Time Series Forecasting in Energy Field

Jyoti Zeilinger

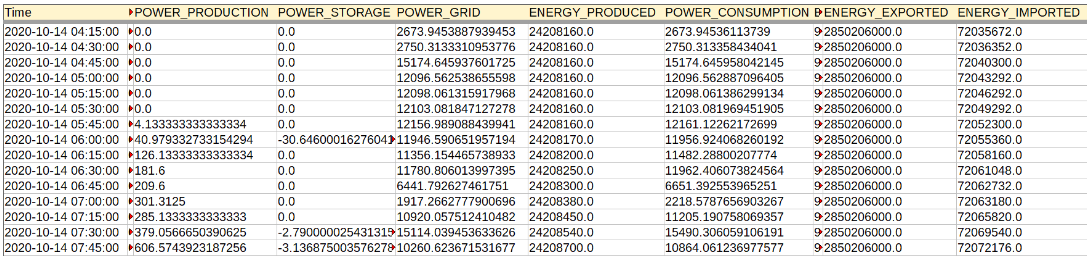
January 17, 2022

1 Introduction

A farmer owns 30 photovoltaic (PV) cells. The PV system is grid-connected, has a battery and a 6.5 kW peak. Given is a time series of the PV system's energy data from two consecutive weeks in October 2020.

2 Data visualization and description

The time series is given in 15-minute intervals, see Figure 1.



Time	POWER_PRODUCTION	POWER_STORAGE	POWER_GRID	ENERGY_PRODUCED	POWER_CONSUMPTION	E	ENERGY_EXPORTED	ENERGY_IMPORTED
2020-10-14 04:15:00	0.0	0.0	2673.9453887939453	24208160.0	2673.94536113739	9	2850206000.0	72035672.0
2020-10-14 04:30:00	0.0	0.0	2750.3133310953776	24208160.0	2750.313358434041	9	2850206000.0	72036352.0
2020-10-14 04:45:00	0.0	0.0	15174.645937601725	24208160.0	15174.645958042145	9	2850206000.0	72040300.0
2020-10-14 05:00:00	0.0	0.0	12096.562538655598	24208160.0	12096.562887096405	9	2850206000.0	72043292.0
2020-10-14 05:15:00	0.0	0.0	12098.061315917968	24208160.0	12098.061386299134	9	2850206000.0	72046292.0
2020-10-14 05:30:00	0.0	0.0	12103.081847127278	24208160.0	12103.081969451905	9	2850206000.0	72049292.0
2020-10-14 05:45:00	4.133333333333334	0.0	12156.989088439941	24208160.0	12161.12262172699	9	2850206000.0	72052300.0
2020-10-14 06:00:00	40.979332733154294	-30.6460001627604	11946.590651957194	24208170.0	11956.924068260192	9	2850206000.0	72055360.0
2020-10-14 06:15:00	126.13333333333334	0.0	11356.154465738933	24208200.0	11482.28800207774	9	2850206000.0	72058160.0
2020-10-14 06:30:00	181.6	0.0	11780.806013997395	24208250.0	11962.406073824564	9	2850206000.0	72061048.0
2020-10-14 06:45:00	209.6	0.0	6441.792627461751	24208300.0	6651.392553965251	9	2850206000.0	72062732.0
2020-10-14 07:00:00	301.3125	0.0	1917.2662777900696	24208380.0	2218.5787656903267	9	2850206000.0	72063180.0
2020-10-14 07:15:00	285.1333333333333	0.0	10920.057512410482	24208450.0	11205.190758069357	9	2850206000.0	72065820.0
2020-10-14 07:30:00	379.0566650390625	-2.790000025431319	15114.039453633626	24208540.0	15490.306059106191	9	2850206000.0	72069540.0
2020-10-14 07:45:00	606.5743923187256	-3.136875003576278	10260.623671531677	24208700.0	10864.061236977577	9	2850206000.0	72072176.0

