

02417 Times Series Analysis - Assignment 3

Authors:

- Jeppe Klitgaard <s250250@dtu.dk>
- Yunis Wirkus <s250700@dtu.dk>

Date: 2025-04-21

1 Simulations

We are given an AR(2) process $\{X_t\}$ with residual term ε_t arising from a stochastic process $\{\varepsilon_t\} \in \mathcal{N}(0, \sigma_\varepsilon^2 = 1)$:

$$X_t + \phi_1 X_{t-1} + \phi_2 X_{t-2} = \varepsilon_t \quad (1)$$

Our implementation will be based on the `statsmodels` library in Python, which offers an implementation of the SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors) model.

Importantly, the SARIMAX model follows an alternative formulation of the model:

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varepsilon_t \quad (2)$$

Where we have used the two different variants of the letter phi to denote the different formulations.

Comparing Eq. 1 and Eq. 2 we find that the two formulations are equivalent given a sign change:

$$\varphi_1 = -\phi_1, \quad \varphi_2 = -\phi_2 \quad (3)$$

1.1

We simulate 5 realisations of the process up to $n = 200$ observations using a burn-in period of $N_B = 10000$ with parameters $\phi_1 = -0.6$ and $\phi_2 = 0.5$ and plot the results in Figure 1.

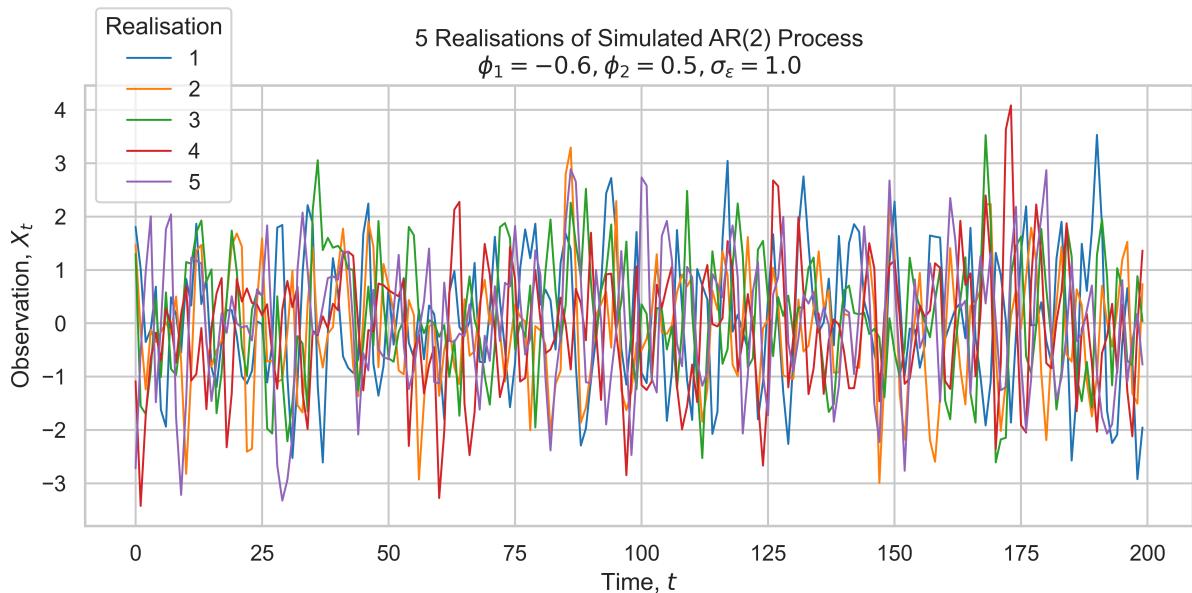


Figure 1: Simulation and Autocorrelation function of AR(2) process with $\phi_1 = -0.6, \phi_2 = 0.5$.

We find that the results look as expected.

1.2

We recall the definition of *autocorrelation* for a *stationary process*:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} \quad (4)$$

Where $\gamma(k)$ is the *autocovariance* for a timeshift k :

$$\gamma(k) = \text{Cov}[X_t, X_{t+k}] \quad (5)$$

We consider the autocorrelation of an AR(2) process by solving for X_t in Eq. 1:

$$X_t = -\phi_1 X_{t-1} - \phi_2 X_{t-2} + \varepsilon_t \quad (6)$$

And then inserting this in Eq. 4:

$$\begin{aligned}
\rho(k) &= \frac{\text{Cov}[X_t, X_{t+k}]}{\gamma(0)} = \frac{\text{Cov}[X_t, -\phi_1 X_{t-1+k} - \phi_2 X_{t-2+k} + \varepsilon_t]}{\gamma(0)} \\
&= \frac{\text{Cov}[X_t, -\phi_1 X_{t-1+k}]}{\gamma(0)} + \frac{\text{Cov}[X_t, -\phi_2 X_{t-2+k}]}{\gamma(0)} + \frac{\text{Cov}[X_t, \varepsilon_t]}{\cancel{\gamma(0)}} \\
&= -\phi_1 \frac{\text{Cov}[X_t, X_{t-1+k}]}{\gamma(0)} - \phi_2 \frac{\text{Cov}[X_t, X_{t-2+k}]}{\gamma(0)} \\
&= -\phi_1 \frac{\gamma(k-1)}{\gamma(0)} - \phi_2 \frac{\gamma(k-2)}{\gamma(0)} \\
\rho(k) &= -\phi_1 \rho(k-1) - \phi_2 \rho(k-2)
\end{aligned} \tag{7}$$

Notably, by stationary it follows $\rho(-k) = \rho(k)$ and from Eq. 4 we find $\rho(0) = 1$, which allows us to build a recursive relation for $\rho(k)$ from:

$$\begin{aligned}
\rho(0) &= 1 \\
\rho(1) &= -\phi_1 \rho(0) - \phi_2 \rho(-1) = -\phi_1 - \phi_2 \rho(1) = \frac{-\phi_1}{1 + \phi_2}
\end{aligned} \tag{8}$$

We can compute the empirical autocorrelation function (ACF) using the `plot_acf` function from `statsmodels`, which we compute for each of the realisations shown in Figure 1.

This is plotted in Figure 2 along with the theoretical autocorrelation function, $\rho(k)$, as given by Eq. 8.

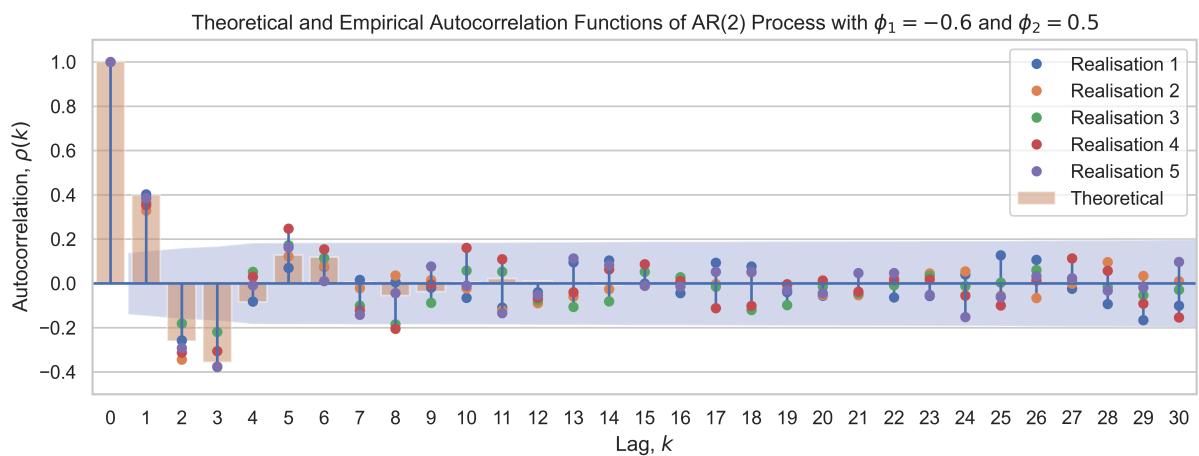


Figure 2: Theoretical and empirical autocorrelation functions of an AR(2) process with $\phi_1 = -0.6, \phi_2 = 0.5$. Shaded blue region is the 95% confidence interval for the empirical ACF.

We observe good agreement between the theoretical and empirical autocorrelation functions. The deviations observed after $k = 4$ are results of the finite sample size of the simulation, which is also reflected in the shaded blue confidence intervals. The damped oscillations of the autocorrelation functions are expected, as also indicated in Table 2 on Page 16.

1.3

We now repeat the simulation with $\phi_1 = -0.6$ and $\phi_2 = -0.3$ and plot the results in Figure 3 and Figure 4.

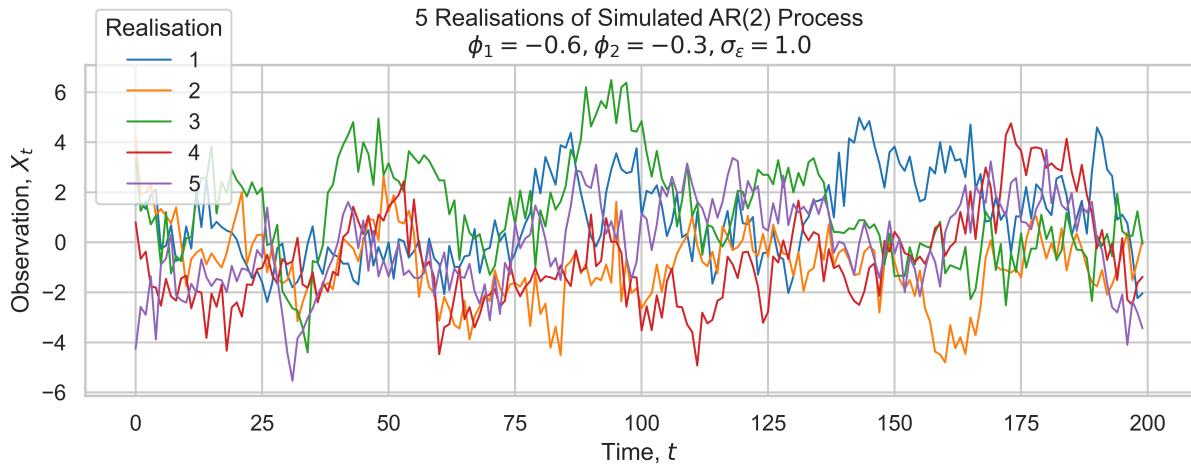


Figure 3: Simulation and Autocorrelation function of AR(2) process with $\phi_1 = -0.6, \phi_2 = -0.3$.

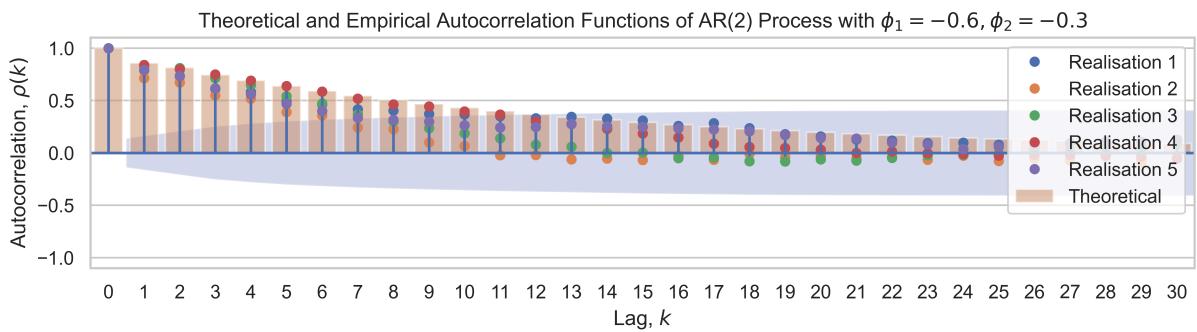


Figure 4: Theoretical and empirical autocorrelation functions of an AR(2) process with $\phi_1 = -0.6, \phi_2 = -0.3$. Shaded blue region is the 95% confidence interval for the empirical ACF.

We observe that the process remains stationary, but the characteristic correlation times are now much longer than for the previous case in Section 1.1 and Section 1.2. In Figure 3 we can clearly see that variations in the time series are much more pronounced than in Figure 1.

1.4

We now repeat the simulation with $\phi_1 = 0.6$ and $\phi_2 = -0.3$ and plot the results in Figure 5 and Figure 6.

