is appropriate.

Alternative: use a Box-Cox transformation, and then use additive decomposition.

Logs turn multiplicative relationship into an additive relationship:

$$y_t = T_t \times S_t \times R_t \quad \Longrightarrow \quad \log y_t = \log T_t + \log S_t + \log R_t.$$

CLASSICAL DECOMPOSITION

The classical method of time series decomposition originated in the 1920s and was widely used until the 1950s.

It still forms the basis of many time series decomposition methods.

The first step in a classical decomposition is to use a moving average method to estimate the trend-cycle.

MOVING AVERAGE

A moving average of order m (called m-MA) can be written as:

$$\hat{T}_t = \frac{1}{m} \sum_{j=-k}^{k} y_{t+j},$$

where $m = 2k + 1$.

MOVING AVERAGE: ODD

For $m = 3$, $k = \dfrac{3-1}{2} = 1$, and 3-MA can be written as:

$$\hat{T}_t = \frac{1}{3}\sum_{j=-1}^{1} y_{t+j} = \frac{1}{3}\left(y_{t-1} + y_t + y_{t+1}\right).$$

MOVING AVERAGE: ODD

For $m = 3$, $k = \dfrac{3-1}{2} = 1$, and 3-MA can be written as:

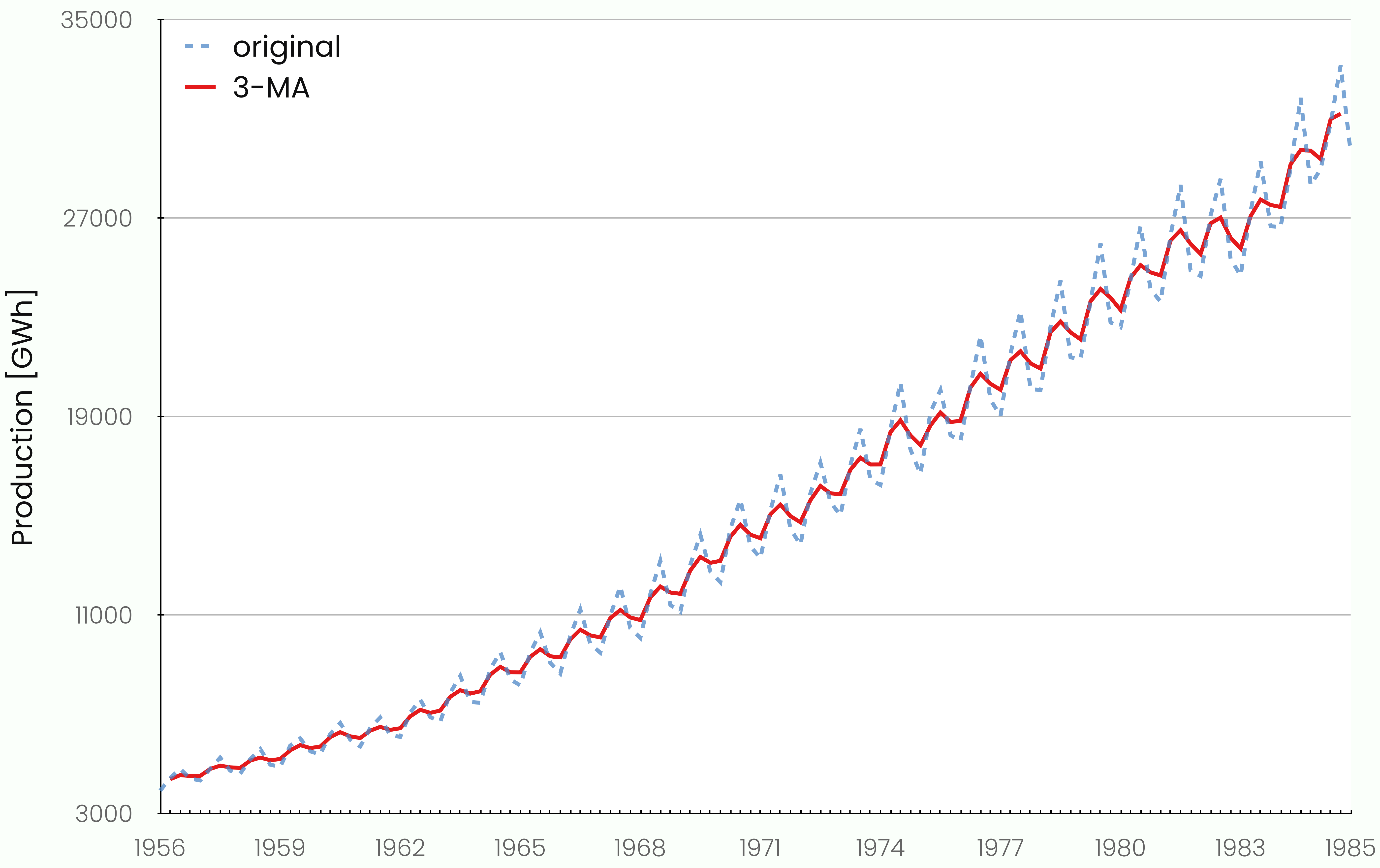
$$\hat{T}_t = \frac{1}{3}\sum_{j=-1}^{1} y_{t+j} = \frac{1}{3}\left(y_{t-1} + y_t + y_{t+1}\right).$$

Similarly, for $m = 5$, $k = (5-1)/2 = 2$, and 5-MA can be written as:

$$\hat{T}_t = \frac{1}{5}\sum_{j=-1}^{1} y_{t+j} = \frac{1}{5}\left(y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2}\right).$$

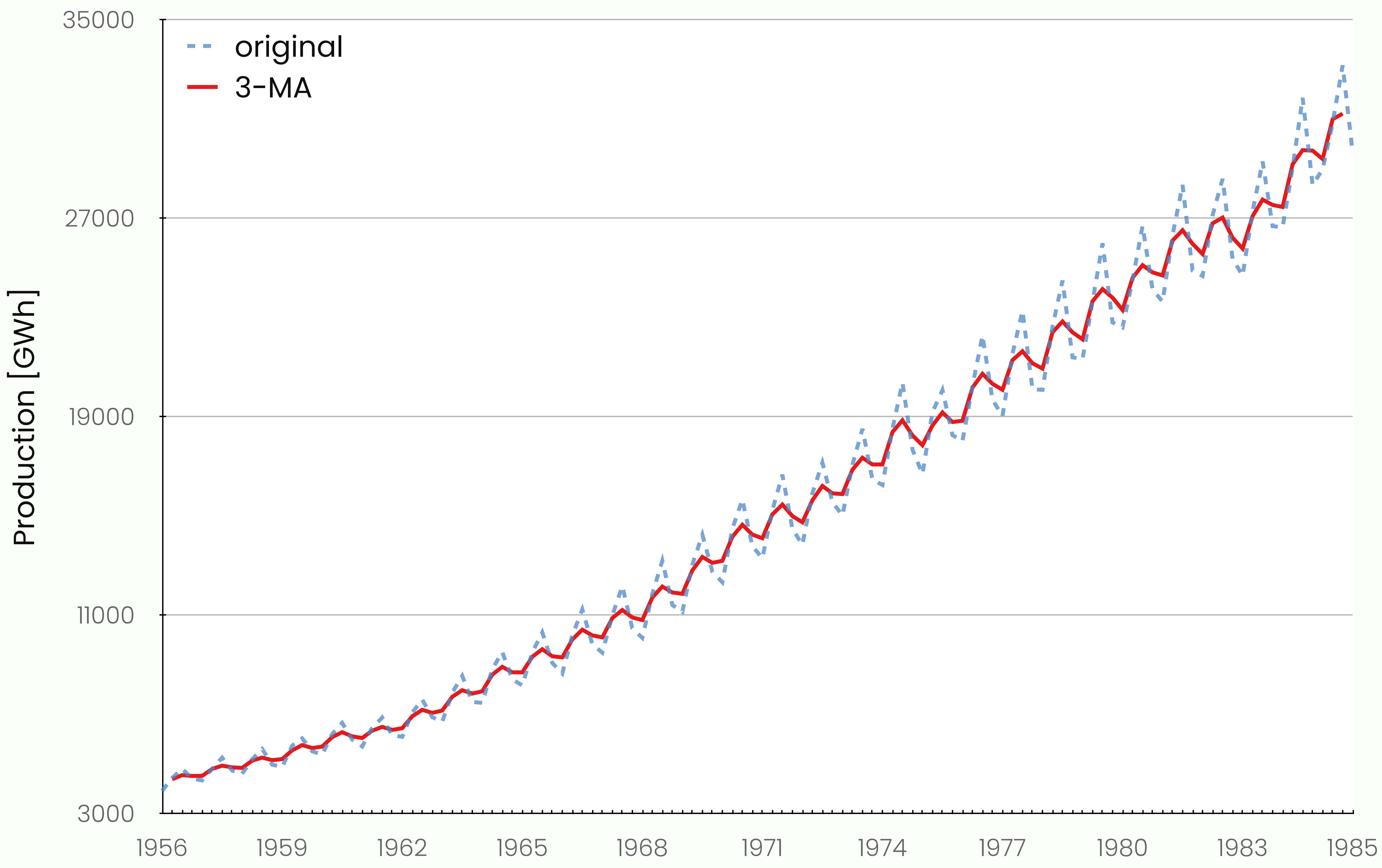
MOVING AVERAGE: ODD (2)

Year	Production	3-MA
Q1 1956	3923	
Q2 1956	4436	
Q3 1956	4806	
Q4 1956	4418	
Q1 1957	4339	
Q2 1957	4811	
...	...	
Q2 1985	30838	
Q3 1985	33165	
Q4 1985	29648	



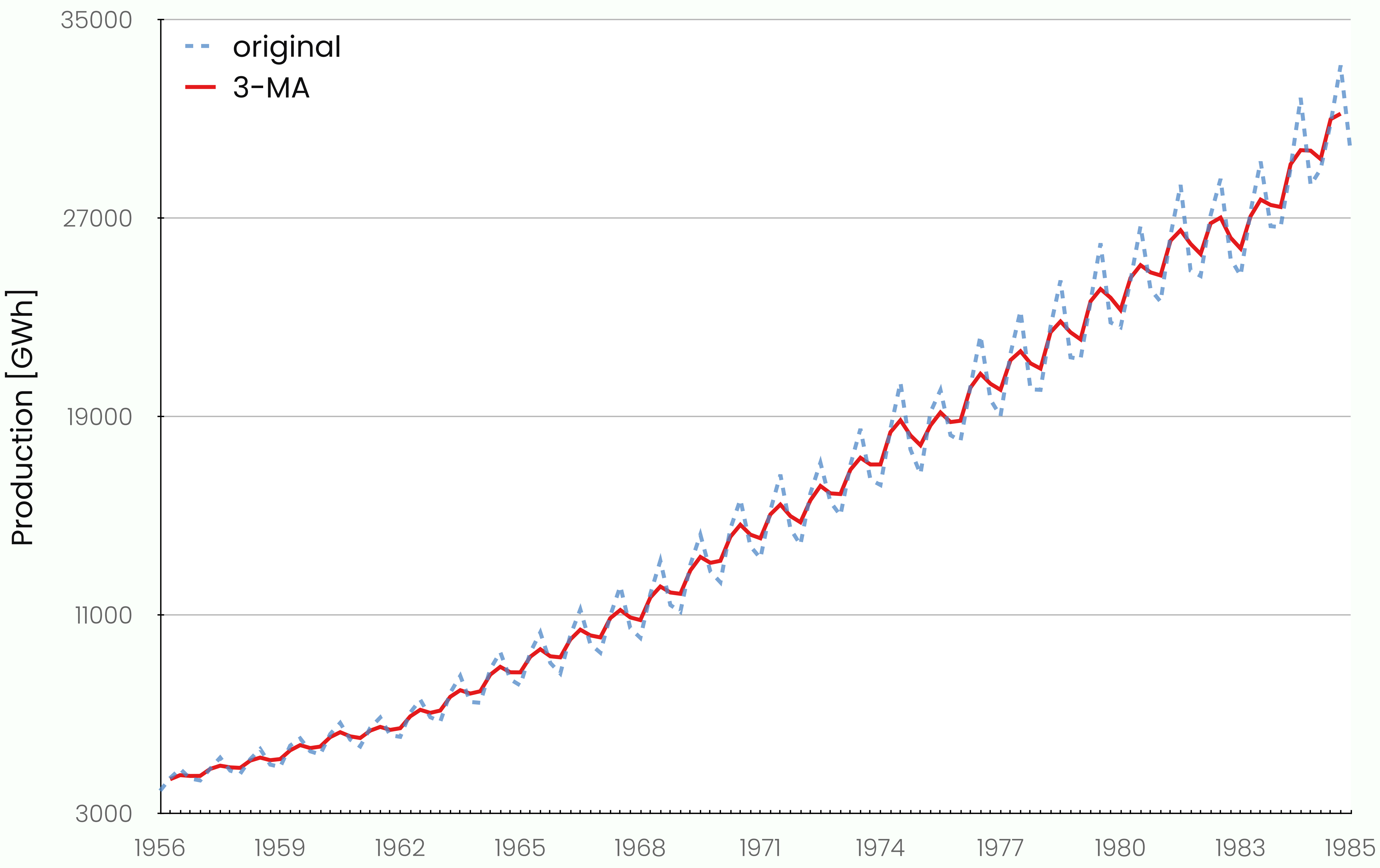
MOVING AVERAGE: ODD (2)

