# Estimating unmetered photovoltaic power consumption using causal models

**Keywords:** *ARIMAX, energy accounts, solar energy, time series*

## 1 Introduction

Energy accounting encompasses the compilation of coherent statistics on energy related issues in countries, including the production and consumption of electricity. A complete picture of demand and supply of electricity must include data on electricity production outside the energy industries, such as electricity produced by domestic photovoltaic (PV) installations. These PV installations are rarely metered by distribution net operators, hence, their production remains invisible to statistical agencies responsible for the energy accounts. Consequently, the renewable electricity production is difficult to estimate while monitoring it is crucially important for climate policy evaluation.

In the Netherlands—the country studied in this article—an incomplete register of PV installations is available. Such registers can be used to estimate power produced by PV installations using a modelling approach relating installed capacity to produced electricity. In the present article we propose inferring solar power production from causal relations between solar irradiance and consumption of grid power. Since the production of solar power by domestic PV installations results in a reduced consumption of electricity from the high-voltage grid the combination of time series of electricity exchange on the high-power grid and series of solar irradiance contain a hidden signal of unmetered solar power produced by domestic PV installations. In this paper a causal model for these time series to estimate unmetered solar power production is developed.

## 2 Methods

When PV installations produce a lot of electricity, consumers need less electricity from the high voltage grid. The total electricity use in the country is catered for partly by small PV installations, and complemented with grid power. For a given—but unknown—installed PV capacity, the solar power is proportional to the solar irradiance. In this paper the amount of produced domestic solar power is estimated by combining time series of electricity exchange from the high power grid in combination with time series on meteorological data, in particular solar irradiance. Data on electricity exchange on the high power grid in MWh at a daily frequency have been used covering the period from Jan 1st, 2004 through Dec 31st, 2017, which are freely available from the website of the Dutch Transmission System Operator (Tennet). Data on solar irradiance (in J/cm$^2$) and the temperature (in 0.1$^o$C) at a daily level as well as day length were obtained from the Royal Netherlands Meteorological Institute for the same period.

For solar energy, the most important exogenous variable is the solar irradiance, $I_{\odot t}$. The subscript $t$ indicates that the accumulated total for time period $t$ in days and $\odot$ is the symbol for the Sun. Interest is in the effect of solar irradiance on public grid demand $Y_t$. The effect of $I_{\odot t}$ depends on how much solar power ($P_{\odot t}$) is generated. Public grid demand is not only determined by solar power, it mainly dependent on the

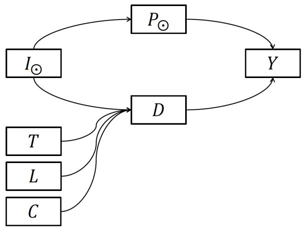


Figure 1: Directed acyclic graph (DAG) for the solar power causal model, with $I_{\odot t}$ solar irradiance, $P_{\odot t}$ solar power, $Y$ grid power, $D$ total demand, $T$ temperature, $L$ length of day and $C$ calender effects.

total electricity demand, $D_t$ and factors like solar irradiance $I_{\odot t}$ weather conditions like the average temperature $T_t$, length of day $L_t$ and calendar effects $C_t$. Figure 1 shows the causal relationships among these variables by means of a directed acyclic graph (DAG), see (Pearl, 1995). Interest is in estimation of $P_{\odot t}$ based on observations of $I_{\odot t}$ and $Y_t$. From the DAG (Figure 1) it is clearly seen that there are two causal paths between $I_{\odot t}$ and $Y_t$,

$$I_{\odot t} \rightarrow P_{\odot t} \rightarrow Y_t \tag{1}$$
$$I_{\odot t} \rightarrow D_t \rightarrow Y_t. \tag{2}$$

The role played by $I_{\odot t}$ renders $P_{\odot t}$ and $D_t$ only conditionally independent; $P_{\odot t} \perp D_t | I_{\odot t}$, which complicates estimation of the direct effects of $I_{\odot t}$ on $Y_t$, in particular the effect of $P_{\odot t}$ on $Y_t$. In order to obtain this desired estimate, the causal path (2) must be closed. This is achieved by conditioning on—or adjusting for—$D_t$. This can be conducted using data where $P_{\odot t} = 0$.