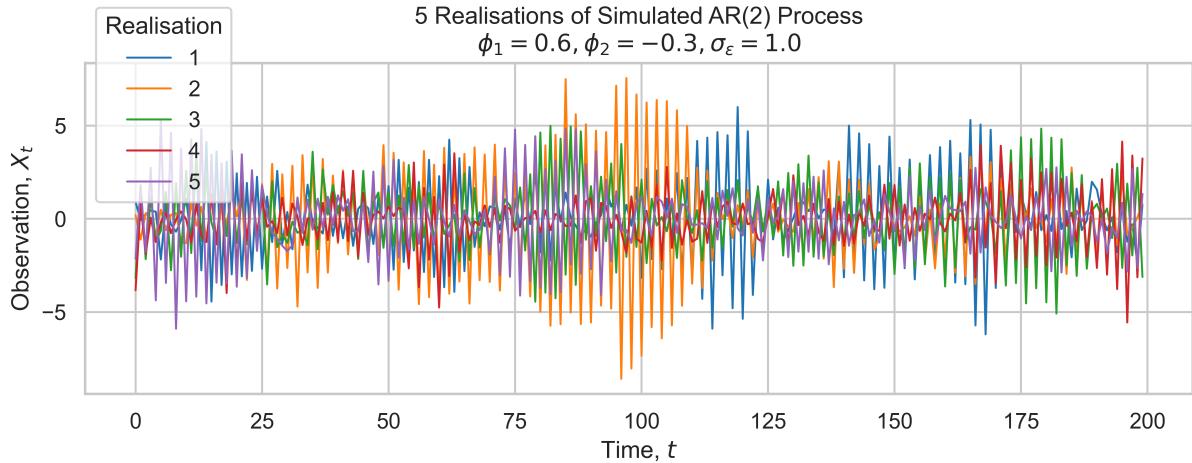


Figure 5: Simulation and Autocorrelation function of AR(2) process with $\phi_1 = 0.6, \phi_2 = -0.3$.

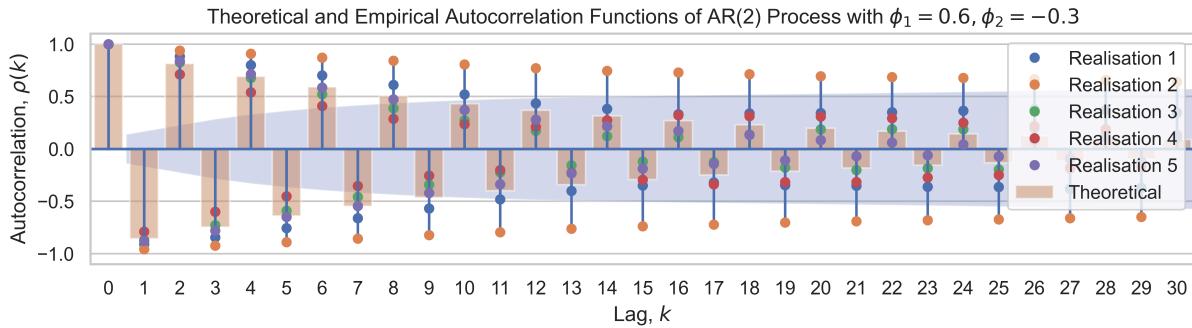


Figure 6: Theoretical and empirical autocorrelation functions of an AR(2) process with $\phi_1 = 0.6, \phi_2 = -0.3$. Shaded blue region is the 95% confidence interval for the empirical ACF.

We again observe that the process is stationary, but find that the sign change in ϕ_2 with respect to Section 1.3 has introduced a fast oscillation in the time series.

We find that for realisations 1 and 3 the autocorrelation does decay slower than expected by theory. This is again attributed to the finite sample size of the simulation, the artifacts of which become more pronounced for processes with longer characteristic correlation times such as the one observed in Figure 5.

We propose increasing the number of simulated observations, n , to a larger number for such processes if the empirical autocorrelation function is to be used for identification or interpretation.

1.5

We again repeat the simulation with $\phi_1 = -0.7$ and $\phi_2 = -0.3$ and plot the results in Figure 7 and Figure 8.

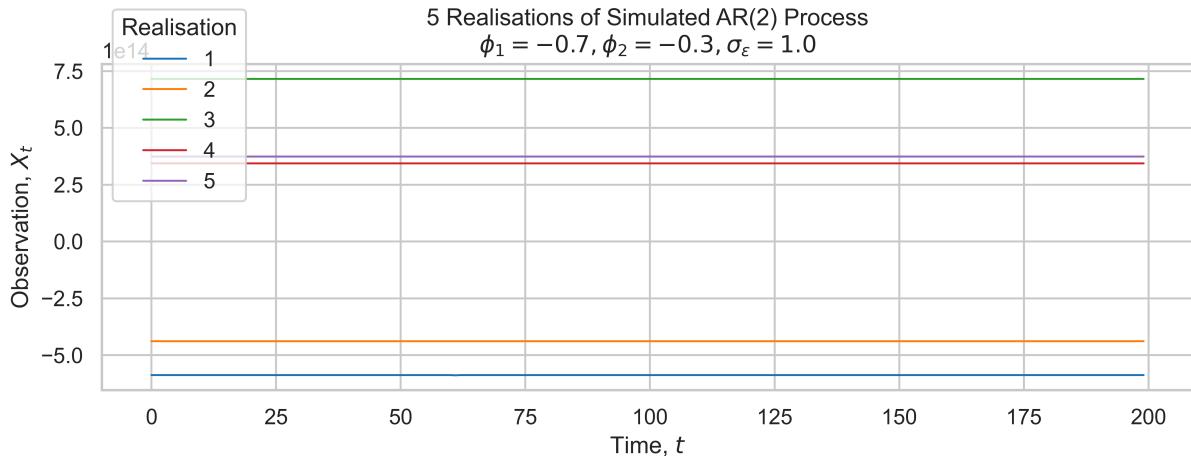


Figure 7: Simulation and Autocorrelation function of AR(2) process with $\phi_1 = -0.7, \phi_2 = -0.3$.

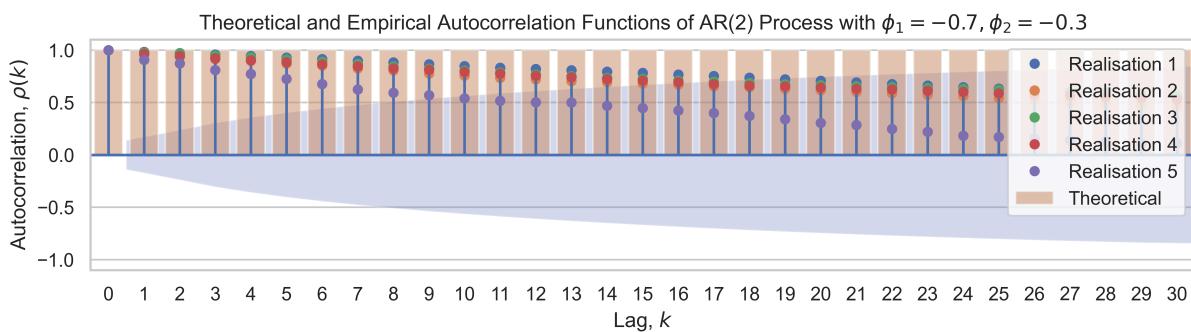


Figure 8: Theoretical and empirical autocorrelation functions of an AR(2) process with $\phi_1 = -0.7, \phi_2 = -0.3$. Shaded blue region is the 95% confidence interval for the empirical ACF.

In figure Figure 7 we observe that the process is exactly at the critical point of stationarity, where the theoretical autocorrelation function does not decay to zero.

This can be understood by considering the stationarity conditions for an AR(2) process, which require that the roots, $z_{1,2}$, of the *characteristic polynomial* lie outside the unit circle, such that $|z_{1,2}| > 1$.

For the AR(2) process in Eq. 1 the characteristic polynomial is given by:

$$P(z) = 1 + \phi_1 z + \phi_2 z^2 \quad (9)$$

Where the roots are then given by:

$$z_{1,2} = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2} \quad (10)$$

We find:

$$|z_{1,2}| = \{1, 3.33\} \quad (11)$$

Where clearly, $|z_1| = 1 \not> 1$, violates the strict inequality of the stationarity condition, the consequence of which is a process where shocks do not decay away over time.

We note that the process does not explode because we are exactly at the point of criticality.

1.6

Once more we repeat the simulation, this time with $\phi_1 = -0.75$ and $\phi_2 = -0.3$ and plot the results in Figure 9 and Figure 10.

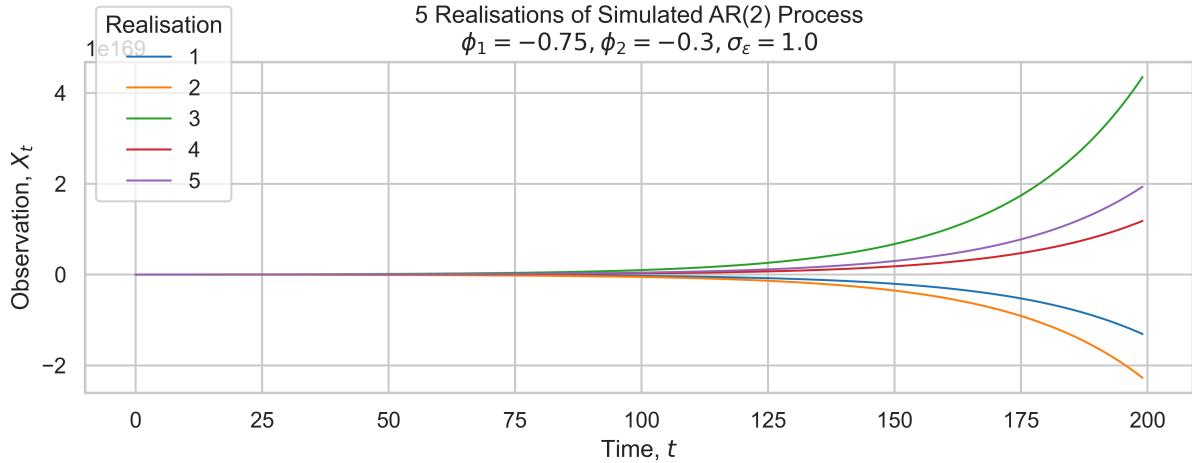


Figure 9: Simulation and Autocorrelation function of AR(2) process with $\phi_1 = -0.75, \phi_2 = -0.3$.

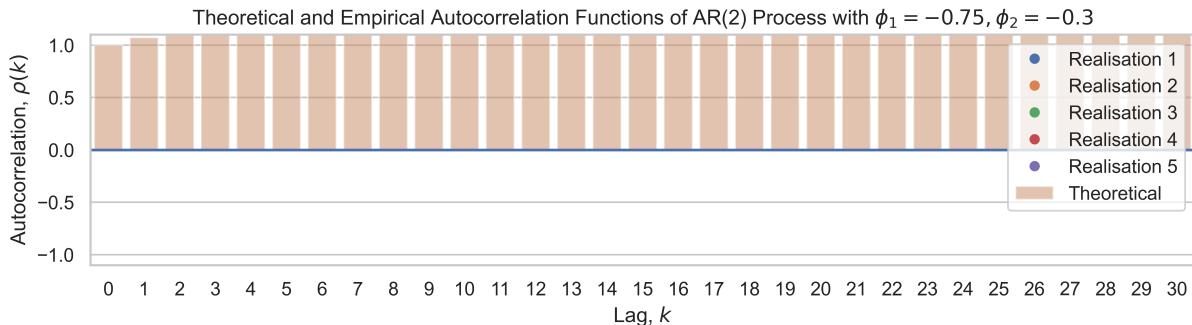


Figure 10: Theoretical and empirical autocorrelation functions of an AR(2) process with $\phi_1 = -0.75, \phi_2 = -0.3$.

Using Eq. 10 we find:

$$|z_{1,2}| \approx \{0.963, 3.463\} \quad (12)$$

Which reveals that the process is decidedly not stationary and should blow up as time progresses, which is also what we observe in Figure 9 and Figure 10.

Figure 10 fails to compute the empirical ACF, while the theoretical ACF quickly increases with time.

1.7

While a lot of information may be gleaned from the autocorrelation function alone, it is useful to also plot the time series itself to get a better understanding of the process, particularly for processes with long characteristic correlation times. For instance, the autocorrelation does not give information about the scale amplitude of the process.

For higher order AR processes, the autocorrelation function may be difficult to interpret, in which case it is again useful to plot the time series itself.

Lastly, the autocorrelation function provides no information about the process realisations at specific points in time, which is often useful for understanding or interpreting the underlying process.

2 Predicting Monthly Solar Power

We are given a seasonal AR model:

$$(1 + \phi_1 B)(1 + \Phi_1 B^{12})(\log(Y_t) - \mu) = \varepsilon_t \quad (13)$$

where Y_t is the monthly energy from the plant in MWh, $\{\varepsilon_t\}$ is a white-noise process with variance σ_ε^2 . The parameters $\phi_1 = -0.38$, $\Phi_1 = -0.94$ and $\mu = 5.72$ are assumed to be known. Based on 36 observations, it is found that $\sigma_\varepsilon^2 = 0.22^2$.