Year	Production	3-MA
Q1 1956	3923	
Q2 1956	4436	4388
Q3 1956	4806	
Q4 1956	4418	
Q1 1957	4339	
Q2 1957	4811	
...	...	
Q2 1985	30838	
Q3 1985	33165	
Q4 1985	29648	



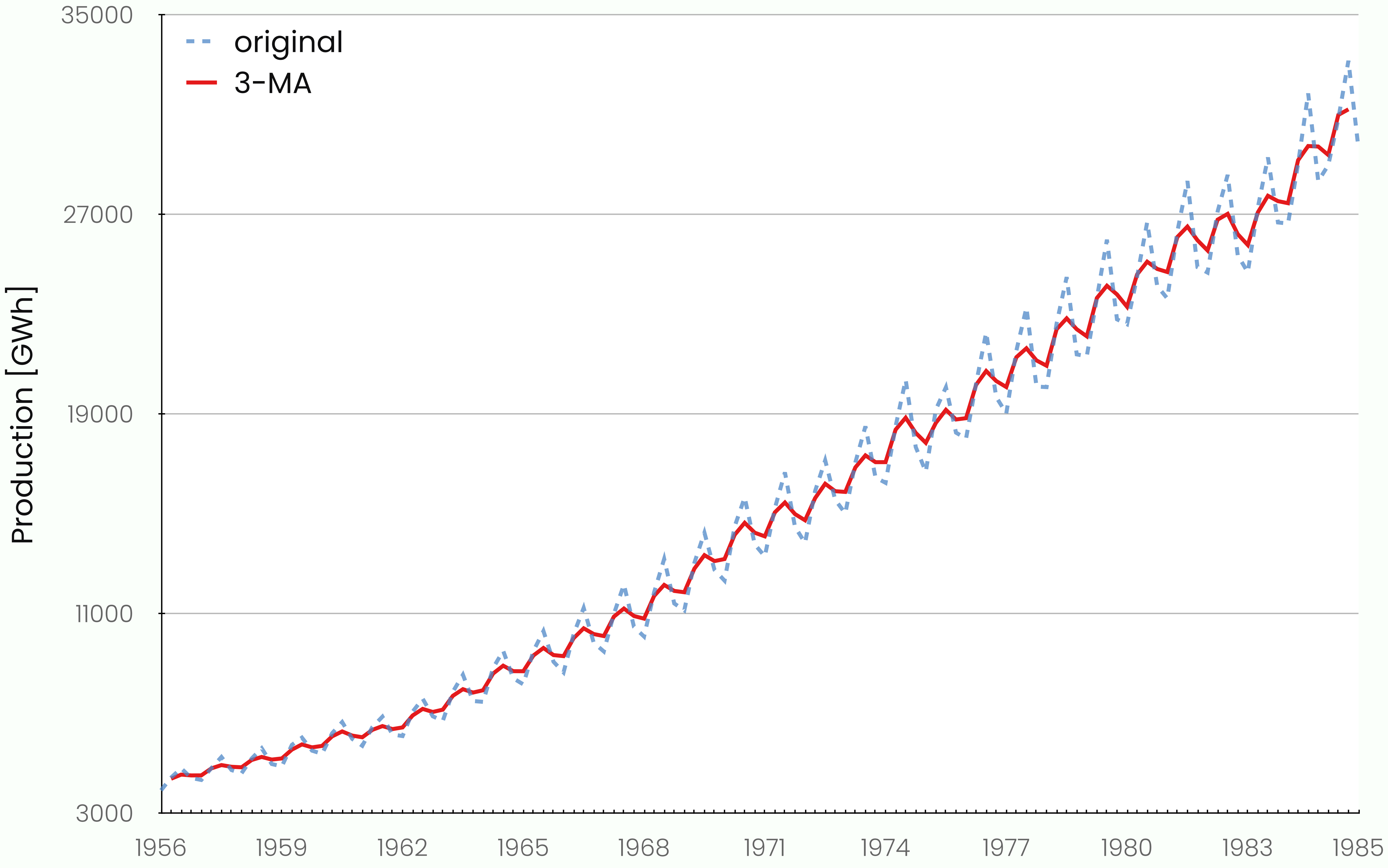
MOVING AVERAGE: ODD (2)

Year	Production	3-MA
Q1 1956	3923	
Q2 1956	4436	4388
Q3 1956	4806	4553
Q4 1956	4418	
Q1 1957	4339	
Q2 1957	4811	
...	...	
Q2 1985	30838	
Q3 1985	33165	
Q4 1985	29648	



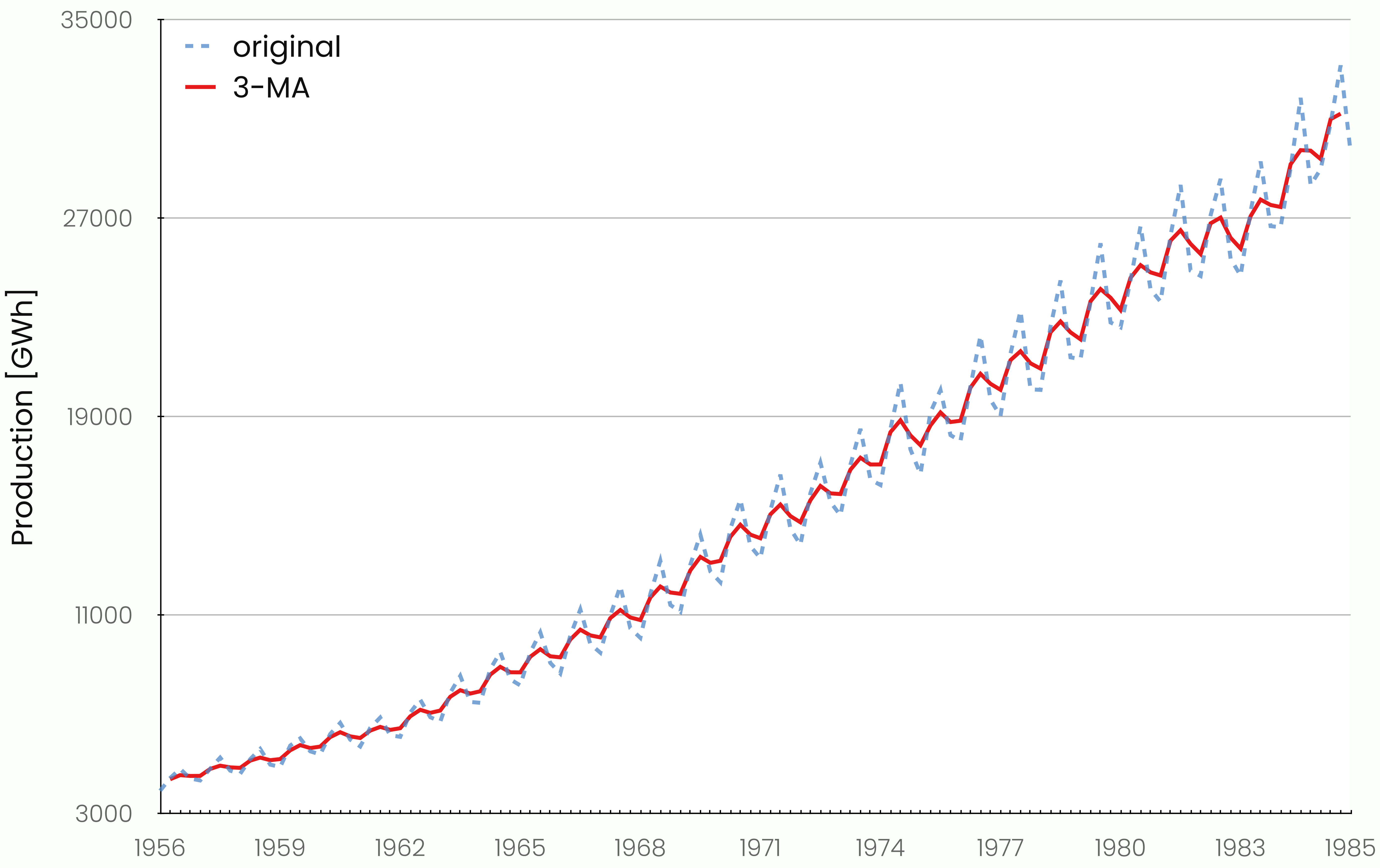
MOVING AVERAGE: ODD (2)

Year	Production	3-MA
Q1 1956	3923	
Q2 1956	4436	4388
Q3 1956	4806	4553
Q4 1956	4418	4521
Q1 1957	4339	
Q2 1957	4811	
...	...	
Q2 1985	30838	
Q3 1985	33165	
Q4 1985	29648	



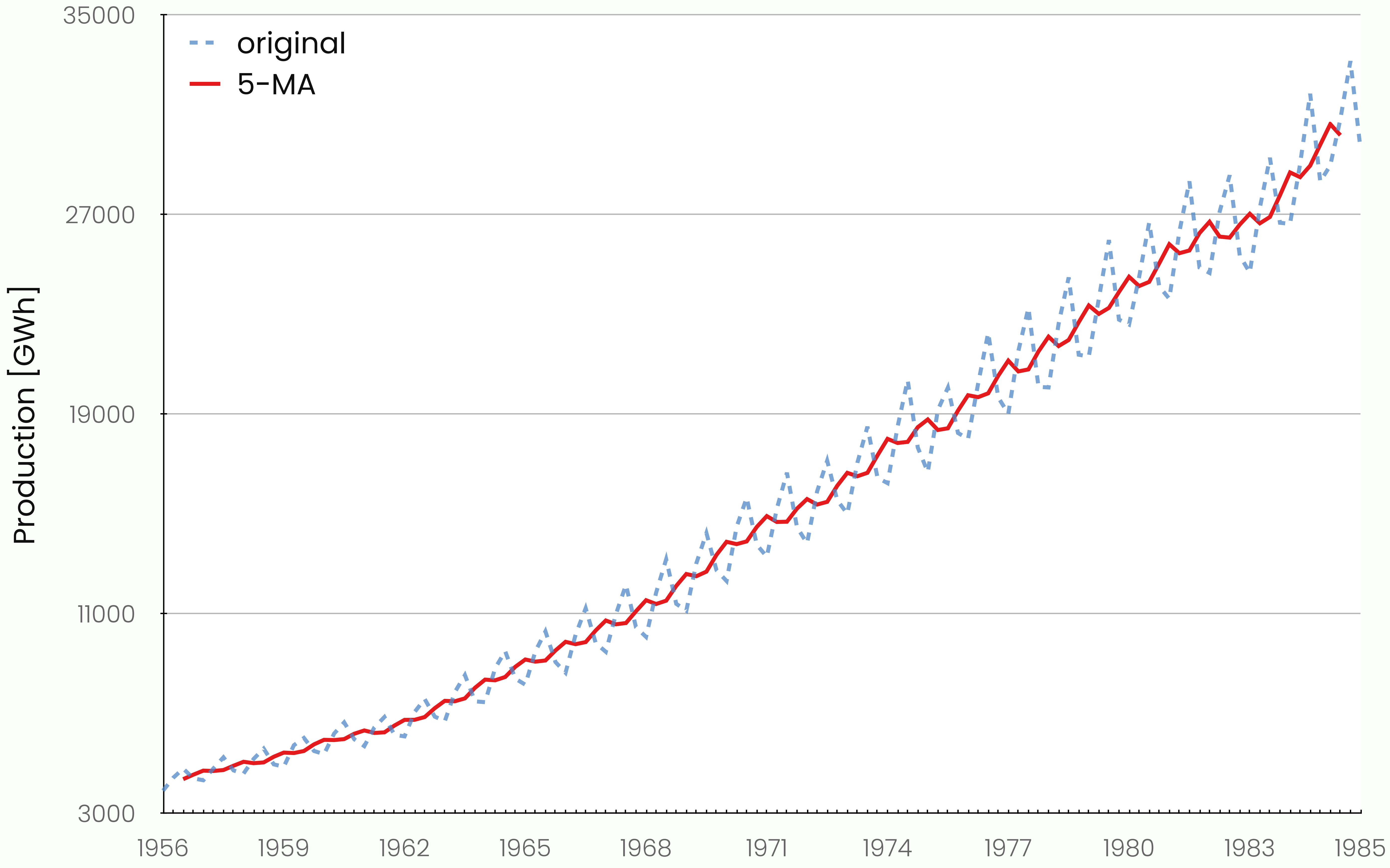
MOVING AVERAGE: ODD (2)

Year	Production	3-MA
Q1 1956	3923	
Q2 1956	4436	4388
Q3 1956	4806	4553
Q4 1956	4418	4521
Q1 1957	4339	4523
Q2 1957	4811	4803
...	...	...
Q2 1985	30838	30995
Q3 1985	33165	31217
Q4 1985	29648	