Figure 1: Time series data set of PV system

The data is given as power data (in W) and energy data (in Wh). The customer's PV system has three main devices: the PV cells, a battery and the common grid. The PV cells use sunlight to produce power, given by POWER_PRODUCTION. Some of the power may be stored in the battery, given by POWER_STORAGE. When needed, power is drawn from the grid, given by POWER_GRID. The PV system produces energy, given by ENERGY_PRODUCED. The customer's total power consumption is given by POWER_CONSUMPTION. The PV system may produce more energy than its owner uses. In this case, the excess energy is exported into the grid, given by ENERGY_EXPORTED. When the PV cells do not produce enough energy, additional energy is imported from the grid, given by ENERGY_IMPORTED.

I now visualize the relationship between the PV system's power production and its energy production, first for one day and then for the whole time series. Figure 2 shows data on one day, namely on 20th October 2020. The upper plot within Figure 2 shows the power production; the lower plot the resulting energy produced. The power production starts before 06:00 in the morning (around sunrise) and continues until just after 15:00 in the afternoon (towards sunset). Outside of this time period, there is no power production. Note that the PV system's power peak of 6.5 kW is reached on this day. The energy production increases precisely during the time where power production takes place. After the 20th October 2020, the energy produced has increased from below 24280 kWh to more than 24310 kWh.

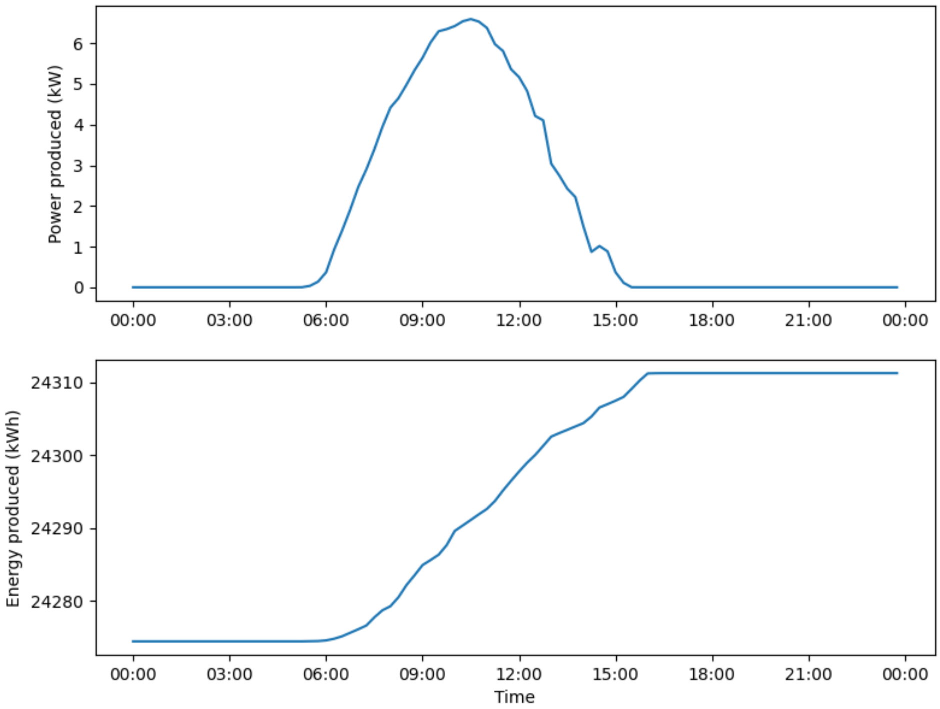


Figure 2: Photovoltaic production on 20th Oct 2020

Figure 3 shows data for the whole two October weeks. Each day, power production starts in the morning and ends in the afternoon. During these time periods, energy is produced. The PV system produces varying amounts of energy per day. Over the course of two weeks, the PV system's power peak of 6.5 kW is reached only once (on 20th October). At the end of the two weeks, the PV system has produced around 200 kWh of total energy.