2.1 Forecasting and Residuals

2.1.a Reformulation

Using a substitution $X_t = \log(Y_t) - \mu$ we can rewrite the model, Eq. 13, as:

$$(1 + \phi_1 B)(1 + \Phi_1 B^{12})X_t = \varepsilon_t \quad (14)$$

We are then asked to rewrite the model in order to calculate the residuals $\hat{\varepsilon}_{t+1|t}$, which we understand to be the estimated one-step-ahead forecast error, such that $\hat{\varepsilon}_{t+1|t} := X_{t+1} - \hat{X}_{t+1}$. Recalling $B^q X_t := X_{t-q}$, we expand Eq. 14 to find:

$$\begin{aligned} & (1 + \phi_1 B)(1 + \Phi_1 B^{12})X_t = \varepsilon_t \\ & X_t + \phi_1 B X_t + \Phi_1 B^{12} X_t + \phi_1 \Phi_1 B^{13} X_t = \varepsilon_t \\ & X_t + \phi_1 X_{t-1} + \Phi_1 X_{t-12} + \phi_1 \Phi_1 X_{t-13} = \varepsilon_t \end{aligned} \quad (15)$$

Multiplying through by B^{-1} we find:

$$X_{t+1} + \phi_1 X_t + \Phi_1 X_{t-11} + \phi_1 \Phi_1 X_{t-12} = \varepsilon_{t+1} \quad (16)$$

Solving for X_{t+1} we find:

$$X_{t+1} = \varepsilon_{t+1} - \phi_1 X_t - \Phi_1 X_{t-11} - \phi_1 \Phi_1 X_{t-12} \quad (17)$$

We then construct the *optimal predictor* for X_{t+1} , [1, eq. 5.139]:

$$\begin{aligned} \hat{X}_{t+1} &= \mathbb{E}[X_{t+1}|X_t, X_{t-1}, \dots] \\ &= -\hat{\phi}_1 X_t - \hat{\Phi}_1 X_{t-11} - \hat{\phi}_1 \hat{\Phi}_1 X_{t-12} \quad \mathbb{E}[\varepsilon] = 0 \end{aligned} \quad (18)$$

Where we have used the fact that the $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. Thus the one-step-ahead forecast error becomes:

$$\begin{aligned} \hat{\varepsilon}_{t+1|t} &= X_{t+1} - \hat{X}_{t+1} \\ &= X_{t+1} + \hat{\phi}_1 X_t + \hat{\Phi}_1 X_{t-11} + \hat{\phi}_1 \hat{\Phi}_1 X_{t-12} \end{aligned} \quad (19)$$

2.1.b Verifying Residuals in Dataset

As already noted in the section above, we assume that the residuals are independent and identically distributed (i.i.d.) with a zero-mean normal distribution.

We can verify that this is indeed the case by constructing the appropriate seasonal AR model using the SARIMAX class from statsmodels. Again we take careful notice of the signs of the parameters ϕ_1, Φ_1 , both of which must be negated to match the convention of the SARIMAX class.

Utilising the built-in `plot_diagnostics` function of the fitted model, we produce Figure 11.

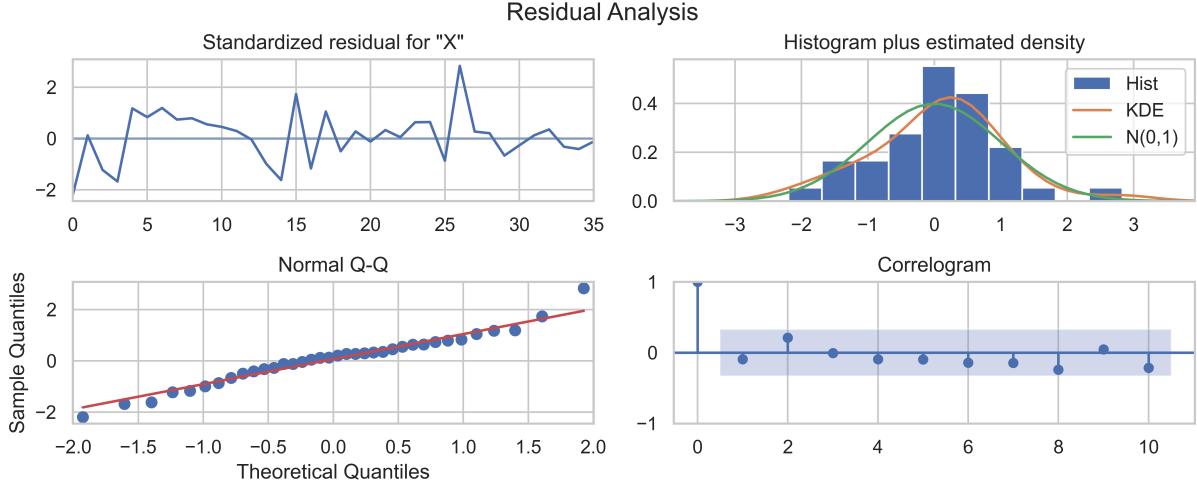


Figure 11: Residual analysis of the solar power dataset using $\phi_1 = -0.38$, $\Phi_1 = -0.94$ and $\sigma_\varepsilon^2 = 0.22^2$ in the notation established by Eq. 14. Note that the residual analysis is performed on the transformed data $X_t = \log(Y_t) - \mu$.

By inspection of Figure 11 we find no evidence that the residuals violate the assumption of i.i.d. with a zero-mean normal distribution. In particular, we find that the standardised residuals do not exhibit strong systematic patterns, although the first summer (observations 5-10) does appear to be skewed slightly positive in the residuals. Additionally, we find the residuals to be well-described by a normal distribution as seen in the Q-Q plot and histogram.

One could carry out a Ljung-Box test to further substantiate this claim, but we find that this is not necessary in this case.

2.2 Forecasting

Using the same model that we fitted in Section 2.1.b, we produce a 12 month forecast ($k = 12$) using the forecast method.

Importantly, the forecasted data must be transformed back into the original scale of the dataset:

$$\begin{aligned} X_t &:= \log(Y_t) - \mu && \text{Forward} \\ Y_t &= e^{X_t + \mu} && \text{Backward} \end{aligned} \tag{20}$$

Time	Power, Y_t [MWh]	X_t
2011-01-01	101.28	-1.10
2011-02-01	147.23	-0.73
2011-03-01	156.92	-0.66
2011-04-01	142.55	-0.76
2011-05-01	355.31	0.15
2011-06-01	538.02	0.57
2011-07-01	611.21	0.70
2011-08-01	631.89	0.73
2011-09-01	630.11	0.73
2011-10-01	508.16	0.51
2011-11-01	375.01	0.21
2011-12-01	280.49	-0.08

Table 1: Twelve month forecast using model given in Eq. 13.

Which we can plot alongside the original data to produce Figure 12.

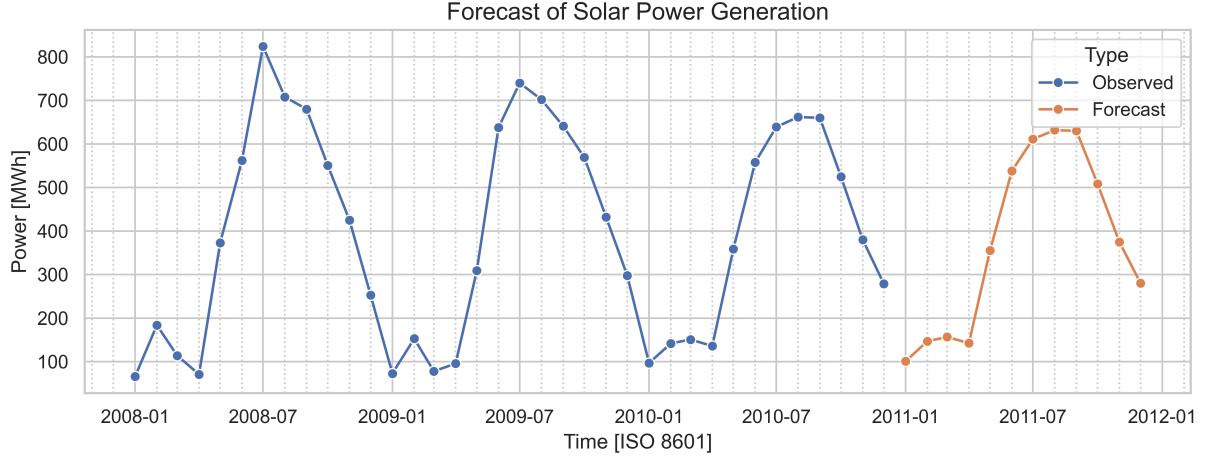


Figure 12: Twelve month forecast using model given in Eq. 13.

We find that the forecasted data matches our intuition of the data well and follows the seasonal pattern observed in the historical data of power generation.

2.3 Prediction Intervals

We now wish to consider the *variance* in our *predictions*, or rather we seek to obtain the *variance of the prediction error*, Eq. 19:

$$\text{Var}(\hat{\varepsilon}_{t+1|t}) = \text{Var}(X_{t+1} - \hat{X}_{t+1}) \quad (21)$$

In order to carry out this calculation, we convert our process to a Moving-Average (MA) process:

$$\begin{aligned} X_t &= \frac{\varepsilon_t}{(1 + \phi_1 B)(1 + \Phi_1 B^{12})} & \Leftrightarrow & (13) \\ &= (1 + \psi_1 B + \psi_2 B^2 + \dots) \varepsilon_t + \dots & \text{Definition of MA process} & (22) \end{aligned}$$

From Eq. 22 we recover the constraint:

$$\begin{aligned} (1 + \psi_1 B + \psi_2 B^2 + \dots)(1 + \phi_1 B)(1 + \Phi_1 B^{12}) &= 1 \\ (1 + \psi_1 B + \psi_2 B^2 + \dots)(1 + \phi_1 B + \Phi_1 B^{12} + \phi_1 \Phi_1 B^{13}) &= 1 \end{aligned} \quad (23)$$

Which imposes that all coefficients of the power series of B must vanish. Comparing powers of B^i we find the following constraints on the weights ψ_i of the transfer function:

$$\psi_i = \begin{cases} (-\phi_1)^i & 1 \leq i \leq 11 \\ (-\phi_1)^i - \Phi_1(-\phi_1)^{i-12} & 12 \leq i \leq 23 \\ \vdots & \vdots \end{cases} \quad (24)$$

Using our MA formulation we again consider the expression of X_{t+1} using the treatment in [1, sec: 5.7.1]:

$$X_{t+k} = \varepsilon_{t+k} + \psi_1 \varepsilon_{t+k-1} + \dots + \psi_k \varepsilon_t + \psi_{k+1} \varepsilon_{t-1} + \dots \quad (25)$$

We consider the expectation value of the residual ε_{t+k} . For past observations, when $k \leq 0$, have a realisation of the residual and thus this will be our expected value. For future observations, when $k > 0$, we have no realisation of the residual and thus the expected value is the mean of the distribution of ε , which we recall to be zero:

$$\mathbb{E}[\varepsilon_{t+k}|X_t, X_{t-1}, \dots] = \begin{cases} \varepsilon_{t+k} & k \leq 0 \\ 0 & k > 0 \end{cases} \quad (26)$$

From which we can obtain our k -step predictions, $X_{t+k|t}$:

$$\begin{aligned}
X_{t+k|t} &= \mathbb{E}[X_{t+k}|X_t, X_{t-1}, \dots] \\
&= \underline{\mathbb{E}[\varepsilon_{t+k}|X_t, X_{t-1}, \dots]} + \psi_1 \underline{\mathbb{E}[\varepsilon_{t+k-1}|X_t, X_{t-1}, \dots]} + \dots \quad \Leftarrow (26) \\
&\quad + \psi_k \mathbb{E}[\varepsilon_t|X_t, X_{t-1}, \dots] + \psi_{k+1} \mathbb{E}[\varepsilon_{t-1}|X_t, X_{t-1}, \dots] + \dots \\
&= \psi_k \varepsilon_t + \psi_{k+1} \varepsilon_{t-1} + \dots
\end{aligned} \tag{27}$$

Using our previous definition of the prediction error, Eq. 19, we find:

$$\begin{aligned}
\hat{\varepsilon}_{t+k|t} &= X_{t+k} - \hat{X}_{t+k} \\
&= \varepsilon_{t+k} + \psi_1 \varepsilon_{t+k-1} + \dots + \underline{\psi_k \varepsilon_t + \psi_{k+1} \varepsilon_{t-1} + \dots} \\
&\quad - (\underline{\psi_k \varepsilon_t + \psi_{k+1} \varepsilon_{t-1} + \dots}) \\
&= \varepsilon_{t+k} + \psi_1 \varepsilon_{t+k-1} + \dots + \psi_{k-1} \varepsilon_{t+1}
\end{aligned} \tag{28}$$

Where the variance of this prediction error at time $t + k$ is given by:

$$\begin{aligned}
\sigma_k^2 &= \text{Var}(\hat{\varepsilon}_{t+k|t}) \\
&= \text{Var}(\varepsilon_{t+k} + \psi_1 \varepsilon_{t+k-1} + \dots + \psi_{k-1} \varepsilon_{t+1}) \\
&= \text{Var}(\varepsilon_{t+k}) + \psi_1^2 \text{Var}(\varepsilon_{t+k-1}) + \dots + \psi_{k-1}^2 \text{Var}(\varepsilon_{t+1})
\end{aligned} \tag{29}$$

We consider that $\hat{\varepsilon}_{t+k|t} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ and thus find $\text{Var}(\varepsilon_{t+k}) = \sigma_\varepsilon^2$ for all $k > 0$. Plugging this into Eq. 28 yields:

$$\sigma_k^2 = \sigma_\varepsilon^2(1 + \psi_1^2 + \dots + \psi_{k-1}^2) \tag{30}$$

This enables us to calculate the $(1 - \alpha)$ confidence interval of our k -step predictions:

$$\hat{X}_{t+k|t} \pm u_{\alpha/2} \sigma_k = \hat{X}_{t+k|t} \pm u_{\alpha/2} \sigma_\varepsilon \sqrt{1 + \psi_1^2 + \dots + \psi_{k-1}^2} \tag{31}$$

Where $u_{\alpha/2}$ is the $\alpha/2$ quantile of the standard normal distribution. For $\alpha = 5\%$ we find $u_{\alpha/2} \approx 1.96$.