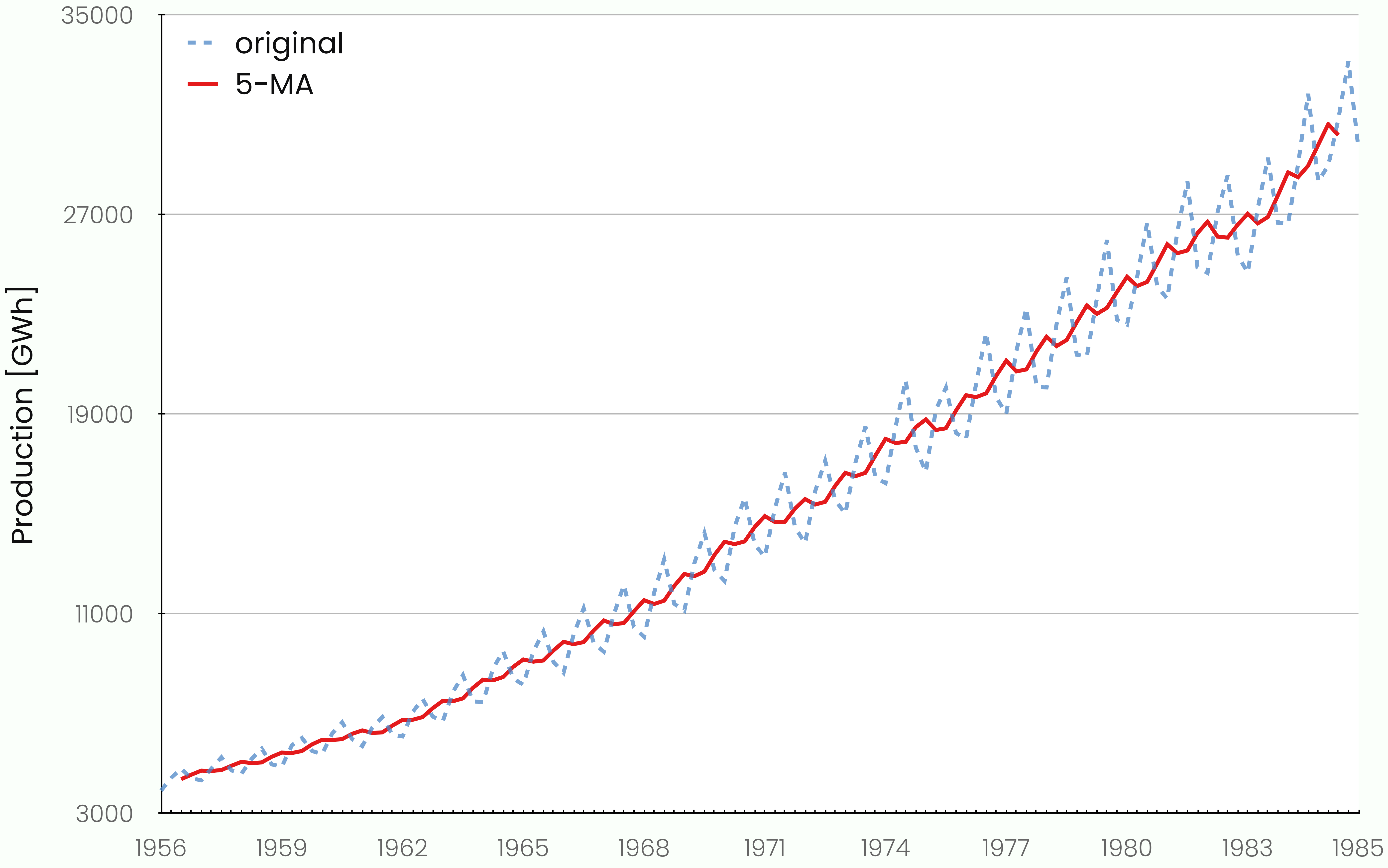
MOVING AVERAGE: ODD (3)

Year	Production	5-MA
Q1 1956	3923	
Q2 1956	4436	
Q3 1956	4806	
Q4 1956	4418	
Q1 1957	4339	
Q2 1957	4811	
...	...	
...	...	
Q2 1985	30838	
Q3 1985	33165	
Q4 1985	29648	



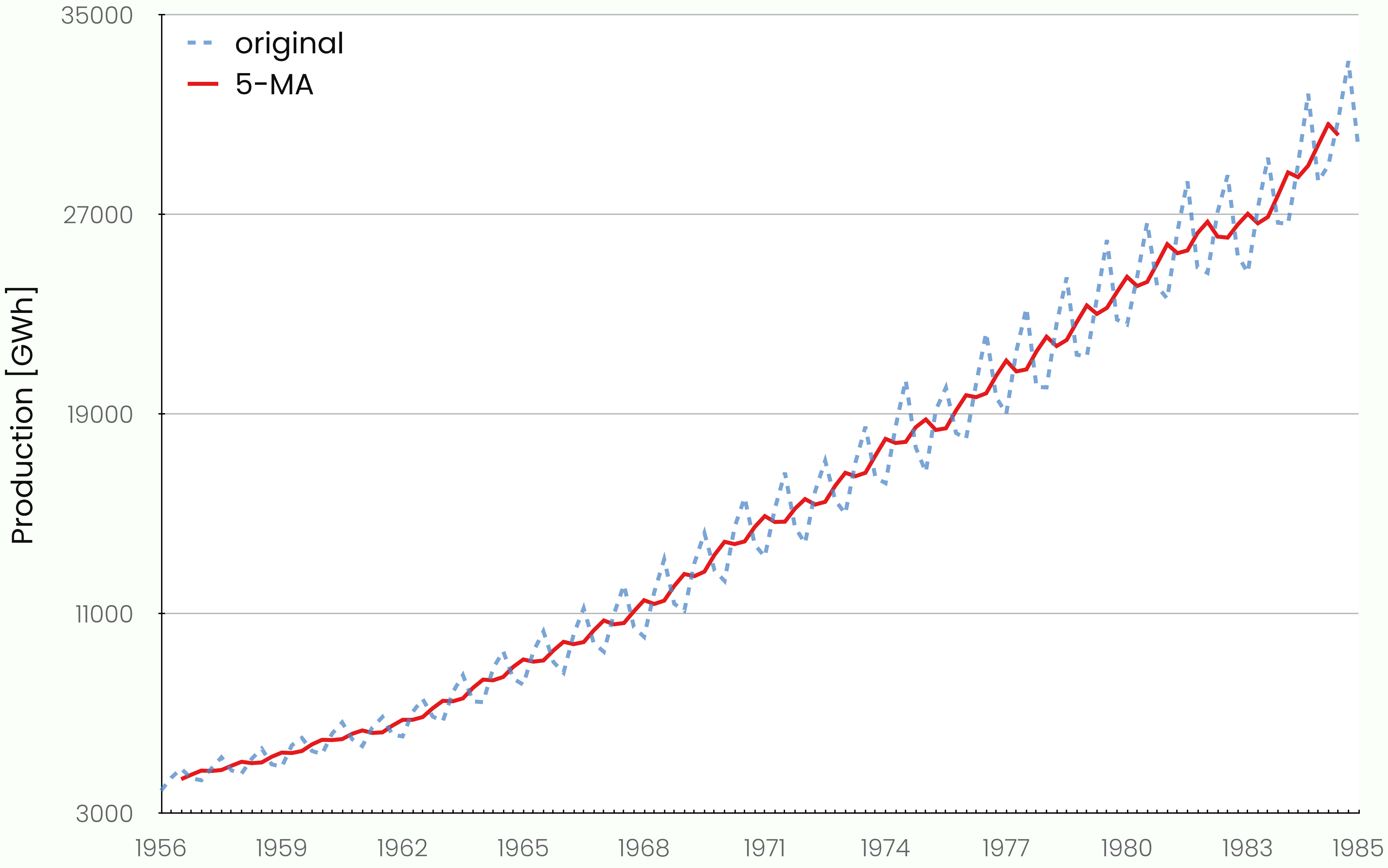
MOVING AVERAGE: ODD (3)

Year	Production	5-MA
Q1 1956	3923	
Q2 1956	4436	
Q3 1956	4806	4553
Q4 1956	4418	
Q1 1957	4339	
Q2 1957	4811	
...	...	
...	...	
Q2 1985	30838	
Q3 1985	33165	
Q4 1985	29648	



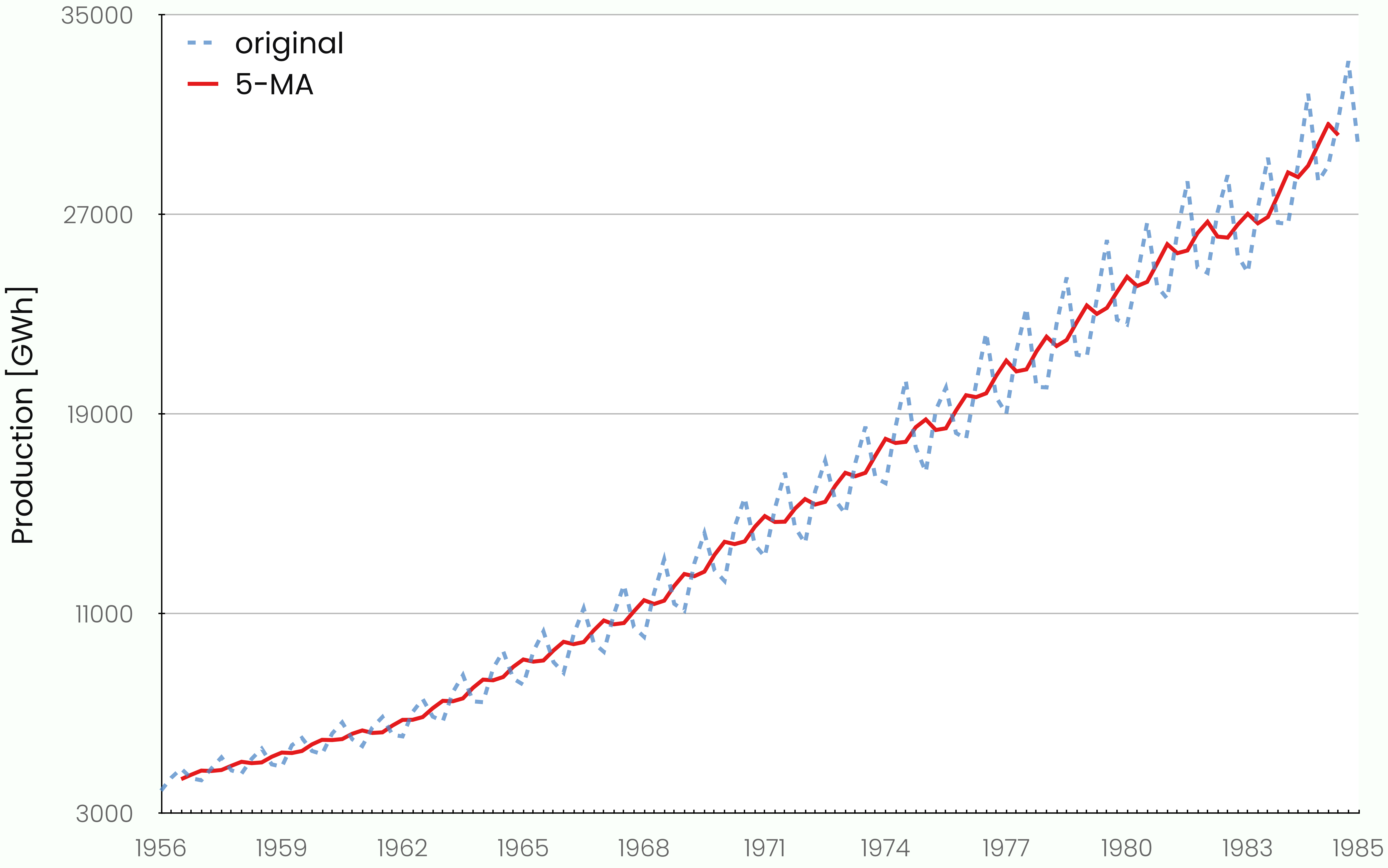
MOVING AVERAGE: ODD (3)

Year	Production	5-MA
Q1 1956	3923	
Q2 1956	4436	
Q3 1956	4806	4553
Q4 1956	4418	4521
Q1 1957	4339	
Q2 1957	4811	
...	...	
...	...	
Q2 1985	30838	
Q3 1985	33165	
Q4 1985	29648	



MOVING AVERAGE: ODD (3)

Year	Production	5-MA
Q1 1956	3923	
Q2 1956	4436	
Q3 1956	4806	4553
Q4 1956	4418	4521
Q1 1957	4339	4523
Q2 1957	4811	4803
...	...	...
...	...	
Q2 1985	30838	30995
Q3 1985	33165	
Q4 1985	29648	



MOVING AVERAGE: EVEN

How about even orders? For example, for $m = 2$, we have:

$$\frac{1}{4}\left(y_{t-2} + y_{t-1} + y_t + y_{t+1}\right) \qquad \text{or} \qquad \frac{1}{4}\left(y_{t-1} + y_t + y_{t+1} + y_{t+2}\right).$$

MOVING AVERAGE: EVEN

How about even orders? For example, for $m = 2$, we have:

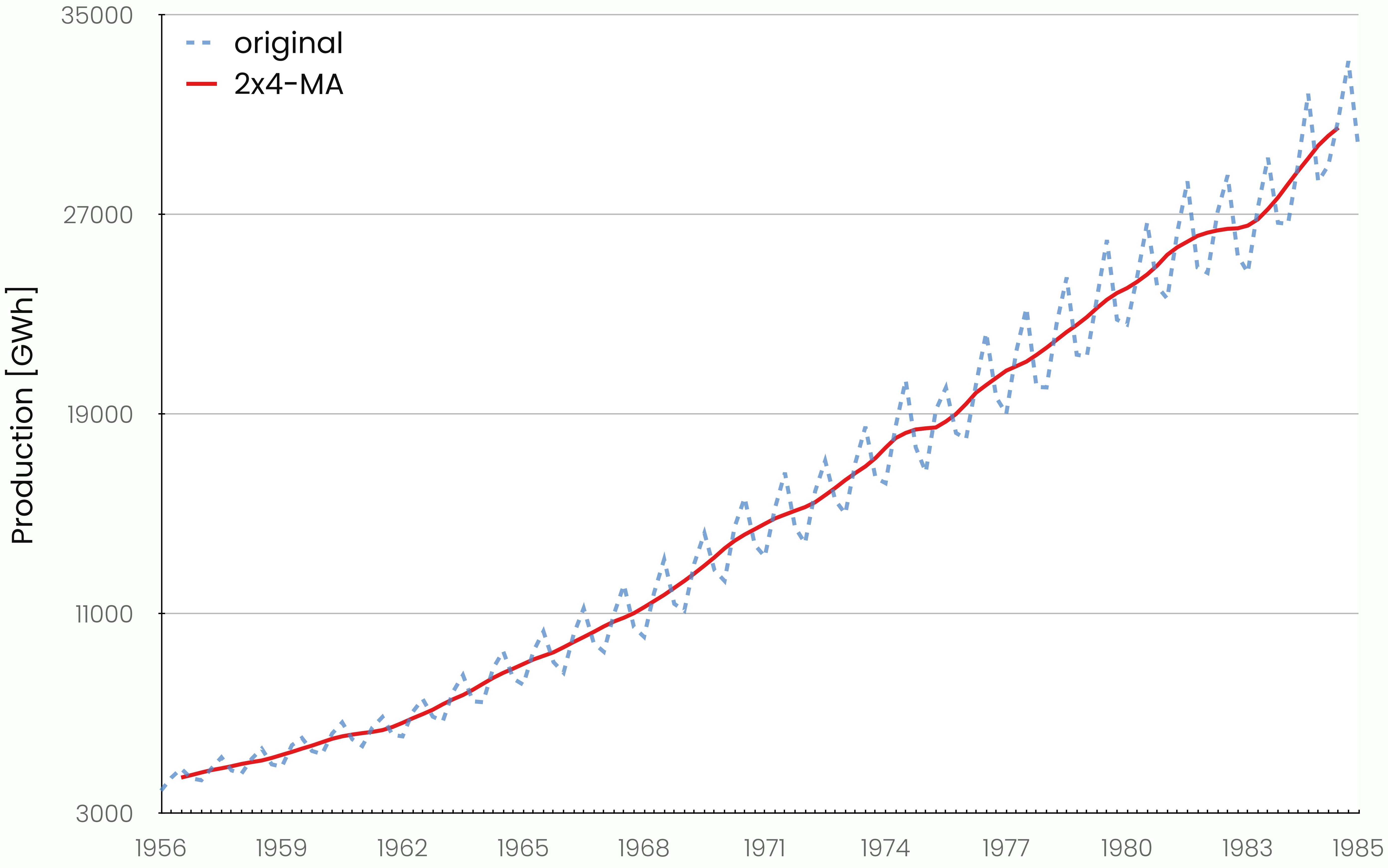
$$\frac{1}{4}\left(y_{t-2} + y_{t-1} + y_t + y_{t+1}\right) \quad \text{or} \quad \frac{1}{4}\left(y_{t-1} + y_t + y_{t+1} + y_{t+2}\right).$$

Solution: Take a further 2-MA to *centre* result:

$$\begin{aligned}\hat{T}_t &= \frac{1}{2}\left\{\frac{1}{4}\left(y_{t-2} + y_{t-1} + y_t + y_{t+1}\right) + \frac{1}{4}\left(y_{t-1} + y_t + y_{t+1} + y_{t+2}\right)\right\} \\ &= \frac{1}{8}y_{t-2} + \frac{1}{4}y_{t-1} + \frac{1}{4}y_t + \frac{1}{4}y_{t+1} + \frac{1}{8}y_{t+2}.\end{aligned}$$

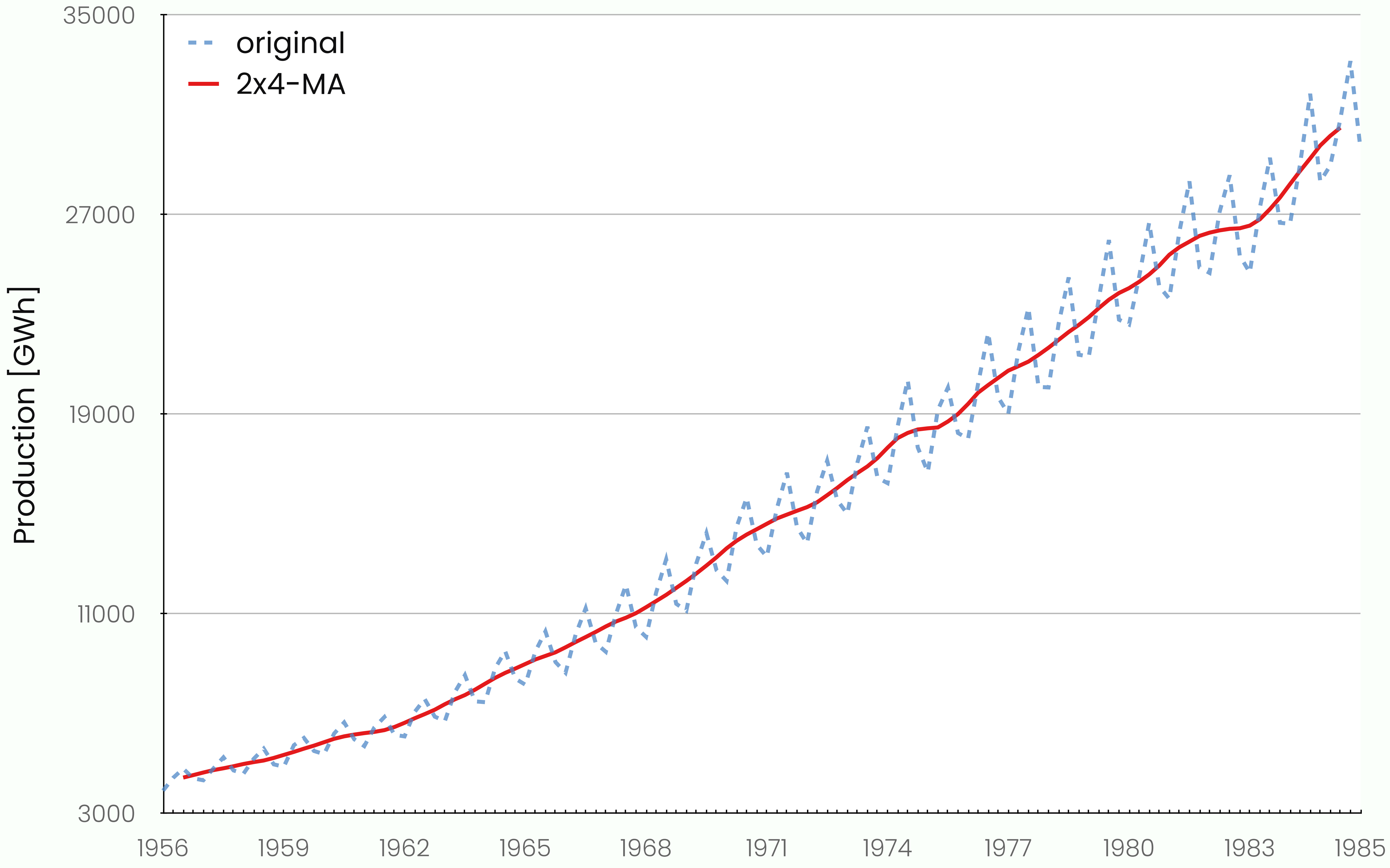
MOVING AVERAGE: EVEN (2)

Year	Production	4-MA	2x4-MA
Q1 1956	3923		
Q2 1956	4436		
Q3 1956	4806		
Q4 1956	4418		
Q1 1957	4339		
Q2 1957	4811		
...	...		



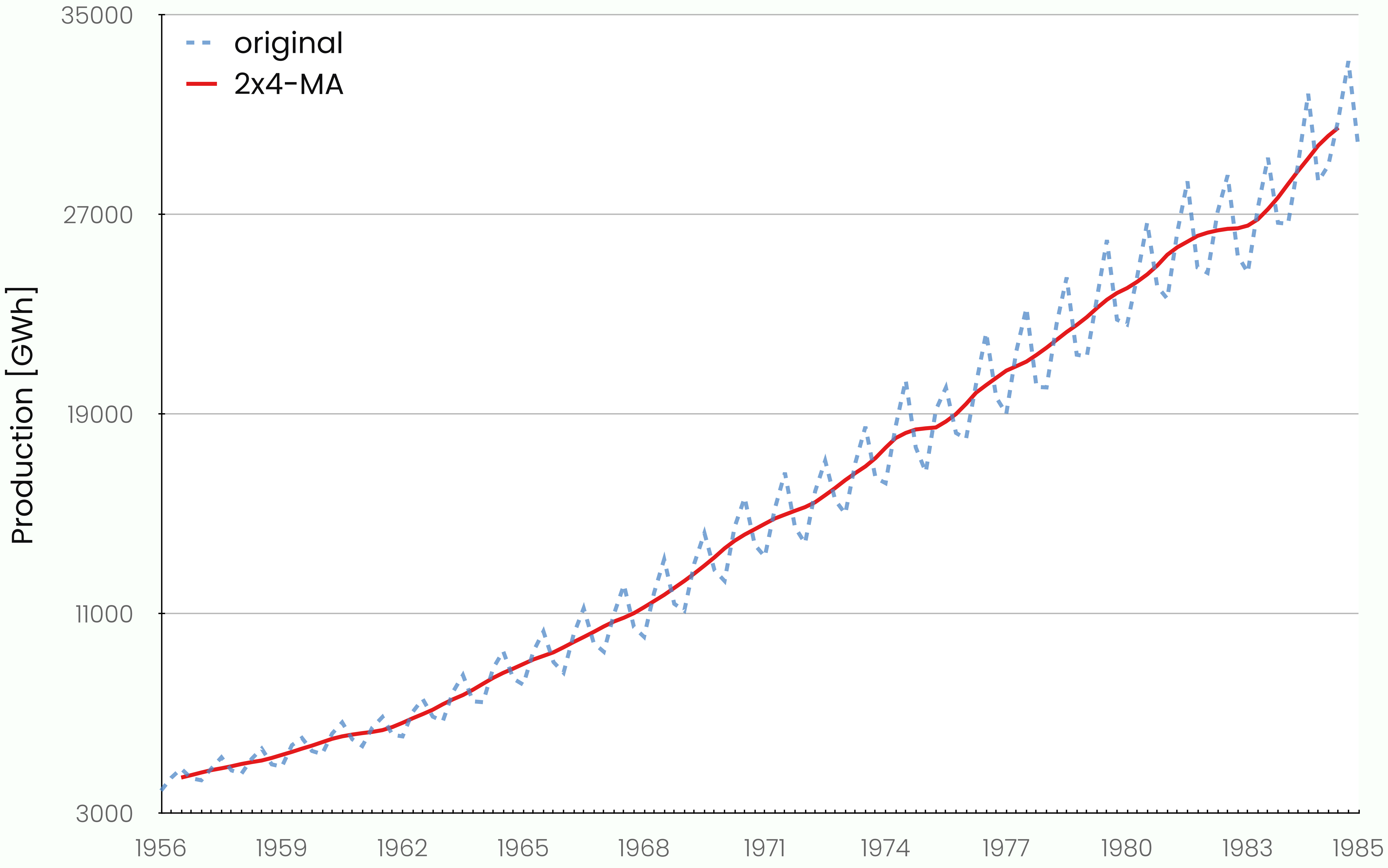
MOVING AVERAGE: EVEN (2)

Year	Production	4-MA	2x4-MA
Q1 1956	3923		
Q2 1956	4436		
		4396	
Q3 1956	4806		
Q4 1956	4418		
Q1 1957	4339		
Q2 1957	4811		
...	...		