In the Netherlands there was no significant presence of domestic PV installations prior to 2010. PV installations were introduced gradually in 2011 and 2012 and started to become more widespread from 2013 onwards. We use the aforementioned observational time series data and construct subset $A$ containing the data from the period 2004—2010 to estimate relation (2), where it can be assumed that $P_{\odot t} = 0$. Subsequently we use subset $B$ for the period 2013—2017 to estimate relation (1) where we control for the effect $I_{\odot t} \rightarrow D_t \rightarrow Y_t$ from the previous analysis.

The time series on electricity exchange $Y_t$ for data set $A$ and $B$ are modelled with autoregressive integrated moving average (ARIMA) models (Box et al., 2015). Relations with $I_{\odot t}$, $P_{\odot t}$, $T_t$, $L_t$ and $C_t$ are included by adding additional covariates terms to the ARIMA model and are known as ARIMAX models. First, an ARIMAX model is fitted to data set $A$, in which no solar panels are present, hence it is known that in this case $P_{\odot t} = 0$. From this model we establish the effect of $I_{\odot t}$ on $D$ through the regression of $Y_t$ on $I_{\odot t}$. The corresponding regression coefficient $\hat{\beta}_I^{[A]}$ can be interpreted as the effect of solar irradiance on the total electricity demand $Y_t$. This relation is used to correct $Y_t$ in data set $B$ for the effect of $I_{\odot t}$ on $D$, i.e. $\tilde{Y}_t = Y_t - \hat{\beta}_I^{[A]} I_{\odot t}$. Then an ARIMAX model is fitted again to the corrected data $\tilde{Y}_t$ in set $B$ and the regression of $\tilde{Y}_t$ on $I_{\odot t}$ is used to derive the production of solar power. Let $\hat{\beta}_{I,y}^{[B]}$ denote the regression coefficients from the regression of $\tilde{Y}_t$ on $I_{\odot t}$ in data set $B$. The subscript $y$ stand for year and denote that separate regression coefficients $\hat{\beta}_{I,y}^{[B]}$ are assumed for the years $y$ = 2013, ..., 2017. Solar power estimates on a daily basis are obtained by $\hat{P}_{\odot t} = \hat{\beta}_{I,y}^{[B]} I_{\odot t}$. Annual estimates are finally obtained by aggregating over the days within a year.

Table 1: Results of the ARIMAX model fit on data set $B$.

| Year | $\hat{\beta}^{[B]}_{I,y}$ | SE | $\hat{P}_{\odot t}$ (MWh) | $\hat{D}$ (MWh) | Percentage solar |
|---|---|---|---|---|---|
| 2013 | -0.390 | 0.787 | 140,877 | 101,554,484 | 0.14% |
| 2014 | -1.296 | 0.797 | 485,381 | 99,549,220 | 0.49% |
| 2015 | -2.004 | 0.755 | 774,212 | 100,436,422 | 0.77% |
| 2016 | -3.409 | 0.828 | 1,275,643 | 102,065,655 | 1.25% |
| 2017 | -5.086 | 0.807 | 1,867,628 | 103,223,204 | 1.81% |

## 3 Results

The order for differencing the series $d$, the required number of AR lags $p$ and MA lags $q$ are chosen by minimising the Akaike Information Criterion. It turns out that first order differencing ($d = 1$) is required to render the $Y$ series stationary. The AR order was determined to be $p = 6$ and the MA order $q = 1$. As covariates $I_{\odot t}$, $T_t$, $L_t$ and $C_t$ are included in the ARIMAX model. We found seasonal components to be insignificant and conjecture the reason for this to be that calendar effects are present in the model, essentially fulfilling a role similar to seasonal components.

Daily solar power is estimated following the method described in Section 2 and is shown in Figure 2. There is a clearly increase over time due to the increasing number of PV installations in the country, collectively producing more and more power. The estimated regression coefficients and their standard errors are given in Table 1. In addition, the table contains the estimated annual solar power $P_{\odot t}$, the estimated total electricity demand $\hat{D}$ which is the sum of total grid demand and solar power, and in the last column the solar power as a percentage of the total. In the years 2013 and 2014 the regression coefficient is not significantly different from zero. However, considering the five years together a gradual uptake of solar power is clearly visible. In 2017, unmetered solar power is estimated to account for just under 2 percent of total electricity use.