With this, we are able to calculate the prediction intervals for our forecasted data using by leveraging the expression for the weights ψ_i in Eq. 24, which yields the 95% prediction intervals shown in Figure 13.

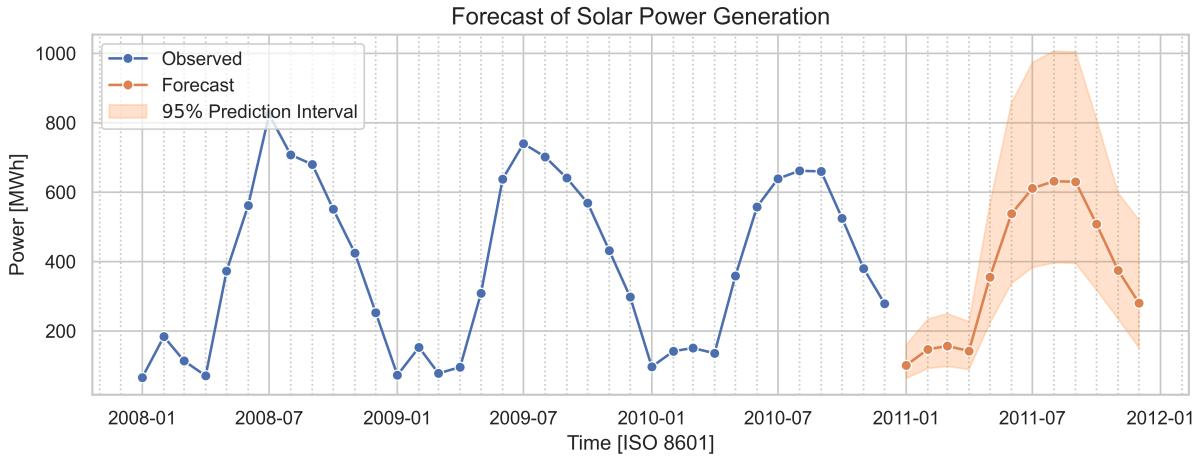


Figure 13: Twelve month forecast using model given in Eq. 13 with 95% prediction intervals.

We note that the for $1 \leq k \leq 11$ the seasonal component does not contribute to the prediction error as also hinted in the assignment description, but for $k = 12$ we find a linear dependence on the coefficient of the seasonal component, Φ_1 . This is a significant difference when contrasted against not including the seasonal component.

2.4 Forecast Commentary

We consider the forecast found in Figure 13 to be a relatively accurate description of the expected power generation of the solar plant during the year of 2011. In the original dataset we observe a slight decline in

the peak power generation as a function of time, which is also observed in the forecast. This can be attributed to expected degradation of the photo-voltaic solar panels over time.

Notably, the prediction intervals are not symmetric over the expected value, which is a result of the non-linear transformation of the data as described in Eq. 20.

Additionally, we observe that the prediction intervals are quite wide, which again can be understood in part by referring to the exponential transformation of the data. Additionally we observe large variations in the shape of the observations of the 3 year history of the solar plant, which pushes up the estimated variance of the residuals, $\sigma_e^2 = 0.22^2$. This strongly influences the width of the prediction intervals, which is particularly notable at power generation observations due to the transformation. This is simply a consequence of linearising the data in order to fit it to our linear model.

From inspection, we find that the prediction intervals are likely too wide during the summer months where power generation is at its peak, due to the overly large effect of the variance during the winter months on the residual error estimates.

3 An ARX Model for the Heating of a Box

We are given a data set of hourly measurements from a box instrumented with a heater, an internal temperature sensor and an external temperature sensor.

The box has window on its south-facing wall, onto which the vertical solar radiation is also recorded in the data set.

The internal temperature of the box was kept approximately constant using a thermostatic control of the internal heater.

In this section, we will be working AutoRegressive models with eXogenous variables (ARX), for which we will adopt a notation in which $\text{AR}(p)\text{-X}(e_1, e_2, \dots)$ refers to an ARX where p determines the order of the AutoRegressive (AR) part of the model while e_j for $j \in \mathbb{N}_+$ refers to the order of the j th exogenous variable X_j .

This yields the following generic ARX model:

$$Y_t = c - \underbrace{\sum_{i=1}^p \phi_i Y_{t-i}}_{\text{AR Part}} + \underbrace{\sum_{j=1}^J \sum_{k=0}^{e_j} \omega_{j,k} X_{j,t-k}}_{\text{Exogenous Part}} + \varepsilon_t \quad (32)$$

Where c is a constant offset, ϕ_i are the AR coefficients, $\omega_{j,k}$ are the exogenous coefficients and ε_t is a white-noise such that $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. J refers to the number of exogenous variables such that $j \in \{1, 2, \dots, J\}$ become the indices of the exogenous variables.

To further unify notation, we assume a model, where $\{T_t\}$ represents the series T_{delta} , $\{G_t\}$ is the series G_v and $\{P_t\}$ is the series P_h for $t \in \{1, \dots, 167\}$ in the training dataset.

3.1 Exploratory Dependency Analysis

We load the data set and seek to visually explore the dependency between the heater power P_h , temperature delta between inside and outside, ΔT , and the vertical solar radiation G_v .

For the purposes of visualisation, we standardize the data by subtracting the mean and dividing by the standard deviation, which we then plot in Figure 14.

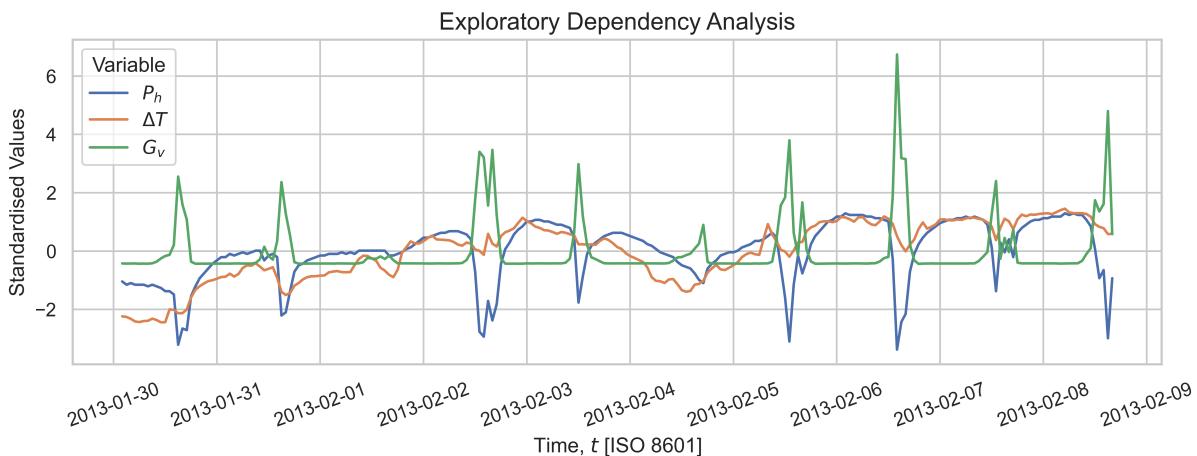


Figure 14: Standardized data set of heater power P_h , temperature delta ΔT and vertical solar radiation G_v .

Figure 14 clearly shows an inverse relationship between the heater power and solar radiation, as we would intuitively expect on physical grounds – for a constant temperature difference, we would expect the sum of the heater power and solar radiation to be constant, thus requiring the heater to compensate for any changes in solar heating.

We additionally observe a positive correlation between the heater power and the temperature difference, which again matches our physical understanding that a larger temperature difference would lead to a larger

heat loss from the box, assuming the inside temperature is larger than the external temperature. Thermostaticity would then require a larger heater power to compensate for the increase in heat loss.

3.2 Test/Train Split

We now split the dataset into two parts, one for training and one for testing. The cut-off point is set to 2013-02-06 00:00, which is the last observation of the training set.

3.3 Further Dependency Investigation

Following on from the investigation in Section 3.1, we plot the three variables against each other in the pair plots found in Figure 15.



Figure 15: Pair plot of the training data set.

In Figure 15 we again observe that the heater power P_h and the temperature difference ΔT are positively correlated, with the heater power being larger when the temperature difference is larger. Interestingly, we find that the temperature difference as a function of the heater power roughly follows one of two lines, which we can further understand when inspecting the pair plot between the temperature difference and the solar radiation G_v , which appears to be roughly bimodal. This does not lend itself to immediate interpretation, but does suggest that there is an underlying structure in the data that may be elucidated with further analysis.

Additionally, we confirm the observation from Section 3.1 that the heater power and solar radiation G_v are negatively correlated.

Additionally we plot the autocorrelations and cross-correlations in a similar matrix to the one found in Figure 15, which reveals characteristic timescales for the three variables and their correlations.

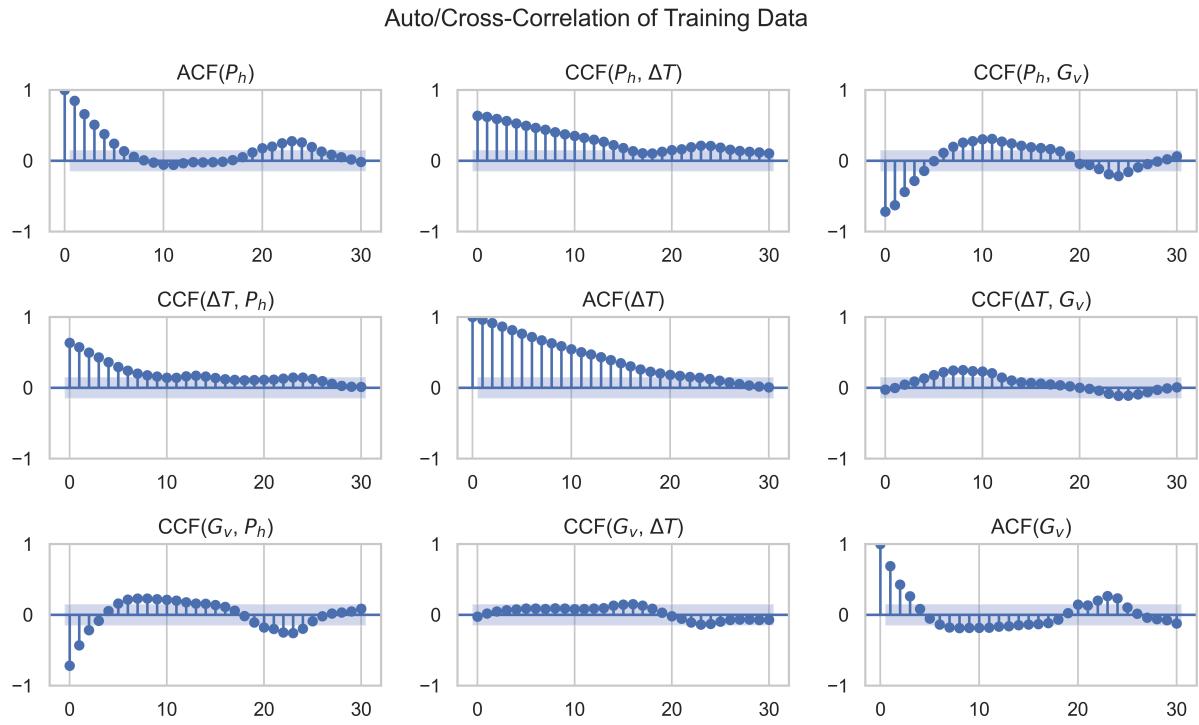


Figure 16: Autocorrelation and cross-correlation of the training data set.

In Figure 16 we find a seasonality of approximately 24 hours in the heater power P_h and solar radiation G_v , which we would expect from the diurnal solar cycle.