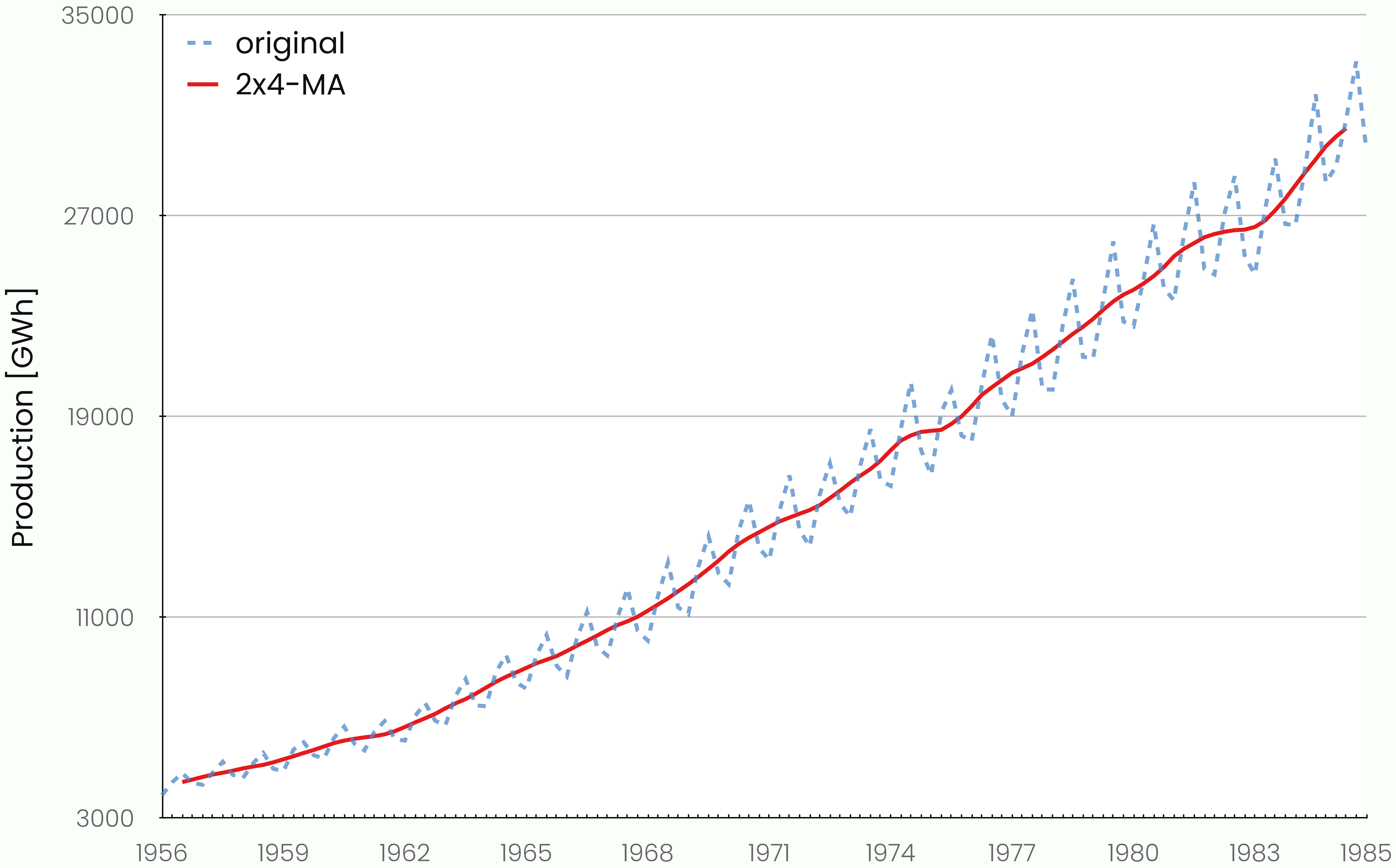
MOVING AVERAGE: EVEN (2)

Year	Production	4-MA	2x4-MA
Q1 1956	3923		
Q2 1956	4436	4396	
Q3 1956	4806	4500	
Q4 1956	4418		
Q1 1957	4339		
Q2 1957	4811		
…	…		



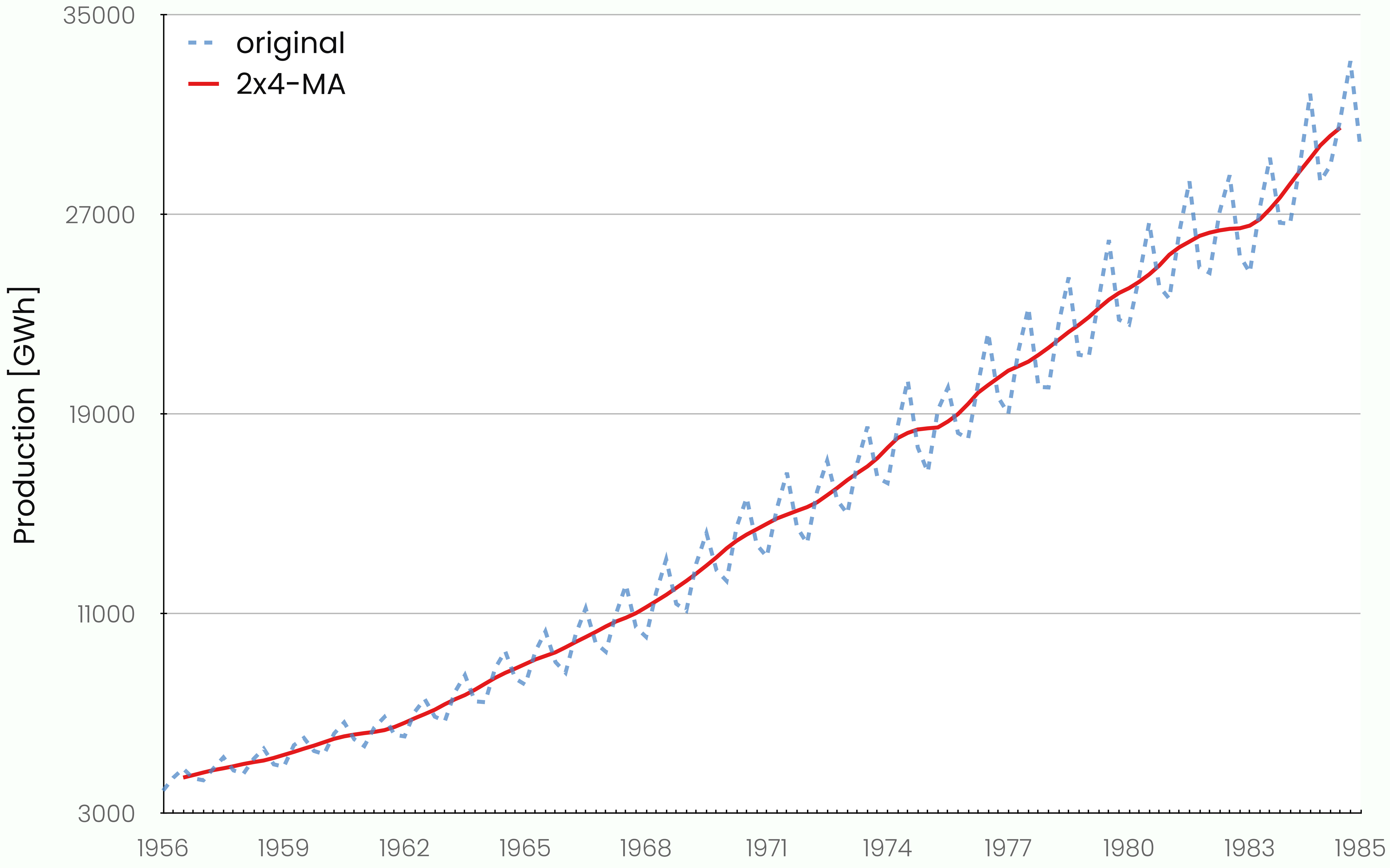
MOVING AVERAGE: EVEN (2)

Year	Production	4-MA	2x4-MA
Q1 1956	3923		
Q2 1956	4436	4396	
Q3 1956	4806	4500	
Q4 1956	4418	4594	
Q1 1957	4339	4707	
Q2 1957	4811	4786	
...	...	...	



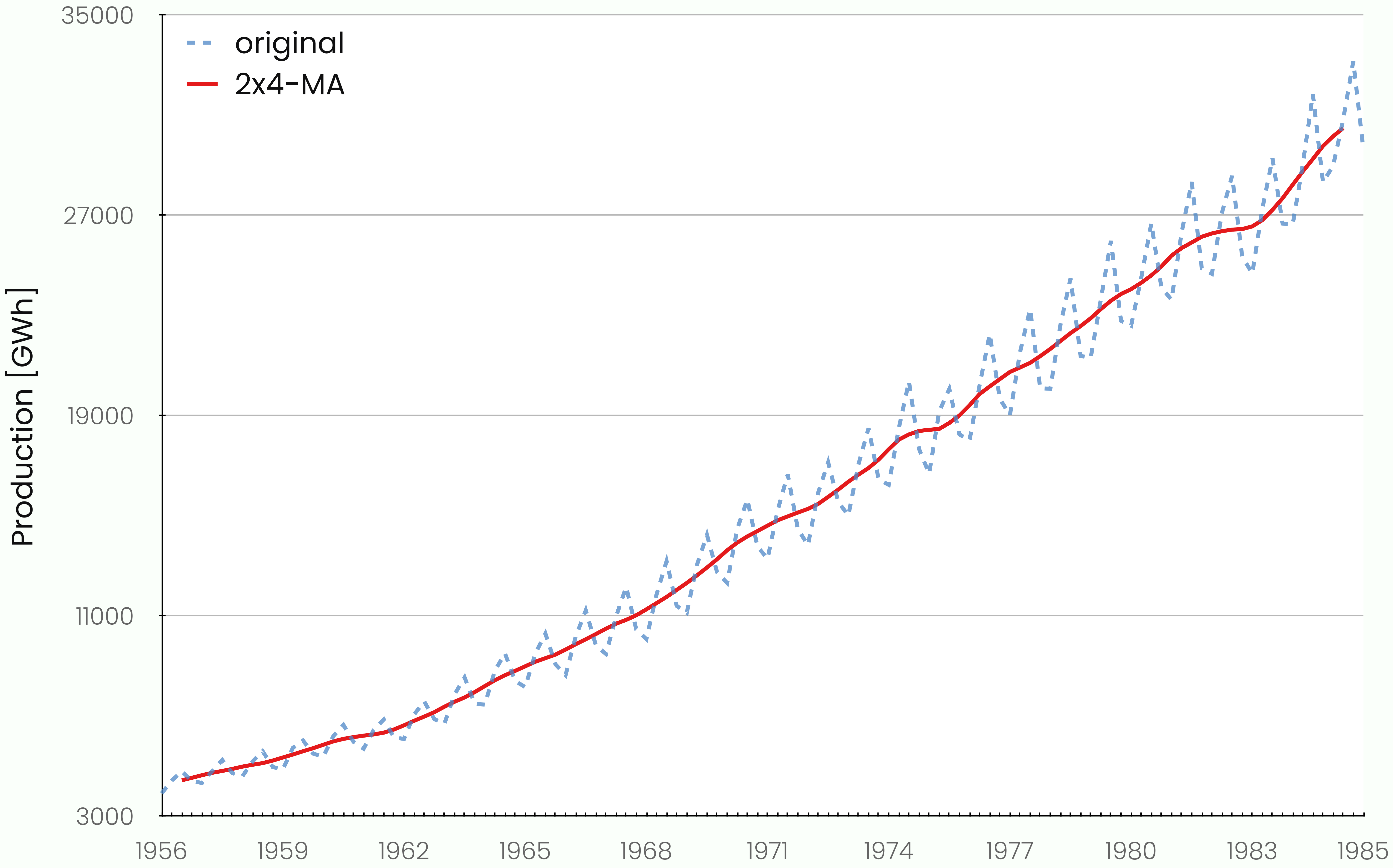
MOVING AVERAGE: EVEN (2)

Year	Production	4-MA	2x4-MA
Q1 1956	3923		
Q2 1956	4436	4396	
Q3 1956	4806	4500	
Q4 1956	4418	4594	
Q1 1957	4339	4707	
Q2 1957	4811	4786	
...	...	...	



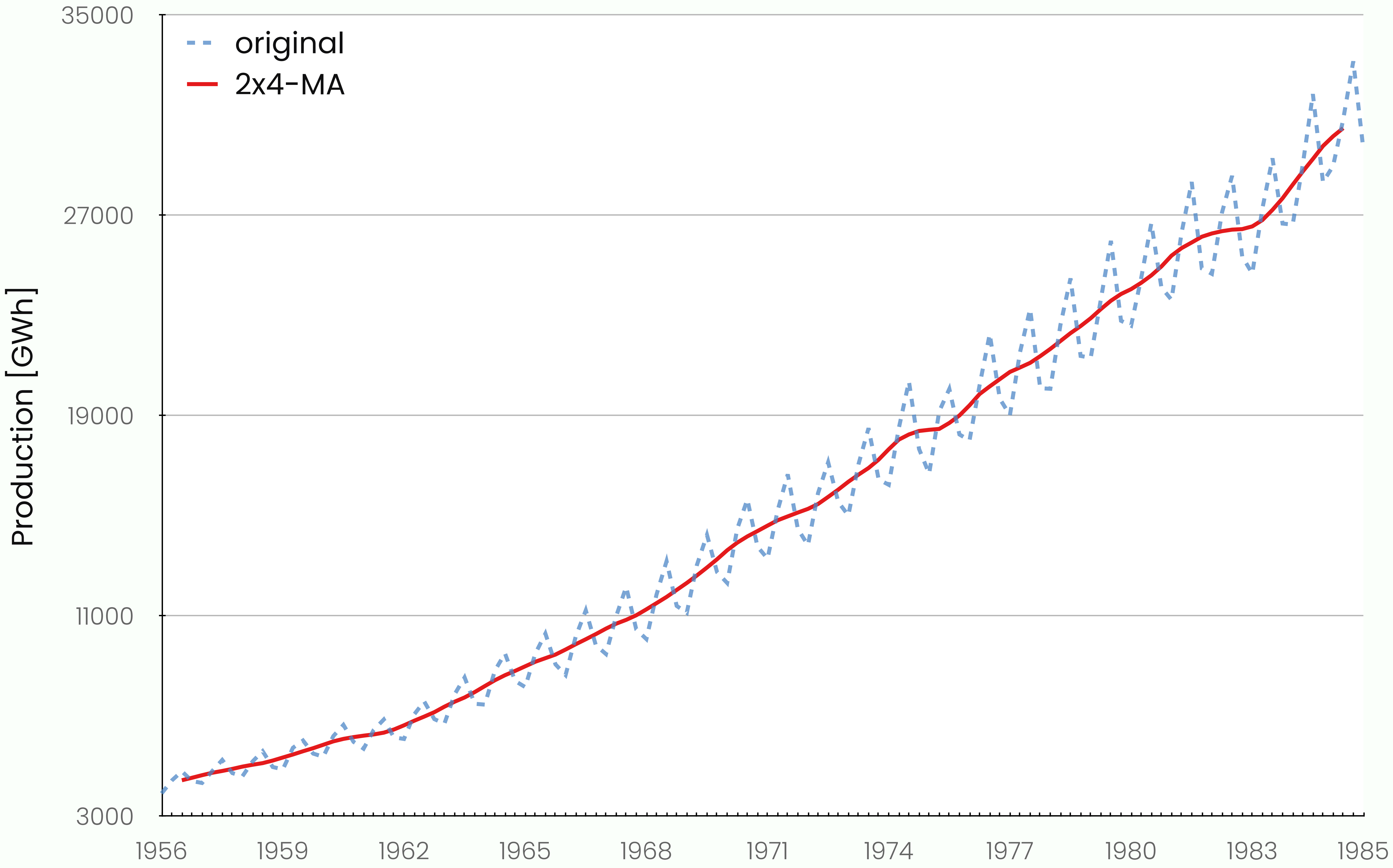
MOVING AVERAGE: EVEN (2)

Year	Production	4-MA	2x4-MA
Q1 1956	3923		
Q2 1956	4436	4396	
Q3 1956	4806	4500	4447.75
Q4 1956	4418	4594	
Q1 1957	4339	4707	
Q2 1957	4811	4786	
...	...	...	