When the PV system's power production is too low, energy must be taken from the grid (or battery). On the other hand, when the PV system's production is too high, energy can be exported. Next I investigate how much of the total consumed power is covered by the grid, PV system or battery. Firstly, I calculate how much percentage of total consumed power comes from the grid, how much is produced locally by the PV cells and how much is stored locally in the battery, see the Python code snippet in Listing 1.

```python
# calculate how much of total power consumption from which power device
df['POWER_RATIO_GRID'] = df['POWER_GRID'] / df['POWER_CONSUMPTION'] * 100
df['POWER_RATIO_PRODUCTION'] = df['POWER_PRODUCTION'] / df['POWER_CONSUMPTION'] * 100
df['POWER_RATIO_STORAGE'] = df['POWER_STORAGE'] / df['POWER_CONSUMPTION'] * 100
```

Listing 1: Calculate how much power from which power source/device

Figure 4 shows how and when the total power consumption is distributed amongst the PV system's devices (for the first eight days of the data set).

The upper plot of Figure 4 shows the percentages of total consumed power supplied by or stored in the different devices. The lower plot shows energy which is exported. In general: the blue line (grid power) is often at 100%. During these times, the grid supplies **all** of the consumed power. During nighttime, 100% of the power consumption is covered by the grid

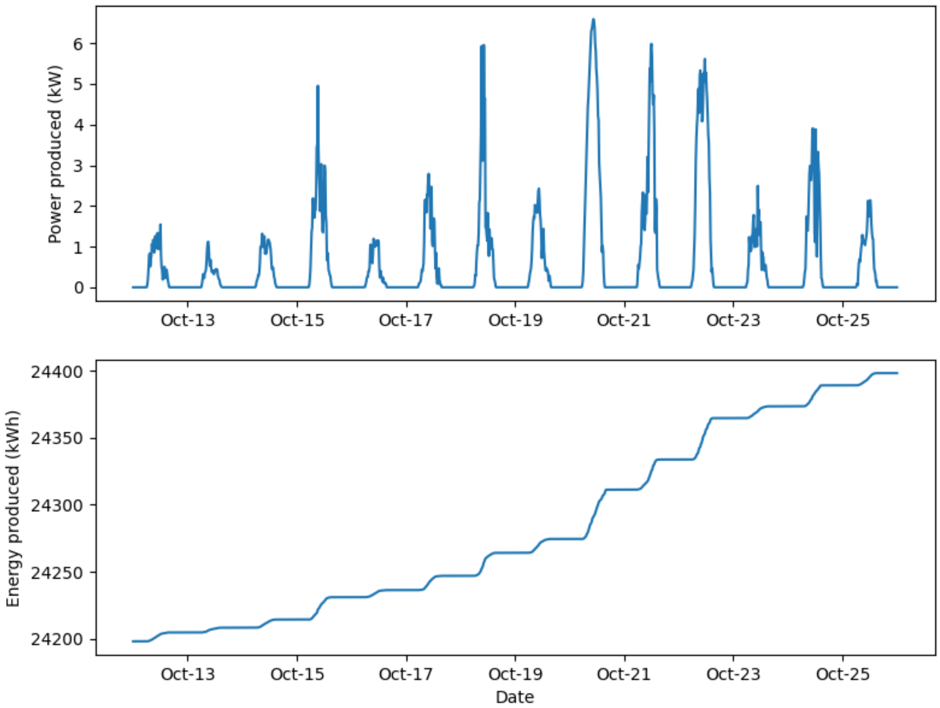


Figure 3: Photovoltaic production during two weeks in Oct 2020

(blue), for sure. During daytime, some of the power consumption is covered locally by the PV cells (orange). More specifically: During daytimes of Oct-12, Oct-13, Oct-14 and Oct-16, upto around 50% of the power consumption can be covered locally. On Oct-15, the PV system produces more power than needed. Some of the power can temporarily be stored in the battery (green). But energy is also exported, corresponding to an increase of exported energy in the lower plot. On Oct-18, there is a high PV power overproduction, leading to a high increase of exported energy in the lower plot. On Oct-19 (and Oct-17), all PV power overproduction can be stored in the battery, i.e. no energy needs to be exported (lower plot stays constant).