The temperature difference does not follow this pattern, partly because it has a much longer timescale in the data set. This could be attributed to more complex weather patterns, such as different wind patterns or nightly cloud cover. The characteristic timescale of the autocorrelation function of the temperature difference reveals that the external temperature varies rather slowly compared to the other two variables.

The inverse relationship between solar radiation and heater power is also confirmed in the cross-correlation plot, where we find a negative correlation between the two variables. We find only a small correlation between the temperature difference and the solar radiation.

Lastly, we inspect the partial autocorrelation functions of the three variables in Figure 17.

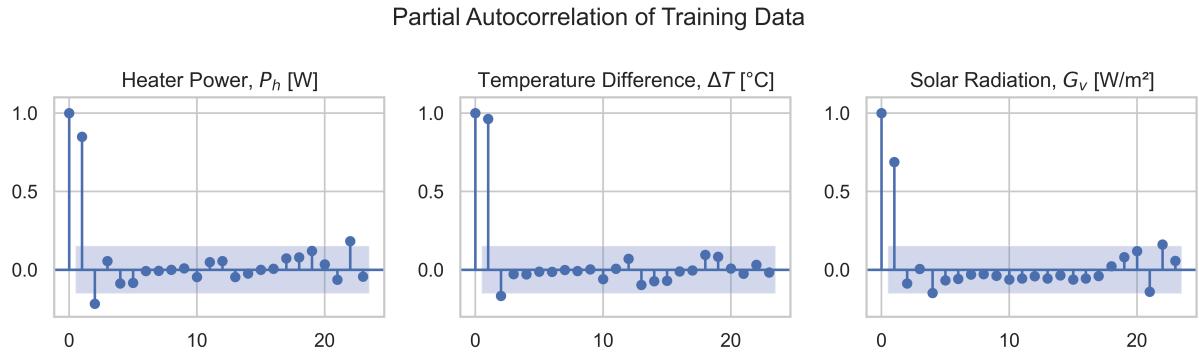


Figure 17: Partial autocorrelation of the training data set.

Comparing Figure 17 with the ARMA order identification table, Table 2, we find that the heater power P_h and temperature difference ΔT may be well described by an AR(2) process, while the solar radiation G_v appears to be well described by an AR(1) process.

The heater power P_h and solar radiation G_v additionally show evidence of seasonality, suggesting a seasonality of 24 hours may be appropriate for modelling their behaviour.

	ACF $\rho(k)$	PACF φ_{kk}
AR(p)	Damped exponential and/or sine functions	$\varphi_{kk} = 0$ for $k > p$
MA(q)	$\rho(k) = 0$ for $k > q$	Dominated by damped exponential and/or sine functions
ARMA(p, q)	Damped exponential and/or sine functions after lag $q - p$	Dominated by damped exponential and/or sine functions after lag $p - qs$

Table 2: Reproduction of Table 6.1 from [1, p. 155] showing the expected behaviour of the autocorrelation function for different ARMA processes

Notably, we cannot reasonably estimate the parameters of, for example, an ARX model from the figures above, will instead require an outright fitting of the model(s) to the data.

3.4 Impulse Response

While the assignment description does not explicitly state the orders of the ARX model, which naturally need to be determined prior to carrying out an impulse response analysis, we consider an AR(1)-X(1, 1) model to be a reasonable choice, which is supported by the analysis of the PACF in Figure 17 above.

We consider the heater power, P_h , to be the *endogenous variable*, while the solar radiation, G_v , and temperature difference, ΔT , are the *exogenous variables*, which yields the following model:

$$P_t = -\phi_1 P_{t-1} + \omega_1 T_t + \omega_2 G_t + \varepsilon_t \quad (33)$$

Where parameters $\phi_1, \omega_1, \omega_2 \in \mathbb{R}$, noting that we have also used the convention describing the different time series described in the introduction in Section 3.

We model two impulse responses in G_t and T_t separately and find that the unit impulse given from the variable T_t seems more significant. As Figure 15 shows, it has a positive correlation with the target series P_t , while Figure 14 clearly indicates, the variable G_t has an inverse relationship with the target P_t . Even before estimating anything, we can assume that a unit impulse from T_t will be impactful, while for G_t the response will decay extremely fast, as the original series have opposing effects.

We will use the the AutoRegclass from `statsmodels.tsa.ar_model` in Python, which allows to provide an exogenous variable and has built in parameter estimation (via OLS and conditional MLE). To double check, we also did our own classic OLS fit on the design matrix: $X = [P_{t-1}, T_t, G_t]$ noted in the form of column vectors of the corresponding series, which yields a parameter vector of $\Theta = [-\phi_1, \omega_1, \beta_1]^T \in \mathbb{R}^{p+e_1+e_2}$.

$$\begin{aligned} \Rightarrow P_{t-p,n} &= X \cdot \Theta + \varepsilon_t = Y \\ \Leftrightarrow \hat{\Theta} &= (X^T X)^{-1} X^T Y \end{aligned} \quad (34)$$

The resulting parameters are given as:

ϕ_1	ω_1	β_1
0.4127	2.3352	-0.0836

Table 3: AR(1)-X(1,1) coefficients with T_d and G_v exogenous of order 1

For the impulse response function, we give the unit impulse first from T_t and use the recursion

$$\begin{aligned} \text{IR (0)} &= 0\phi_1 + 1\omega_1 + 0\beta_1 \\ \text{IR (1)} &= \text{IR (0)}\phi_1 + \text{IR (0)}\omega_1 + \text{IR (0)}\beta_1 \\ \text{IR (2)} &= \text{IR (1)}\phi_1 + \text{IR (0)}\phi_2 + 0\omega_1 + 0\beta_1 \\ \text{IR (3)} &= \text{IR (2)}\phi_1 + \text{IR (1)}\phi_2 + \text{IR (0)}\phi_3 + 0\omega_1 + 0\beta_1 \\ &\vdots \\ \text{IR (}k=15\text{)} &= \text{IR (14)}\phi_1 + \text{IR (13)}\phi_2 + \dots + \text{IR (5)}\phi_{10} + 0\omega_1 + 0\beta_1 \\ &\vdots \end{aligned} \quad (35)$$

Or in shorter sum notation this would be with $\text{IRF}(0) = 1\omega_1 + 0\beta_1$ set as initial value:

$$\text{IRF}(k) = \sum_{i=1}^k \sum_{j=0}^{\min(i,p)} [\phi_j \text{IRF}(i-j-1) \mathbb{I}_{i-j-1 \geq 0}] \quad (36)$$

for $k \in \mathbb{N}$ the lag and p the AR model order.

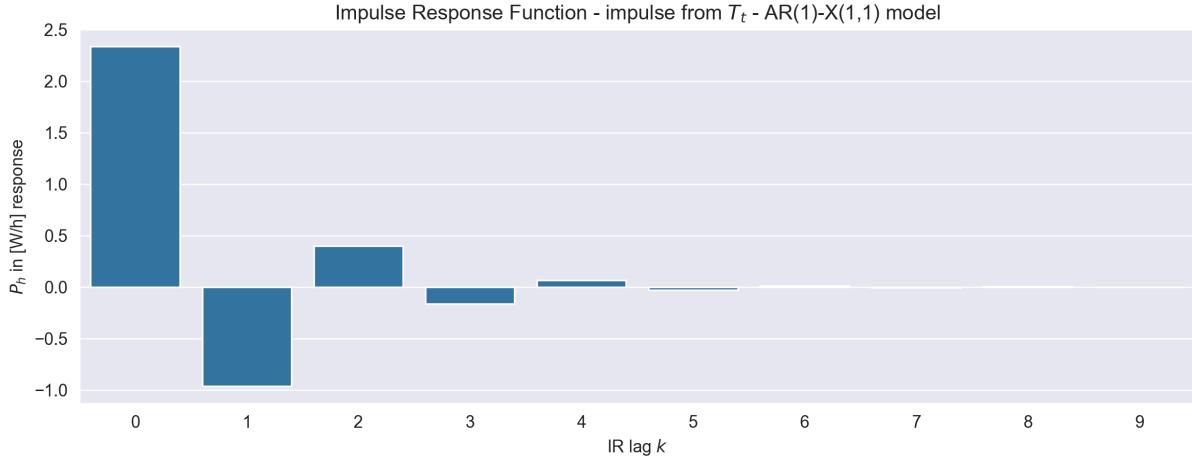


Figure 18: Impulse Response of the AR(1)-X(1,1) with the unit impulse from T_d

Now we give a second plot of a unit impulse from G_v :

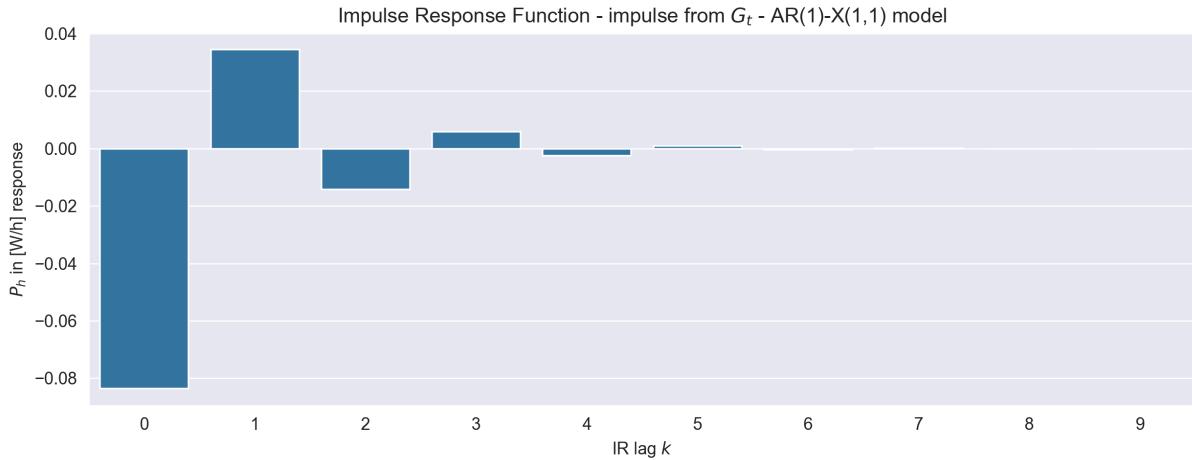


Figure 19: Impulse Response of the AR(1)-X(1,1) with the unit impulse from G_v

The impulse response coefficients v_k are:

impulse	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}
$T_t \rightarrow P_t$	2.3352	-0.9637	0.3977	-0.1641	0.0677	-0.0279	0.0115	-0.0048	0.002	-0.0008
$G_t \rightarrow P_t$	-0.0836	0.0345	-0.0142	0.0059	-0.0024	0.001	-0.0004	0.0002	-0.0001	0.0

Table 4: impulse response coefficients for AR(1)-X(1,1)

As already explained based on Figure 14, the two impulse responses counter-act. The IR (impulse response) for this model decays rather quickly, especially with the impact of an exogenous variable that is reciprocal to the target variable.

We can also confirm that this process is stable given $\sum_{k=1}^{\infty} |v_k| < \infty$.

Because of this stability the model is rather robust to unit shocks from both exogenous variables, more robust from shocks of G_v than from T_t in fact. As a consequence, with this model, we can accept a certain degree of fluctuation in the exogenous variables, without having to worry about a big impact on the

predictive performance of our model. For example, the effect of a measurements error or sensor defect in T_d or G_v , that acts as a shock to our system, will not last for very long. Thus, having a larger order for the AR part of our model protects against shocks from the other variables. On the other hand, the model is more vulnerable to sudden changes in auto-regressive behaviour.

3.5 Linear Regression Model

We are given a linear model with the following form:

$$P_{h,t} = \omega_1 T_{\text{delta},t} + \omega_2 G_{v,t} + \varepsilon_t \quad (37)$$

Where $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ and assumed to be i.i.d. (independent and identically distributed).

We fit such a model to the data and produce the estimation shown in Figure 20.

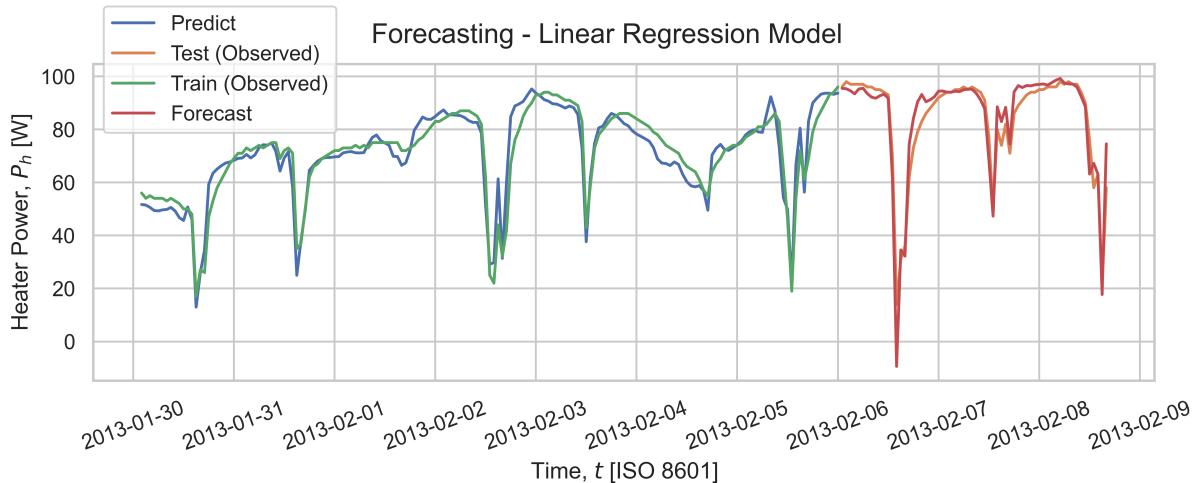


Figure 20: Linear regression model with T_{delta} and G_v as exogenous variables.