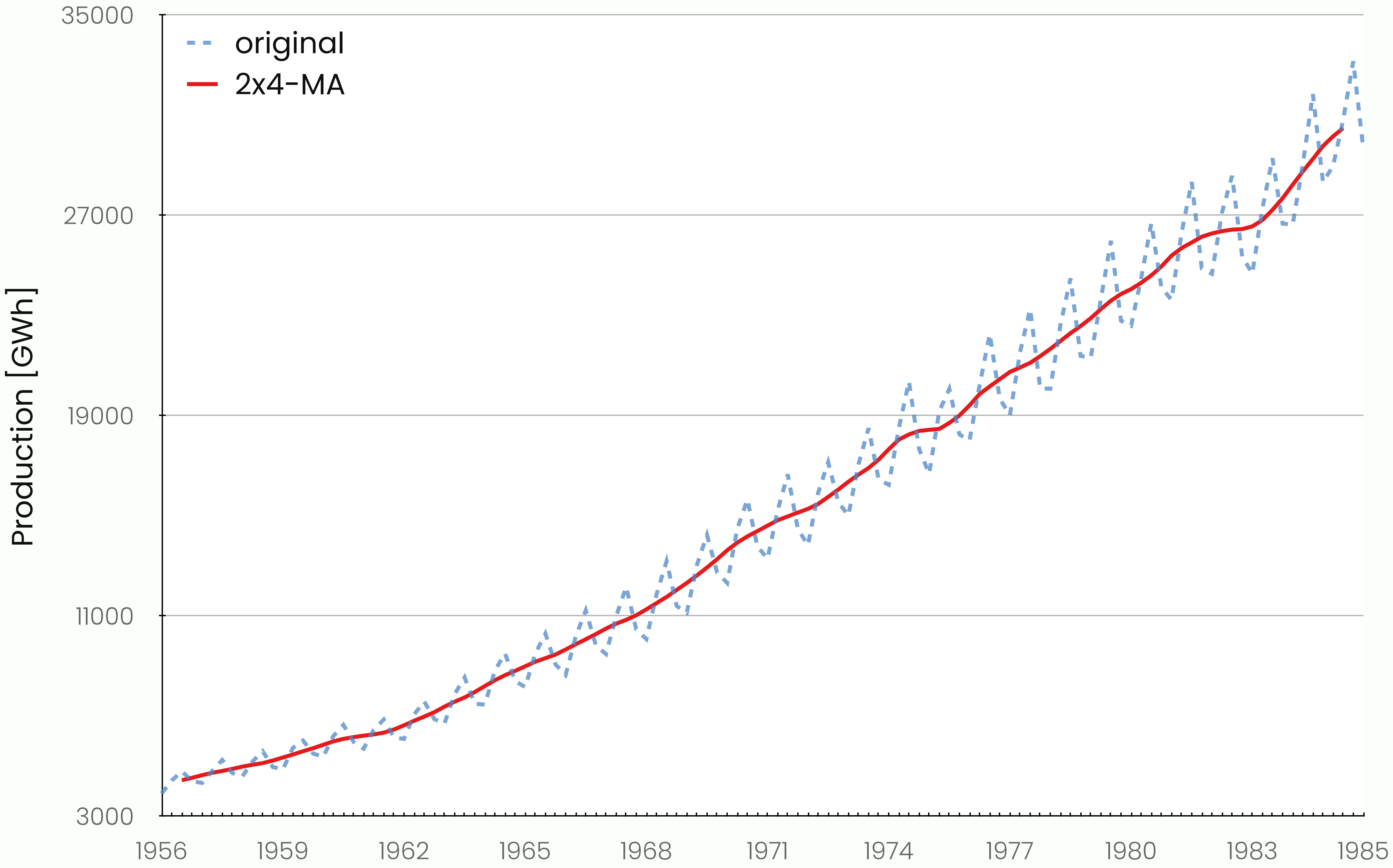
MOVING AVERAGE: EVEN (2)

Year	Production	4-MA	2x4-MA
Q1 1956	3923		
Q2 1956	4436	4396	
Q3 1956	4806	4500	4447.75
Q4 1956	4418	4594	4546.625
Q1 1957	4339	4707	
Q2 1957	4811	4786	
...	...	...	



MOVING AVERAGE: EVEN (2)

Year	Production	4-MA	2x4-MA
Q1 1956	3923		
Q2 1956	4436	4396	
Q3 1956	4806	4500	4447.75
Q4 1956	4418	4594	4546.625
Q1 1957	4339	4707	4650.125
Q2 1957	4811	4786	4746.375
...	...	...	...



ADDITIVE CLASSICAL DECOMPOSITION (ACD)

ORIGINAL SERIES

=

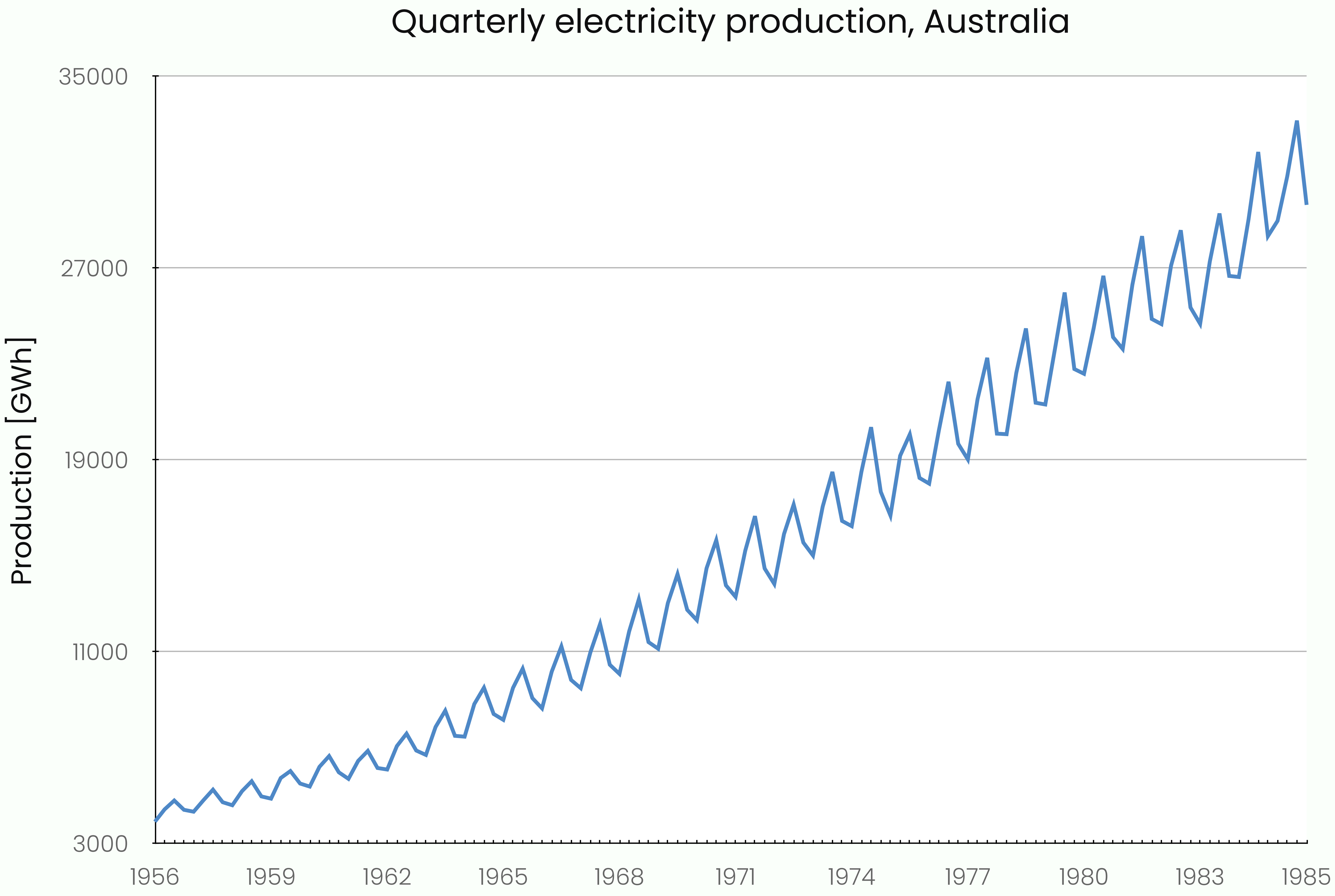
TREND / CYCLE

+

SEASON

+

REMAINDER



ACD: TREND

STEP 1:

If m is an even number, compute the trend-cycle component $\hat{T}_t$ using a $2 \times m$-MA. If m is an odd number, use an m-MA.

ACD: TREND

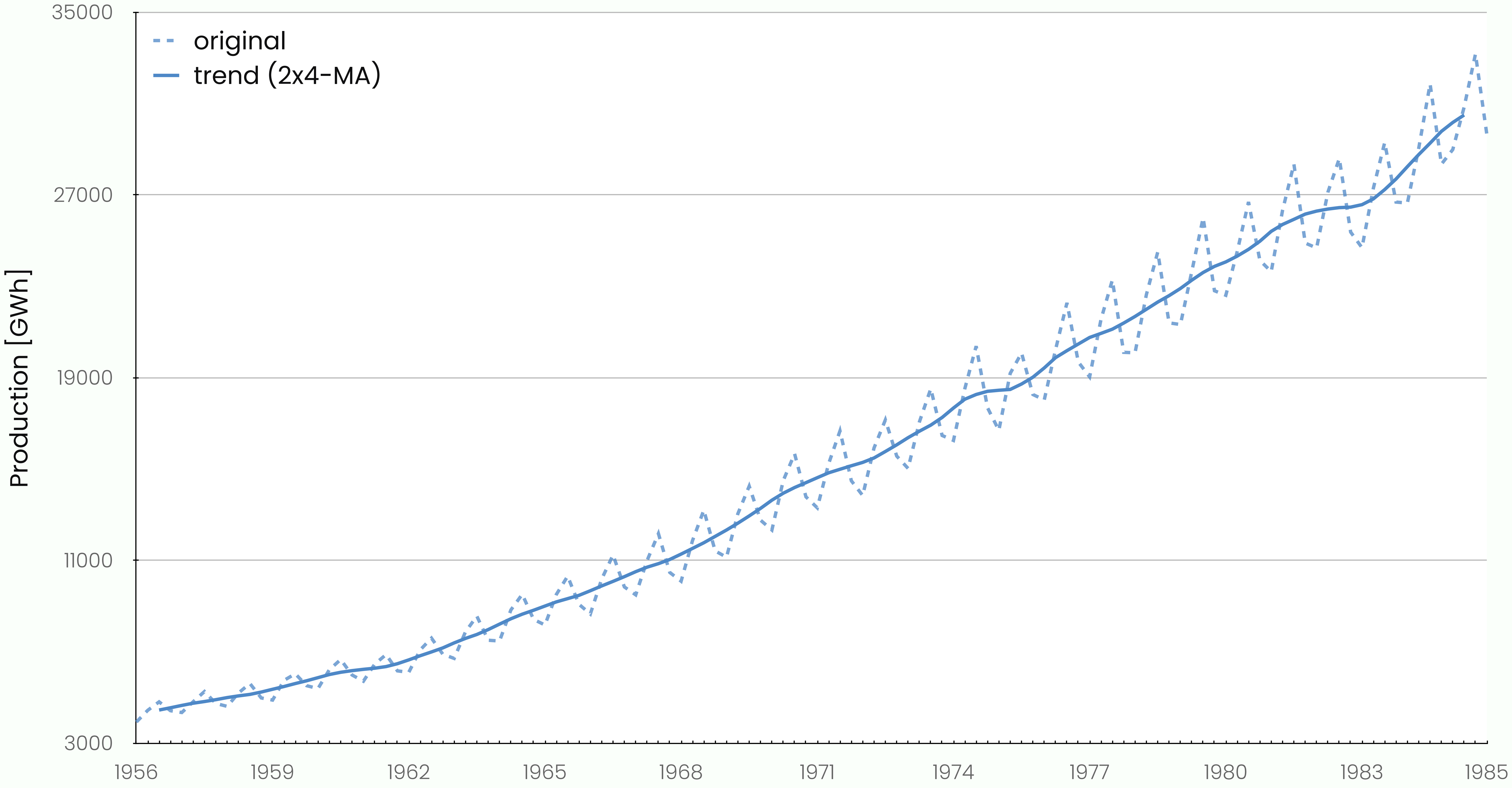
STEP 1:

If m is an even number, compute the trend-cycle component $\hat{T}_t$ using a $2 \times m$-MA. If m is an odd number, use an m-MA.