3 Forecast PV production

Next, I predict the PV system's energy production two days into the future. To do so, I use a classical statistical method, namely the additive Holt-Winters model (from Python's statsmodels library). The Holt-Winters model uses triple exponential smoothing. Furthermore, the (additive) Holt-Winters model assumes a linear trend and a constant seasonal cycle.

Here, the period of a season is one day. Since we have quarter-hourly data, the seasonal period is 96 (hours per day times minutes per day divided by minutes per quarter hour).

First, I do an in-data forecast from the first 12 days (train data set) to the last two days (test data set). I compare the prediction results to the known data points (present in the data set). Then I do an actual out-data forecast into the unknown future (not present in data set).

Figure 5 shows the model fit and in-data forecast from the train data set to the test data set. The model fit is very good (orange and purple lines match). But the forecast is not

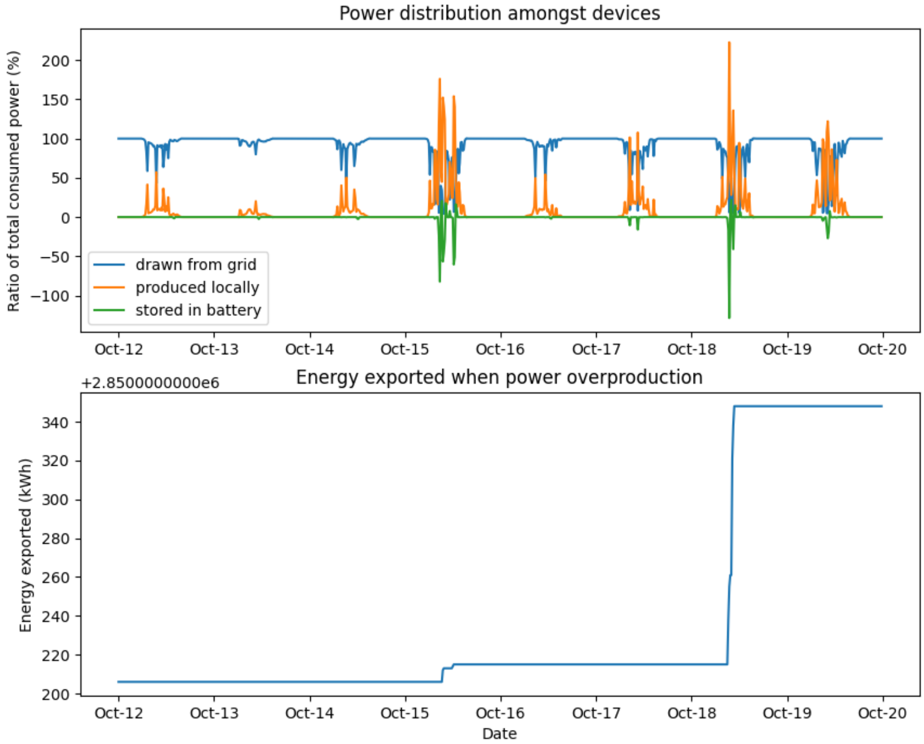


Figure 4: Power overproduction leads to energy export

very satisfactory. The data set has a pattern of being constant most of the time with abrupt increases. This behaviour is not reflected well in the forecast[1]. Rather, the forecast shows a more slow, gradual increase. The prediction level at the end of the first predicted day is met well. For the second day, the prediction level is too optimistic.