To evaluate how well the ARIMAX model fit the series several diagnostic checks are applied to the standardized residuals to test the assumptions that these residuals are identically and normally distributed. In addition the estimate of solar power production based on the ARIMAX model are compared with the official figures of Statistics Netherlands. Figure 3 shows estimated unmetered solar power consumption (our model, solid line) and estimates by CBS of total solar power consumption (dashed line) and of the consumption of power from domestic PV installations (dashed line). The latter is derived from an incomplete register of PV installations and assumptions about the power production of these installations. The trends of the CBS total and our model are remarkably well aligned showing predominantly a level difference. Since the estimated total in the official statistics of CBS includes metered solar power as well—from large solar power farms—it is expected that the unmetered component is lower than the total. The unmetered solar power is often thought to be supplied mainly by PV installations at private homes. In the years 2013, 2014 and 2015 the official estimate of domestic PV consumption is close to our model estimates, but the two diverge in 2016 and 2017. No data are publicly available on the share of metered versus unmetered solar power in the total CBS estimates, nor in the domestic versus business consumption. If all estimates shown in Figure 3 are correct, one may conclude that businesses must have installed smaller—hence unmetered—PV installations in the most recent years, 2016 and 2017, before which it was domestic solar power consumption that accounted for most unmetered solar power.

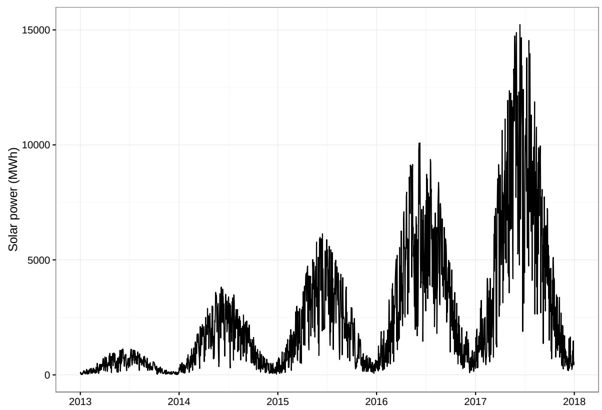


Figure 2: Estimated solar power for the years 2013—2017 in MWh.

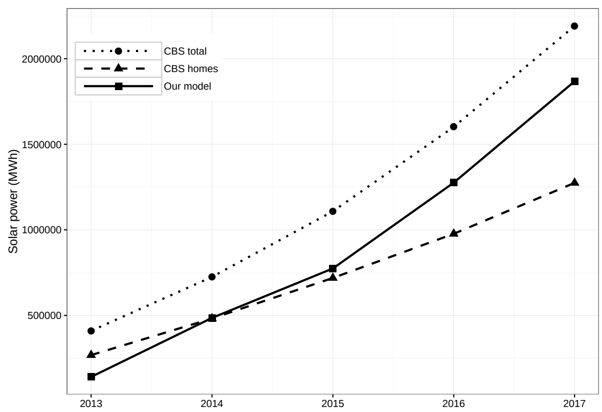


Figure 3: Comparison of our model results (solid line) with official statistics published by CBS on total solar energy consumption (dotted line) and the amount consumed by households (dashed line).

## 4 Conclusions

Reliable statistical information on the use of renewable energy is relevant in order to monitor the implementation of sustainable development. To this end, a time series model is proposed to estimate the unmetered solar power production as a hidden signal in time series of exchange of electricity on the high voltage grid and meteorological time series on solar irradiance, temperature and day length. All time series data used in this analysis are freely available from the internet.

We conclude that our model estimates are not in disagreement with official statistics on solar power consumption. Estimates have been produced on unmetered solar power production and consumption. While official statistics are at annual level, our modelling approach produced daily estimates. In contrast with the regular official statistics, no administrative or survey data on PV installations in the country was required. Hence, the proposed model can be applied easily, quickly and widely, and could be particularly useful in countries where no good estimates of unmetered PV electricity are available yet.

## References


Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.

George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.