Additionally we evaluate the residuals during the training period and one-step prediction errors for the test data set in Figure 21.

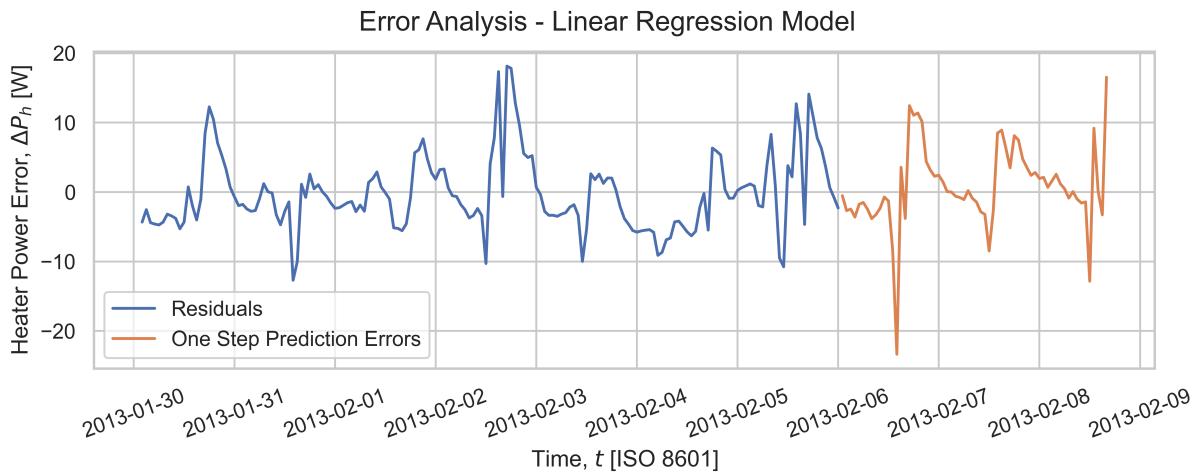


Figure 21: Residual analysis of the linear regression model.

We observe that the errors, here understood to be the residuals for the training data set and the one-step prediction errors for the test data set, look relatively good. We have chosen to plot the test data series and predictions here as well in order to be able to gauge the generalisation performance of the model. We conduct an analysis of the autocorrelation functions in Figure 22.

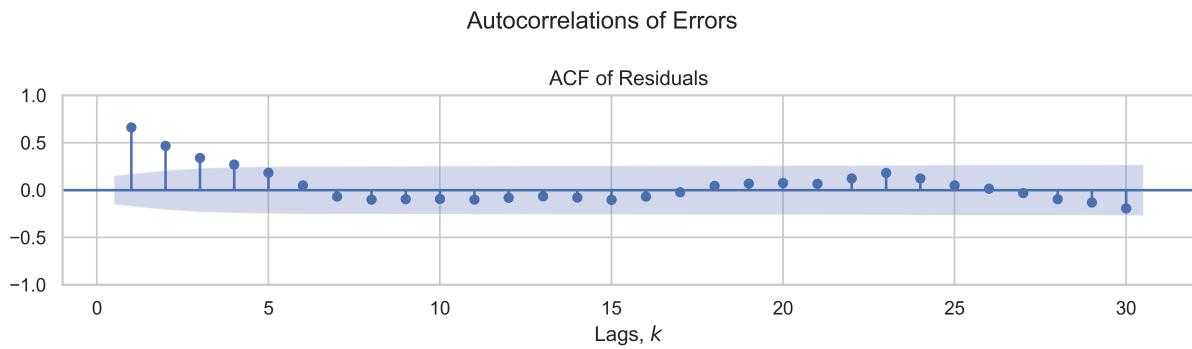


Figure 22: Autocorrelation of the residuals of the linear regression model.

In Figure 22 we observe there remain significant correlations in the residuals, which suggests that the model is not fully capturing the underlying structure of the data.

To further investigate this, we employ a cross-correlation analysis of the residuals and the exogenous variables in Figure 23.

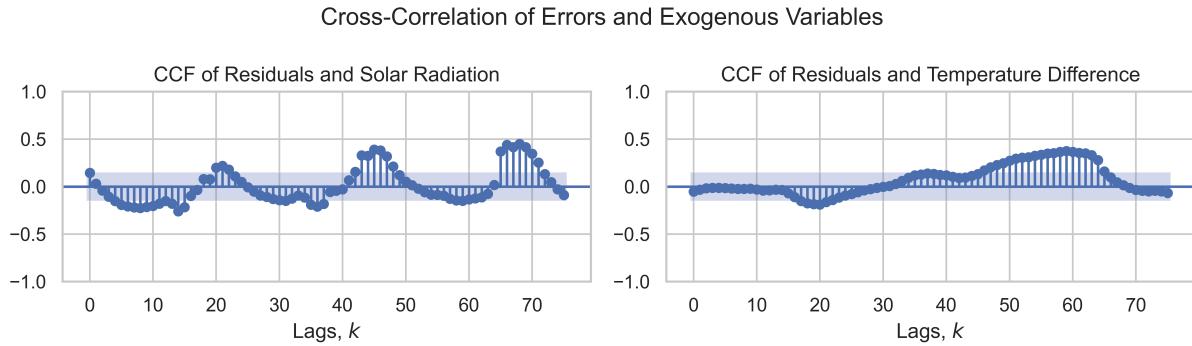


Figure 23: Cross-correlation of the residuals of the linear regression model with the exogenous variables.

Here we observe that some significant correlations remain between the errors and the exogenous variables, which suggests that the model may not be fully capturing the underlying structure of the data and a more complex model may be required.

In order to more formally substantiate this, we employ the Ljung-Box test on the residuals of the model, which reveals that significant correlations remain in the residuals of the model for all choices of lags k .

The fact that the autocorrelation in Figure 22 shows significant correlations with past values strongly suggests that a model with an autoregressive component may be more appropriate. Generally such models of linear time-invariant systems (LTI) are well-described by the use of *transfer functions*, which are a powerful tool for modelling the relationship between the input and output of a system. In the case of an ARX model, its transfer function contains the appropriate AR components expressed as backshifts of the input series.

In order to build up a model which better captures the underlying structure of the data, we may use the idea of *transfer function building*, in which we build up a model from the residuals of a simpler model by employing *pre-whitening*, in which models are used as *filters* and spikes in the CCF plots are used to identify a candidate lag structure.

Pre-whitening would involve fitting an ARMA model on each of the input variables T_t and G_t .

$$\phi(B)G_t = \theta(B)\varepsilon_t \quad \phi_2(B)T_t = \theta_2(B)\varepsilon_t \quad (38)$$

Which when re-arranged will be transformed into the white-noise component:

$$\phi(B)(\theta(B))^{-1}G_t = \varepsilon_t \phi_2(B)(\theta_2(B))^{-1}T_t = \varepsilon_{2,t} \quad (39)$$

Which we can then apply as a ‘filter’ to the input series P_t :

$$\phi(B)(\theta(B))^{-1}P_t = \gamma_t \phi_2(B)(\theta_2(B))^{-1}P_t = \gamma_{2,t} \quad (40)$$

And then calculate the CCF of $\varepsilon_{2,t}$ and $\gamma_{2,t}$ to get the coefficients of a transfer function from T_t to P_t and analogously for G_t .

This is beyond the scope of this assignment, where we will instead follow the model building outlined in the assignment description.

3.6 AR(1)-X(1,1) Model

This time we employ a first order ARX model:

$$P_{h,t} = -\phi_1 P_{h,t-1} + \omega_1 T_{\text{delta},t} + \omega_2 G_{v,t} + \varepsilon_t \quad (41)$$

Where ϕ_1 is the AR coefficient, ω_1 and ω_2 are the exogenous coefficients and ε_t is a white-noise process as before.

We first perform the forecasting and estimation, as shown in Figure 24.

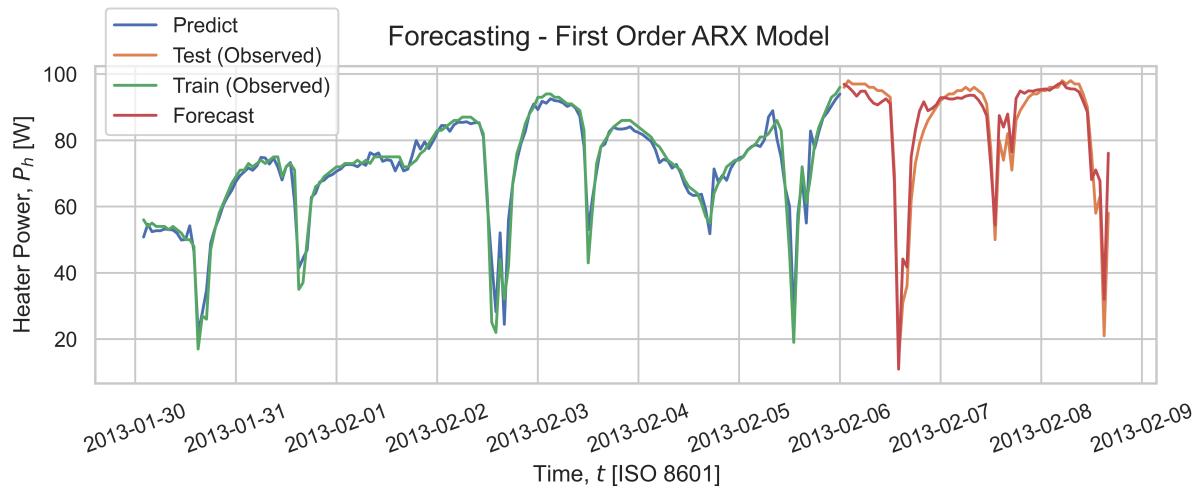


Figure 24: AR(1)-X(1,1) model with T_{delta} and G_v as exogenous variables.

With subsequent error analysis shown in Figure 25.

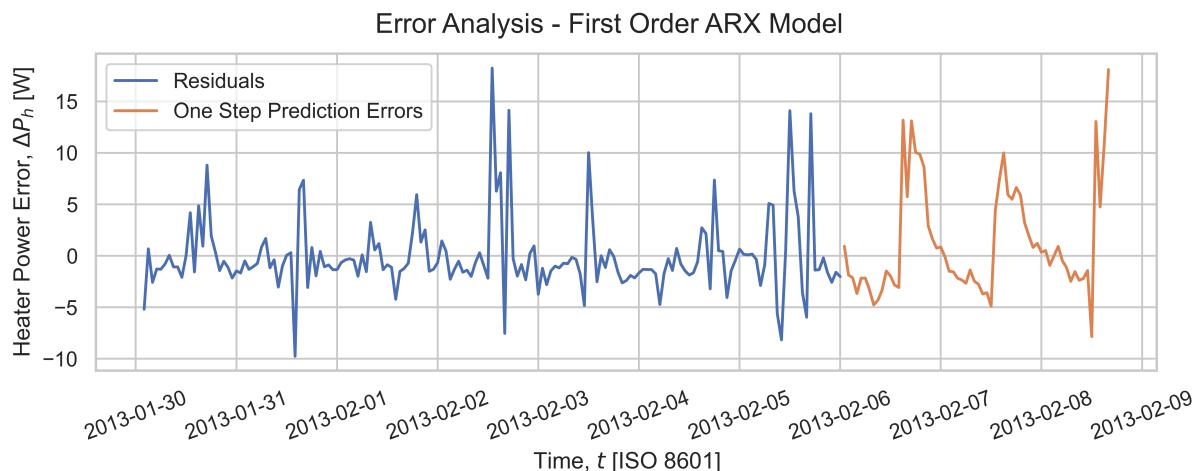


Figure 25: Residual analysis of the AR(1)-X(1,1) model.

With ACF and CCF analysis shown in Figure 26 and Figure 27 respectively.