We have a quarterly data.

Hence, $m = 4$ is even and we apply 2×4-MA.

ACD: TREND (2)



original
trend (2x4-MA)
Production [GWh]
35000
27000
19000
11000
3000
1956
1959
1962
1965
1968
1971
1974
1977
1980
1983
1985

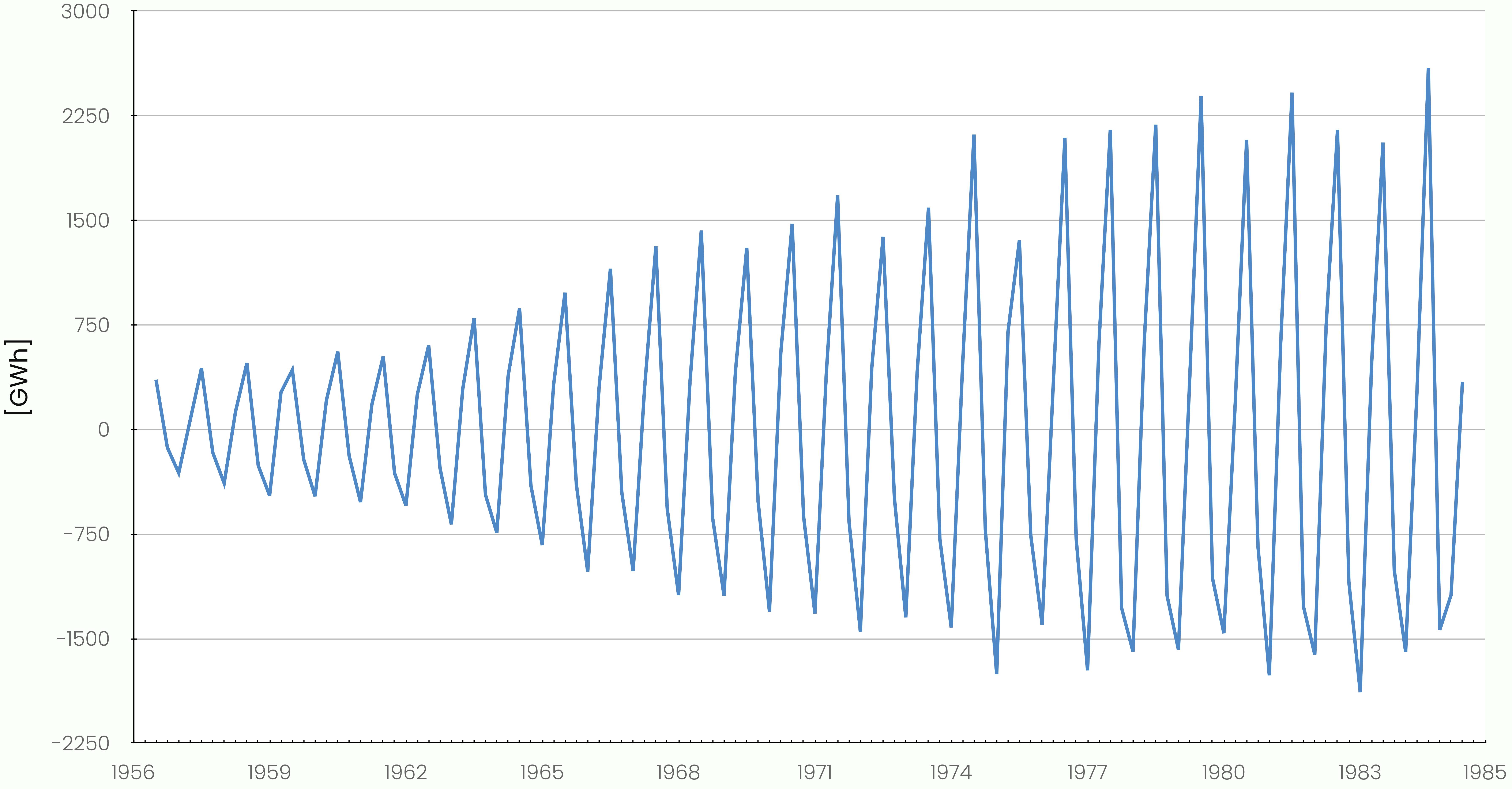
ACD: TREND (3)

STEP 2:

Calculate the de-trended series: $y_t - \hat{T}_t$.

ACD: TREND (4)

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



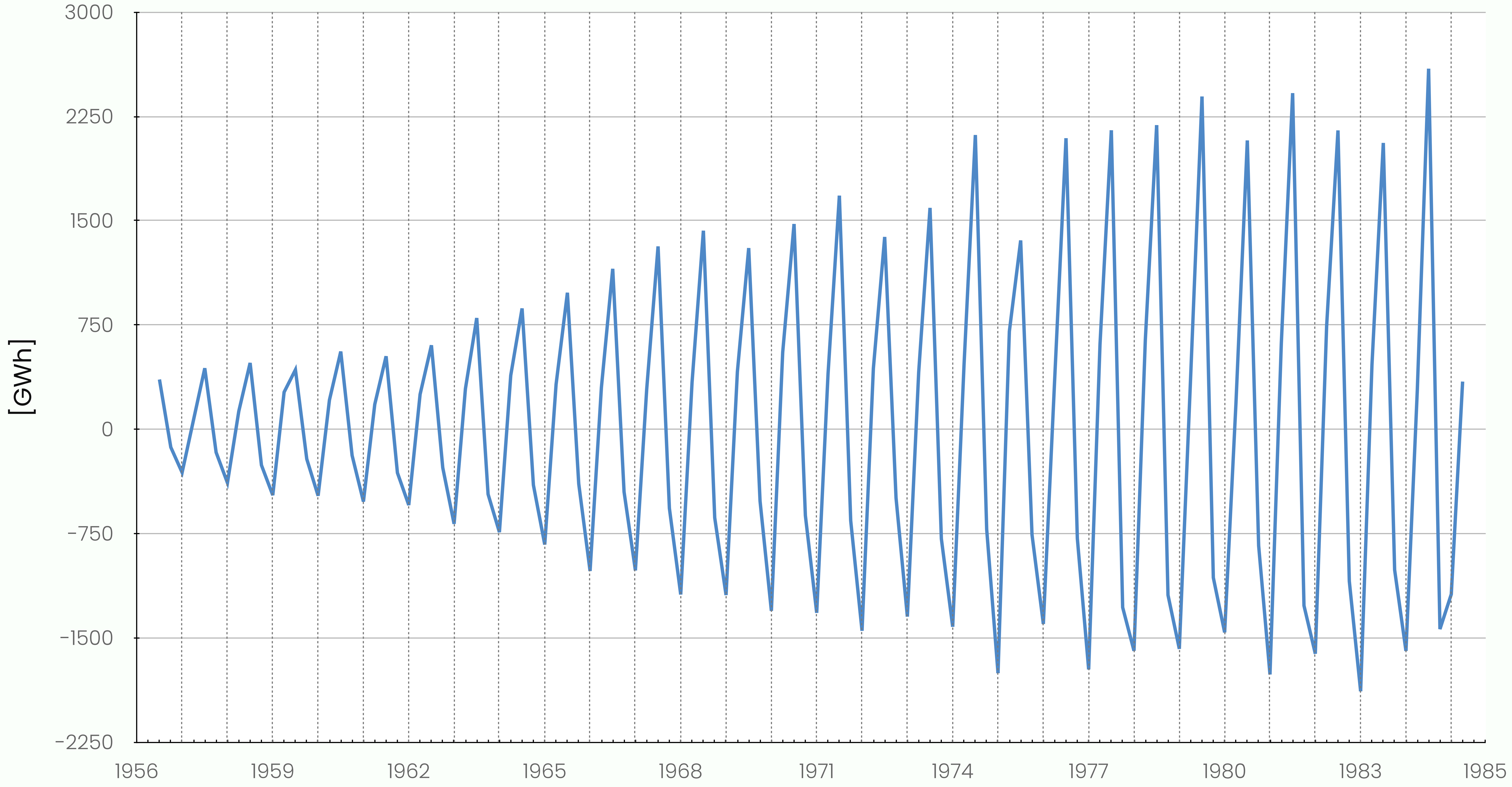
ACD: SEASONAL

STEP 3:

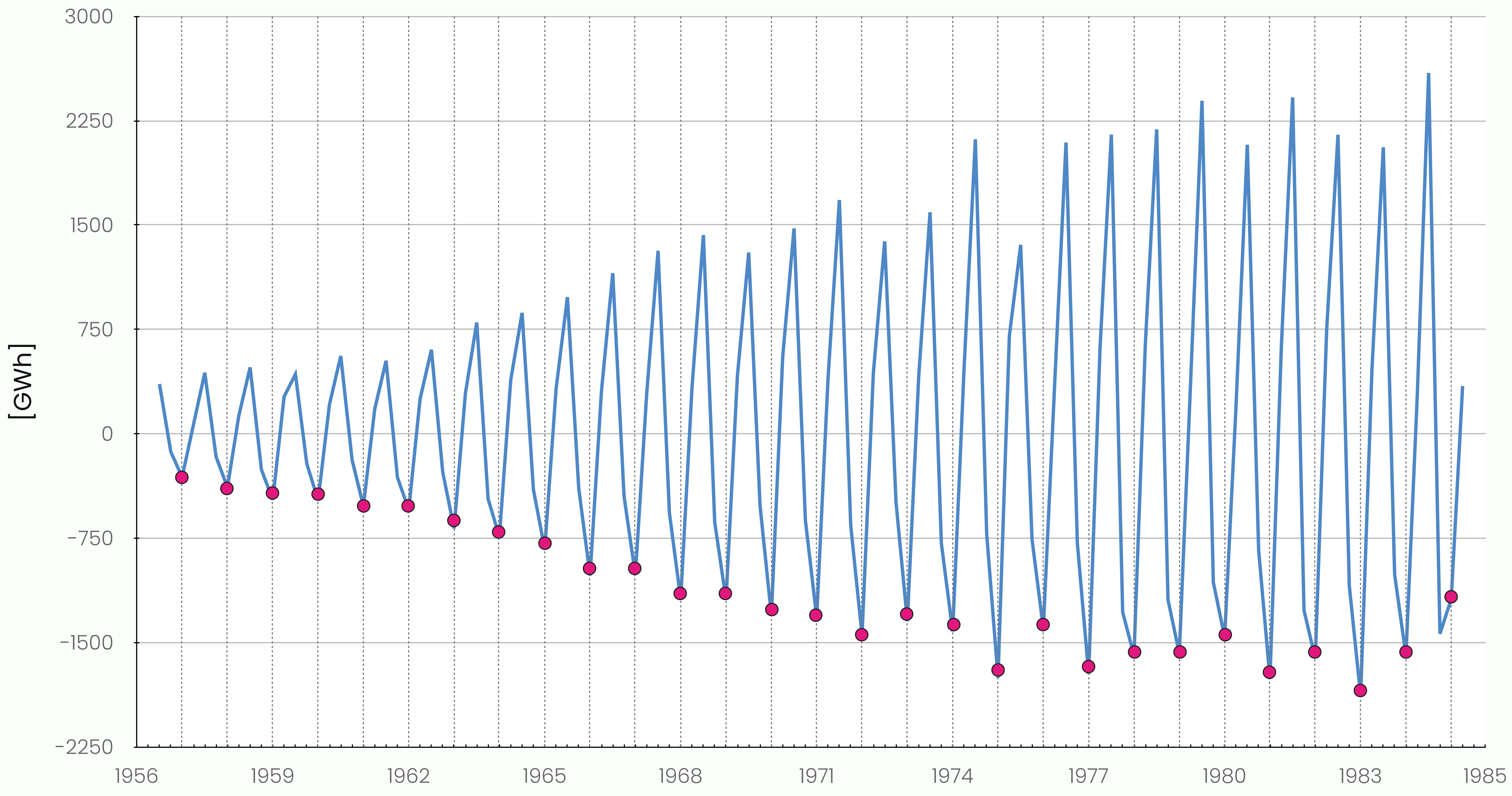
To estimate the seasonal component for each season, average the de-trended values for that season.

These seasonal component values are then adjusted to ensure that they add to zero. The seasonal component is then obtained by stringing together these values, and then replicating the sequence for each time interval. This gives $\hat{S}_t$.