Figure 6 also shows the out-data forecast from the whole data set to two days into the unknown future. Compared to the known jumps (energy production increases) within the data set, the predicted jumps are rather small. Therefore the forecast indicates a small amount of energy production.

In general, I do not expect the best results from this simple Holt-Winters forecast. There is only one input data (time) and only one target (energy produced). For a better forecast, I would use a (higher dimensional) model with exogenous variables: weather and solar data (especially solar flux and sunrise/set data).

[1]Note that there is a warning message for both in-data and out-data forecasts. I suspect that the flat parts of the data are causing the issue behind the warning.

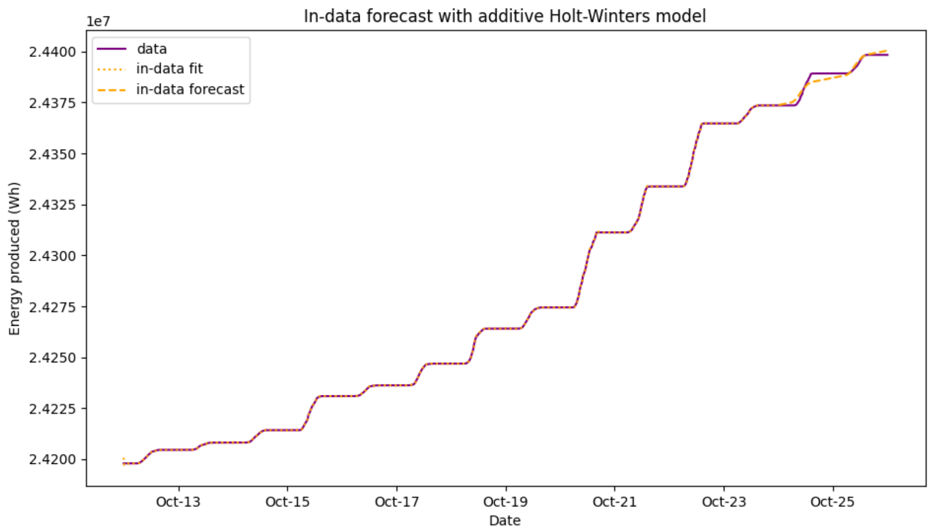


Figure 5: Model forecasts energy levels well but does not catch flat behaviour

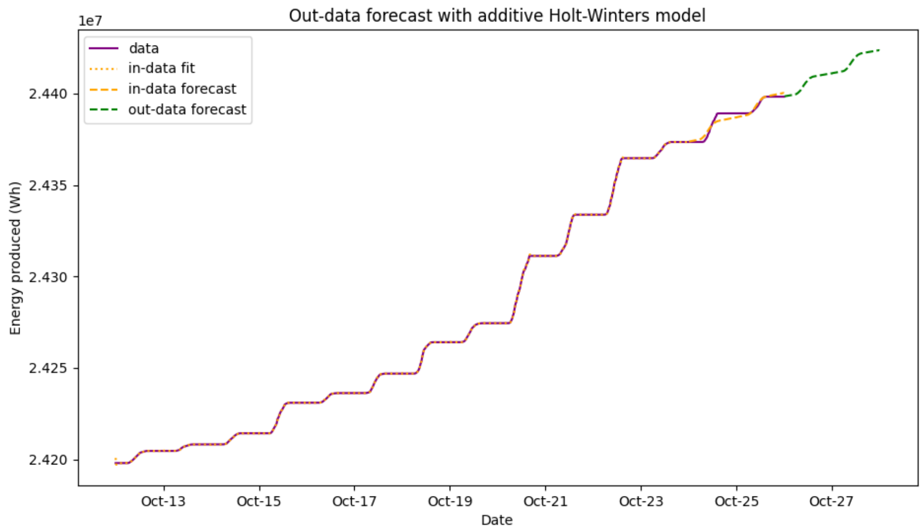


Figure 6: Produced energy for next two future days is forecast to be comparatively low