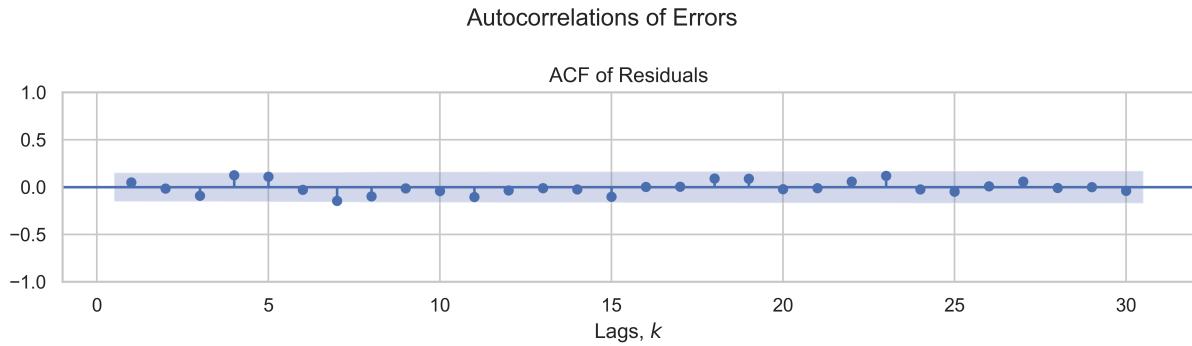


Figure 26: ACF of the residuals of the AR(1)-X(1,1) model.

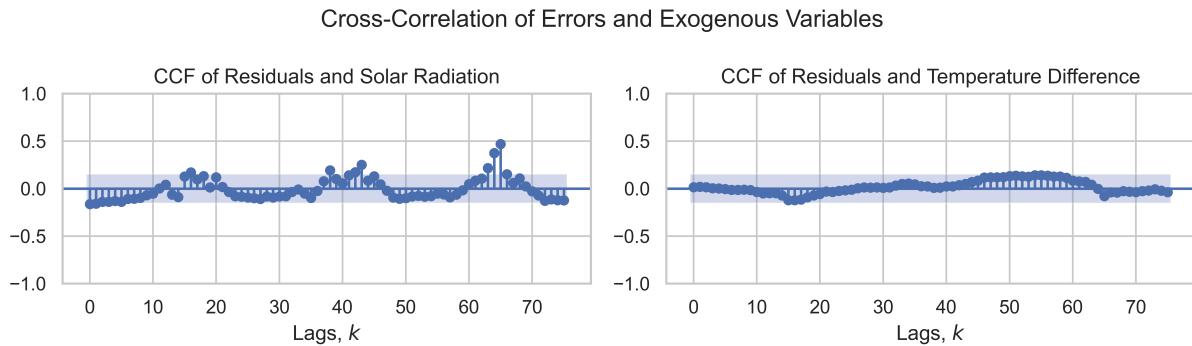


Figure 27: CCF of the residuals of the AR(1)-X(1,1) model with the exogenous variables.

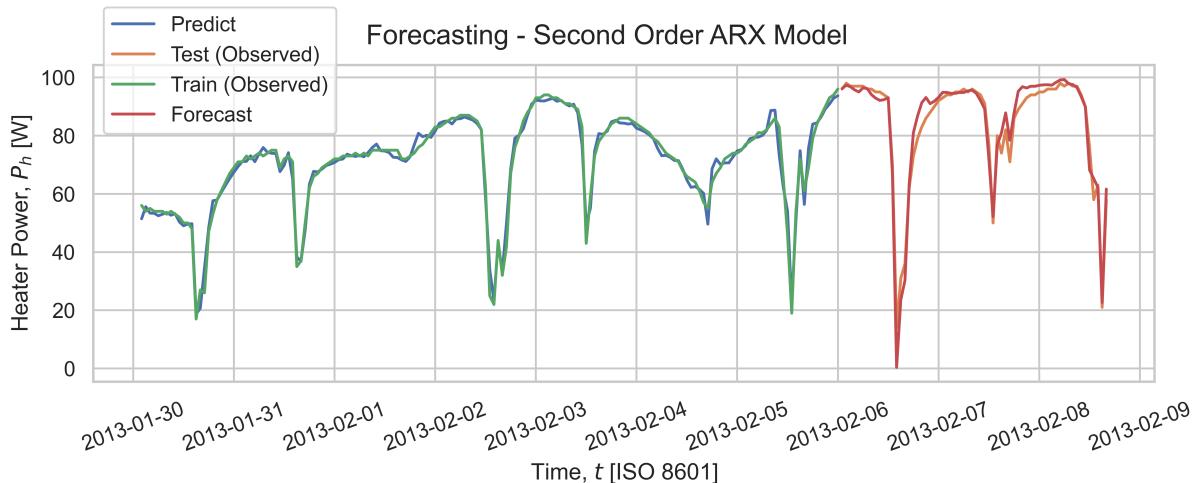
Comparing Figure 22 and Figure 26 we observe that the autocorrelation of the residuals has been significantly reduced, the same of which is true for the cross-correlations shown in Figure 27. This suggests that this only slightly more complex model describes the underlying structure of the process much better than the linear regression model. A Ljung-Box test suggests that the residuals do not have significant autocorrelation left.

3.7 AR(2)-X(2,2) Model

Lastly, we are given a second order ARX model:

$$P_{h,t} = -\phi_1 P_{h,t-1} - \phi_2 P_{h,t-2} + \omega_{1,0} T_{\text{delta},t} + \omega_{1,1} T_{\text{delta},t-1} + \omega_{2,0} G_{v,t} + \omega_{2,1} G_{v,t-1} + \varepsilon_t \quad (42)$$

This yields the following estimations:

Figure 28: AR(2)-X(2,2) model with T_{delta} and G_v as exogenous variables.

Yielding the following error analysis:

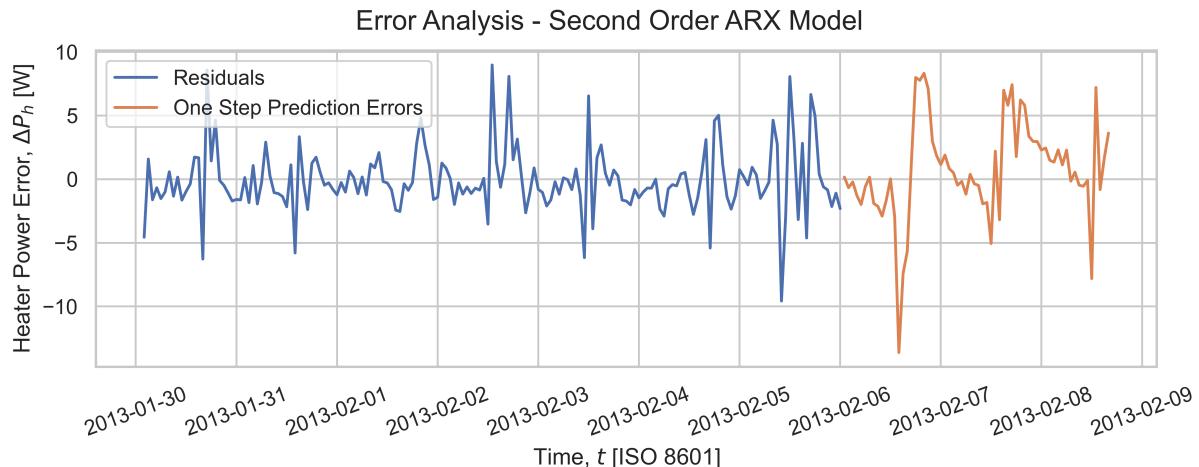


Figure 29: Residual analysis of the AR(2)-X(2,2) model.

With ACF and CCF analysis shown in Figure 30 and Figure 31 respectively:

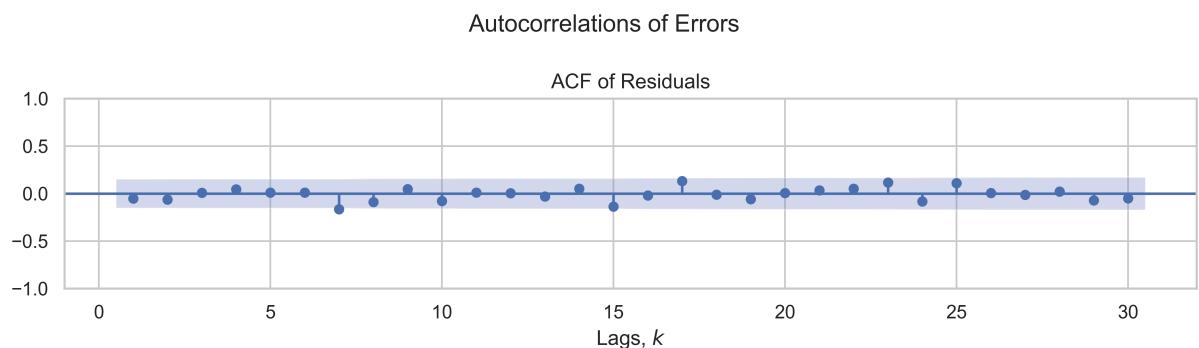


Figure 30: ACF of the residuals of the AR(2)-X(2,2) model.

Cross-Correlation of Errors and Exogenous Variables

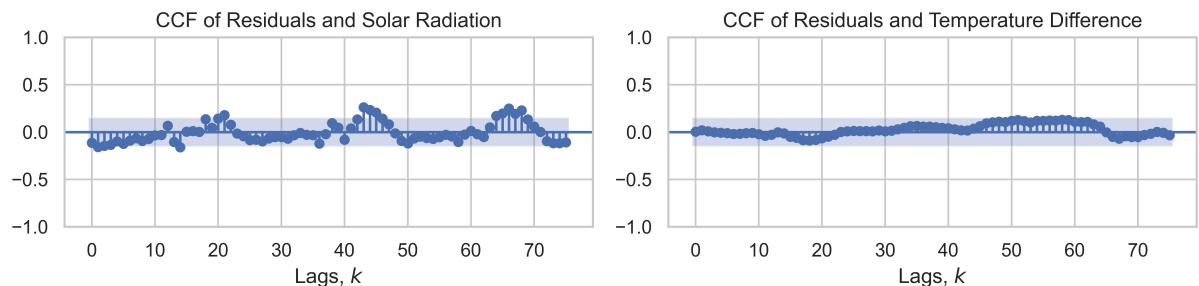


Figure 31: CCF of the residuals of the AR(2)-X(2,2) model with the exogenous variables.

We cannot directly observe an improvement when only looking at the ACF and CCF plots, and it may be difficult to understand whether the increased model complexity is justified. Again a Ljung-Box test suggests that the residuals do not have significant autocorrelation left.

3.7.a AIC and BIC

Now we want to delve a bit deeper into other metrics for model selection, specifically the BIC (Bayesian Information Criterion) and AIC (Akaike Information Criterion). Generally they are calculated as:

$$\begin{aligned} \text{AIC} &= -2 \log L(\mathbf{Y} | \hat{\boldsymbol{\psi}}) + 2p \\ \text{BIC} &= -2 \log L(\mathbf{Y} | \hat{\boldsymbol{\psi}}) + p \log n \end{aligned} \tag{43}$$

with the log-likelihood as:

$$\log L(\mathbf{Y} | \hat{\boldsymbol{\psi}}) = -\frac{n}{2} \left[\log(2\pi) + \log\left(\frac{(\mathbf{Y} - \hat{\mathbf{Y}})^2}{n}\right) + 1 \right] \quad (44)$$

for $\hat{\boldsymbol{\psi}}$ as the MLE estimated parameters, \mathbf{Y} as the modelled time-series, n as the number of observations and p as the number of parameters. In our case for the ARX model, since we only have an AR part, the MLE becomes an OLS estimation analogous to Eq. 34, hence $\hat{\boldsymbol{\psi}} = \hat{\Theta}$.