ACD: SEASONAL (2)



ACD: SEASONAL (2)



ACD: SEASONAL (3)

Quarterly average values are:

$Q1 = -1164.5$

$Q2 = 380.28$

$Q3 = 1412.6$

$Q4 = -654.42$

Year	Detrended	
Q1 1956		
Q2 1956		
Q3 1956	358.25	
Q4 1956	-128.63	
Q1 1957	-311.13	
Q2 1957	64.63	
Q3 1957	439.38	
Q4 1957	-166.38	
Q1 1958	-385.25	
Q2 1958	128.75	
...	...	...

ACD: SEASONAL (3)

Quarterly average values are:

$\boxed{Q1 = -1164.5}$

$Q2 = 380.28$

$Q3 = 1412.6$

$Q4 = -654.42$

Year	Detrended	
Q1 1956		
Q2 1956		
Q3 1956	358.25	
Q4 1956	-128.63	
Q1 1957	$\boxed{-311.13}$	
Q2 1957	64.63	
Q3 1957	439.38	
Q4 1957	-166.38	
Q1 1958	$\boxed{-385.25}$	
Q2 1958	128.75	
...	...	...

ACD: SEASONAL (3)

Quarterly average values are:

$Q1 = -1164.5$

$\boxed{Q2 = 380.28}$

$Q3 = 1412.6$

$Q4 = -654.42$

Year	Detrended	
Q1 1956		
Q2 1956		
Q3 1956	358.25	
Q4 1956	-128.63	
Q1 1957	-311.13	
Q2 1957	$\boxed{64.63}$	
Q3 1957	439.38	
Q4 1957	-166.38	
Q1 1958	-385.25	
Q2 1958	$\boxed{128.75}$	
...	...	...

ACD: SEASONAL (3)

Quarterly average values are:

$Q1 = -1164.5$

$Q2 = 380.28$

$\boxed{Q3 = 1412.6}$

$Q4 = -654.42$

Year	Detrended	
Q1 1956		
Q2 1956		
Q3 1956	$\boxed{358.25}$	
Q4 1956	-128.63	
Q1 1957	-311.13	
Q2 1957	64.63	
Q3 1957	$\boxed{439.38}$	
Q4 1957	-166.38	
Q1 1958	-385.25	
Q2 1958	128.75	
...	...	...

ACD: SEASONAL (3)

Quarterly average values are:

$Q1 = -1164.5$

$Q2 = 380.28$

$Q3 = 1412.6$

$\boxed{Q4 = -654.42}$

Year	Detrended	
Q1 1956		
Q2 1956		
Q3 1956	358.25	
Q4 1956	-128.63	
Q1 1957	-311.13	
Q2 1957	64.63	
Q3 1957	439.38	
Q4 1957	-166.38	
Q1 1958	-385.25	
Q2 1958	128.75	
...	...	...

ACD: SEASONAL (3)

Quarterly average values are:

$Q1 = -1164.5$

$Q2 = 380.28$

$Q3 = 1412.6$

$\boxed{Q4 = -654.42}$

Adjust seasonal components by removing the mean:

$$\mu_t = \text{mean}\left(y_t - \hat{T}_t\right) = -6.51.$$

Year	Detrended	
Q1 1956		
Q2 1956		
Q3 1956	358.25	
Q4 1956	$\boxed{-128.63}$	
Q1 1957	-311.13	
Q2 1957	64.63	
Q3 1957	439.38	
Q4 1957	$\boxed{-166.38}$	
Q1 1958	-385.25	
Q2 1958	128.75	
...	...	...

ACD: SEASONAL (3)

Quarterly average values are:

$$Q1 = -1164.5$$
$$Q2 = 380.28$$
$$Q3 = 1412.6$$
$$\boxed{Q4 = -654.42}$$

$$\xrightarrow{-\mu_t}$$

$$Q1 = -1157.98$$
$$Q2 = 386.79$$
$$Q3 = 1419.11$$
$$Q4 = -647.91$$

Adjust seasonal components by removing the mean:

$$\mu_t = \text{mean}\left(y_t - \hat{T}_t\right) = -6.51.$$

Year	Detrended	
Q1 1956		
Q2 1956		
Q3 1956	358.25	
Q4 1956	$\boxed{-128.63}$	
Q1 1957	-311.13	
Q2 1957	64.63	
Q3 1957	439.38	
Q4 1957	$\boxed{-166.38}$	
Q1 1958	-385.25	
Q2 1958	128.75	
...	...	...

ACD: SEASONAL (3)

Quarterly average values are:

$Q1 = -1164.5$

$Q2 = 380.28$

$Q3 = 1412.6$

$\boxed{Q4 = -654.42}$

$\xrightarrow{-\mu_t}$

$Q1 = -1157.98$

$Q2 = 386.79$

$Q3 = 1419.11$

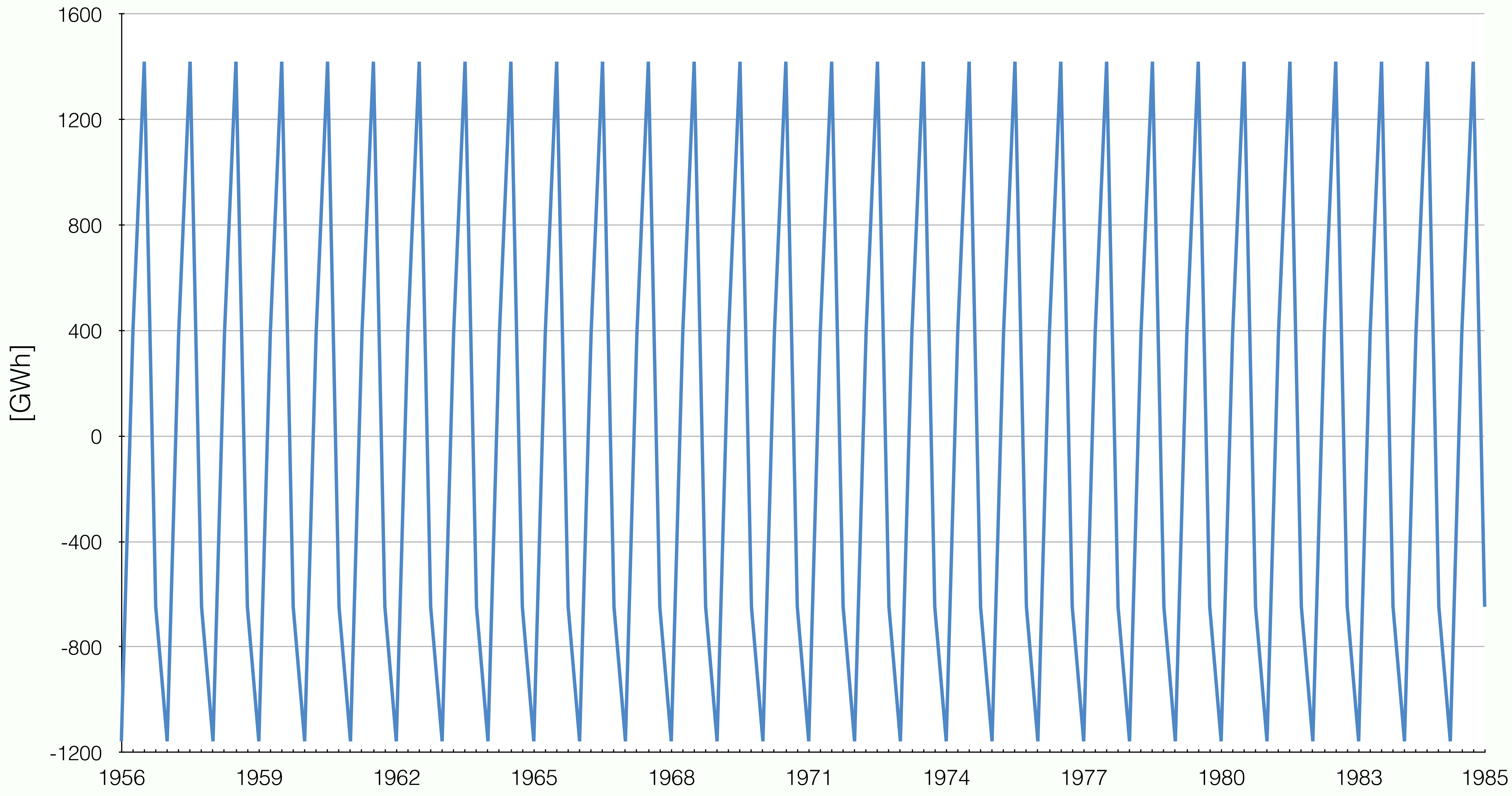
$Q4 = -647.91$

Adjust seasonal components by removing the mean:

$$\mu_t = \text{mean}\left(y_t - \hat{T}_t\right) = -6.51.$$

Year	Detrended	$\hat{S}_t$
Q1 1956		
Q2 1956		
Q3 1956	358.25	1419.11
Q4 1956	-128.63	-647.91
Q1 1957	-311.13	-1157.98
Q2 1957	64.63	386.79
Q3 1957	439.38	1491.11
Q4 1957	-166.38	-647.91
Q1 1958	-385.25	-1157.98
Q2 1958	128.75	386.79
...	...	...

ACD: SEASONAL (4)



ACD: REMAINDER

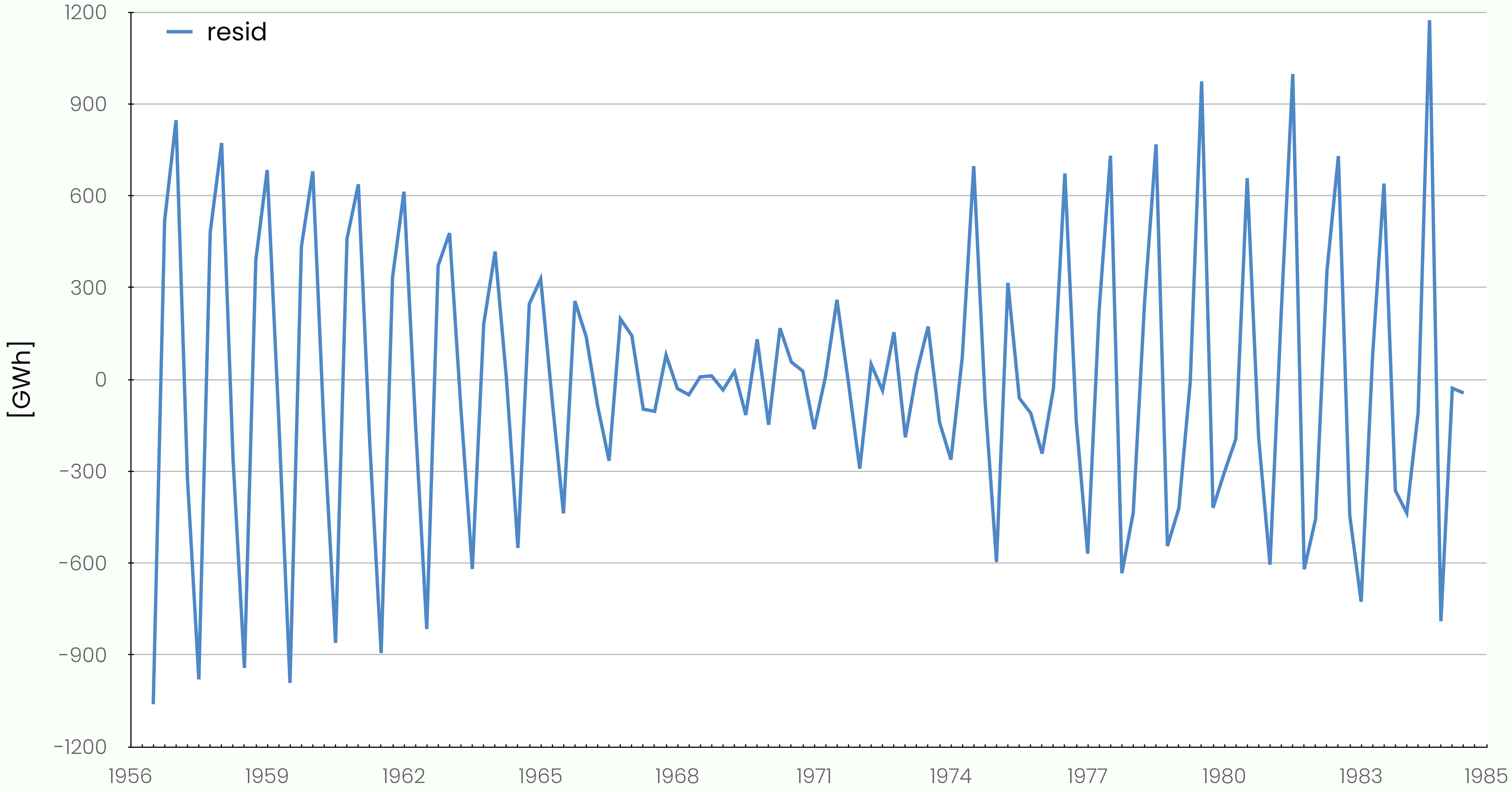
STEP 4:

The remainder component is calculated as:

$$\hat{R}_t = y_t - \hat{T}_t - \hat{S}_t.$$

ACD: REMAINDER (2)

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



HOME ACTIVITIES

- ✳ Experiment with "AU_gas_production_mult_ts.xlsx" data using multiplicative decomposition.
- ✳ For more applications you may read about STL decomposition for anomaly detection.

Thank you!

Questions?