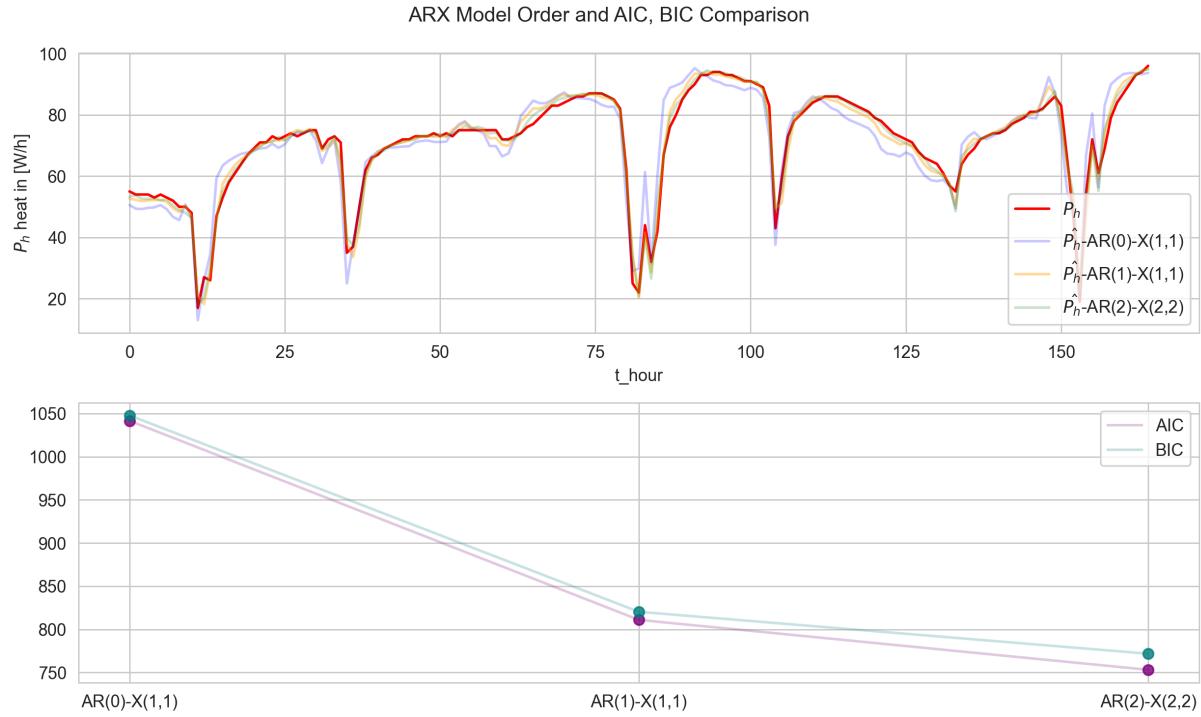


Figure 32: AIC, BIC comparison of different models orders from Section 3.5, Section 3.6, Section 3.7

This is a classic *elbow* curve. In various modelling fields, information theoretical measures or straight residual measures (such as MSE, etc.) are plotted against different model variations. The general idea is, to deduce at which point there is a good trade-off between parameters of the model and the model performance. In this case, we have the x -axis as increasing model order, hence increasing complexity and proneness to overfitting. Hence, we are looking for a trade-off between performance and model complexity. It is an interesting choice of metric to do this kind of plot. AIC and BIC inherently punish model complexity, as long as $\log(n) > 2$, the BIC does even more so; this also explains the slight difference between those two curves. Thus they theoretically already account for one of the decision parameter that such elbow-curve is intended to help with. If we look at the plot, there is a clear trend towards the more complex AR(2)-X(2,2) model. Traditionally, the decision here would be to select the AR(1)-X(1,1) model, as it already performs relatively well and the performance gain to the more complex model is very small.

However, as the AIC and BIC both already punish the increased model complexity of AR(2)-X(2,2) model, but still produce a lower score, which yields the interpretation that the increased performance is a worthwhile trade-off. Therefore, the model selection would fall onto the AR(2)-X(2,2) model.

3.8 RMSE Metric and Prediction on Test Data Set

We now use the on the training dataset estimated parameters $\hat{\Theta}$ and use those for one-step predictions on the test dataset with an accordingly constructed design matrix.



Figure 33: RMSE comparison of different models orders from Section 3.5, Section 3.6, Section 3.7 on concatenation of train and test dataset

We find in Figure 33 that using the RMSE as a metric for model selection yields the same result as the AIC and BIC metrics, which is a good sign that the model selection process is robust.

We already observed on the AIC and BIC on the training dataset that higher model order improves the predictive performance.

Thus, having validated via the RMSE on the test dataset that our model generalises well, we maintain that the AR(2)-X(2,2) model is the best trade-off between performance and complexity.

3.9 k -Step Predictions

We selected the AR(2)-X(2,2) model, mainly because of the AIC, BIC selection criteria. From experience, they produce a very reliable selection process. We provide a selection of k step-widths, which are usually choices that would make sense in real-life settings: 12h, 24h, 48h.

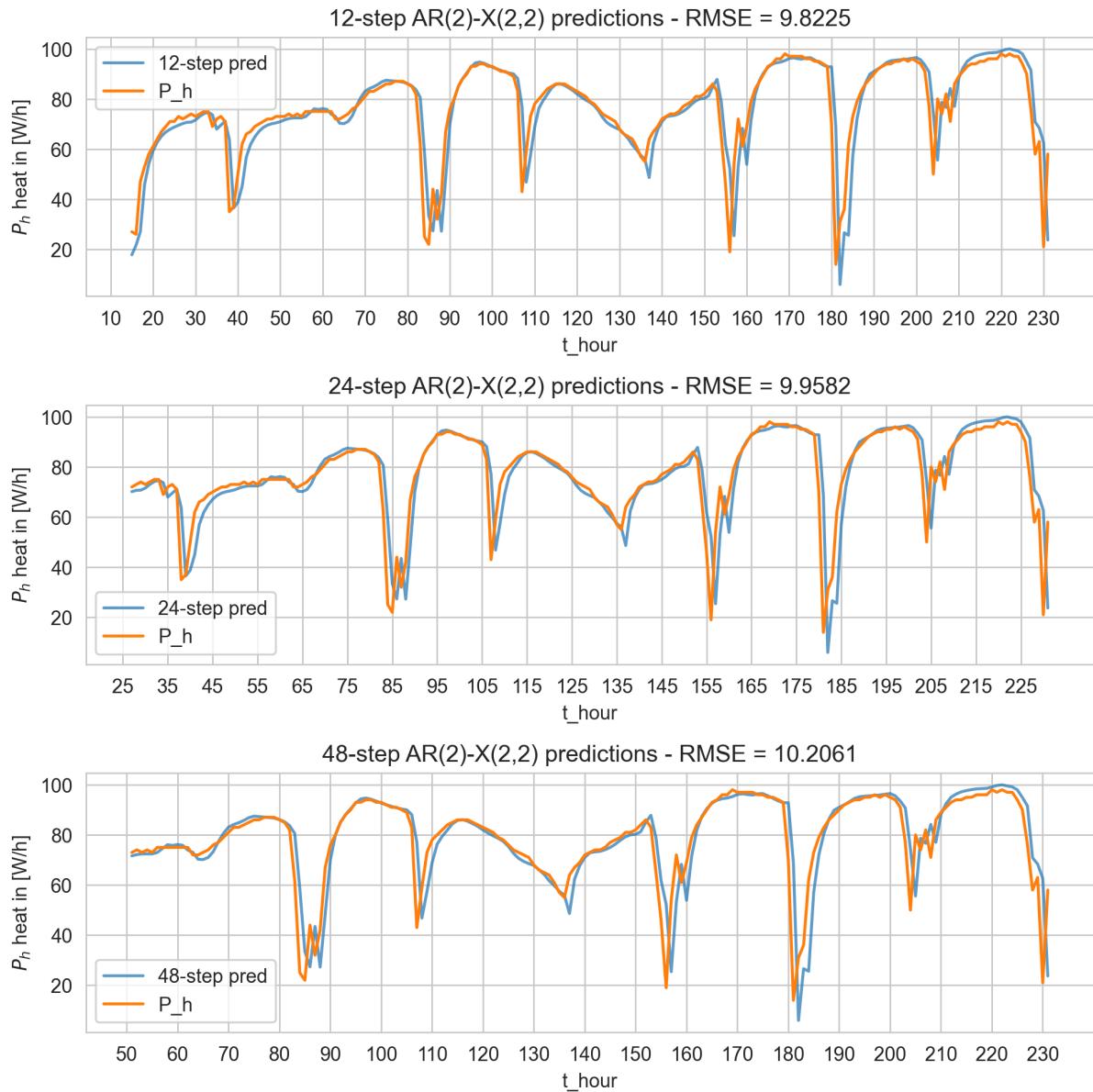


Figure 34: k-step predictions for intervals

Beyond that we look at a plot of the RMSE against the step-width k with the intend to draw some conclusions about the error behaviour over longer periods of prediction.

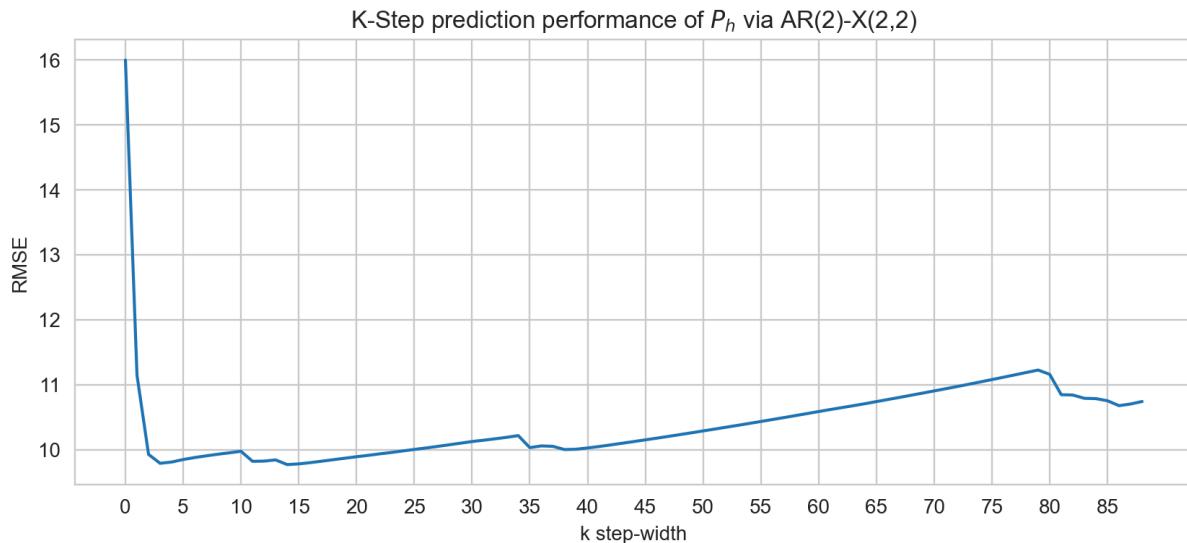


Figure 35: k-step predictions for intervals

As expected, the RMSE reacts heavily to the amount of steps into the future. Naturally, in the beginning there is a huge error in the predictions, as for k -step predictions we make predictions upon predicted values. For $k < 3$ there are not many original values of the target series included, resulting in predictions quickly drifting off.

P_h clearly shows seasonal drops in power (Figure 14). In some intervals however, there are exceptions, where the drop is not as sudden or not as steep. We assume this results in the small buckles in Figure 35. It is a step width, that goes in phase with the seasonal drops in P_h .

In context the most reasonable k-step intervals, as in Figure 34, yield reasonable RMSEs until $k \leq 40$, however, not quite as good as the one-step fits. Yet, it does provide the advantage of being able to forecast k -steps into the future.

One could try different starting points for the k-step predictions, instead of just going through the entire dataset. With each further prediction the uncertainty rises, since the newest predictions depend on the older predictions.

Overall, the model performs well; the residuals approach a white noise like distribution. There is still some visible auto-regressive behaviour present, but statistically the evidence is almost sufficient for independence of residuals. Trying a higher order could resolve that.

Apart from the prediction accuracy (which could potentially be improved), in an operational setting the model would work just fine with multi-step predictions. It is not too computationally expensive and would be easily integrated into a prediction pipeline. Naturally, the further into the future we predict, the higher the uncertainty; that notion is present without calculating the prediction intervals. Beyond that, we saw certain points of the series (towards the drops of P_h) where the predictions tend to deviate (Figure 35), the seasonality of the RMSE over k step-width. This does not improve if we simply choose a k step-width, that produces a good RMSE, since we would work on a continually predicting on estimated values. At some point the system will take a state (the power drops), where the predictions have shown to be weaker. Hence, by the nature of the series, this point of drop in P_h will naturally come, regardless of the k -step, so this cannot be avoided.

Modelling-wise an RLS model with higher “forgetting” coefficient may be a viable option.

Given that knowledge on the historic data, it would be wiser to simply accept an increased uncertainty in specific periods of the series and contextualize consequences: Reduce heating at the drops, but also have power available in case of slight mis-predictions.

It could also be an option to reduce the measurement interval from hourly to 5-min intervals, to be able to react quicker and counter prediction uncertainty.

3.10 Conclusions

We conclude based on AIC, BIC, and RMSE criteria that an AR(2)-X(2,2) model is the best trade-off between performance and complexity for the data set. More complex models may prove superior, but may require larger amounts of data to be able to estimate the parameters of the model accurately.

The model has a high predictive performance even for higher k -step forecasting and captures the underlying structure of the data well. The model is able to accurately predict power generation several hours or even days into the future. The model is stable and robust to unit shocks from the exogenous variables, which is a desirable property for a model of this type.

We find that simpler models without autoregressive components or transfer function features are not able to capture the underlying structure of the data and yield poor predictive performance.

It would be interesting to see where the tipping between predictive power and complexity lies. In the elbow plot Figure 32 we could already see a decreasing gradient.

3.11 Notes

We were unable to attach the code to the report for hand-in. The code is available upon request.

Bibliography

- [1] H. Madsen, *Time Series Analysis*. Chapman & Hall/CRC, 2008.