# Kalman Filter For Renewable Energy Systems

**Okan Coskun**
Student ID: 6369042
Computer Science Msc

## Abstract

Forecasting the output of renewable energy sources and being able to act accordingly plays an important role in todays life. Especially when it comes to integrating them into our energy grid and fully using their potential, it is neccessary to be accurate with the forecasts. Although predictions of such resources already exist, there is still room for improvement. Difficulties arrise mainly because of the unpredictable nature of weather and therefore renewable energy resources. This paper tackles this issue by using the Kalman Filter as a forecasting model for Photovoltaic and Wind energy, making hourly and 24 hour predictions using weather data. Although the hourly estimates perform worse than the daily ones, there are still lessons to be learned when using the Kalman Filter as a forecasting model.

## 1 Introduction

The usage of renewable energy sources becomes more important as greenhouse gases are increasing. It provides a cleaner alternative to fossil fuels (Ang et al., 2022) and reduces the dependence on finite resources. Accurate forecasting of renewable energy is essential to ensure a stable and reliable energy grid system, not only to better integrate such sources but also to provide it according to demand and need. However, as these outputs can be influenced by various factors such as the weather, humidity level and wind, it turns out that having reliable and stable forecasts are challenging.

Two of such renewable energy sources are wind and photovoltaic (PV). The focus of this paper is going to be how the Kalman Filter proposes a solution to the forecasting related issues by using weather forecasts and actual energy output measurements. Although there are more advanced frameworks and deep learning methods for time series forecasting, more basic methods just like the Kalman Filter can be very useful when setting up benchmarks. Another advantage of the filter is its simplicity, meaning that especially in times when transparency plays an important role, particularly in regard to the AI Act, it can be beneficial to have a method that is both promising and well-explainable.

In the beginning, the data and its preparation is going to be presented. After having a better understanding of the underlying structure, the Kalman filter will be introduced. After that, the obtained results will be presented and discussed. Finally, having all the insights and results, a short conclusion about the work and future improvements will be made.

## 2 Related Work

The Kalman Filter has existed for many years. A specific version of it was even used in the Apollo mission in the 1960s (McGee and Schmidt, n.d.). In context of renewable energy systems, the kalman filter is used for various aspects, like wind speed forecasting (Hur, 2021). Often, a modified version of it is used, whether it is the Extended Kalman Filter, or using different adapting filtering methods (Tian et al., 2014). One big challenge is finding the state space representation of a system. This will be discussed later in the Methodology section, however two different and interesting approaches are the N4SID algorithm (Van Overschee and De Moor, 1994), which is used by the Darts library

Seminar Machine Learning for Renewable Energy (Summer Term 2024), Tübingen.

(*Kalman Filter — darts documentation* n.d.), and a closed forumula solution like proposed in Wu et al. (2006), given the fact that we have actual measurements and true values which can be used in order to find the state space representation. Interestingly, the circumstances there have been very similar to this forecasting problem, although their main task was in regard to robotic arm movements, which has no connection to energy forecasting.

## 3 Methodology

The used data holds weather information for 80 different weather stations in Germany, as presented in Fig 1.

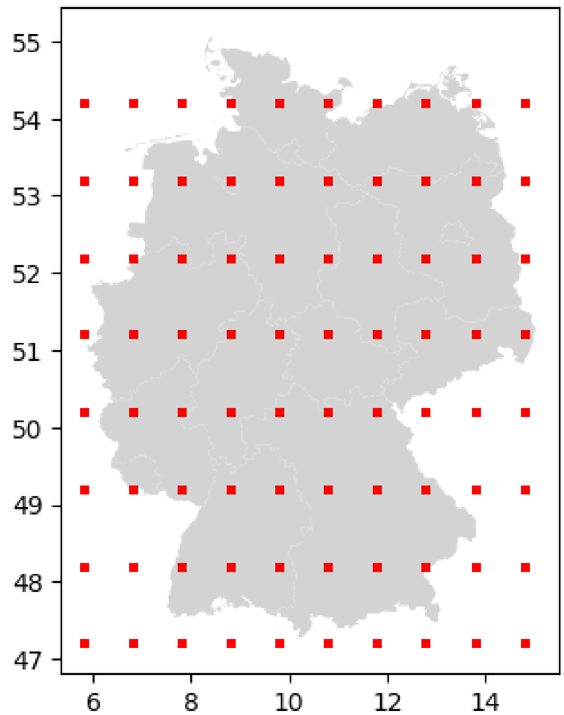


Figure 1: Weather stations in Germany.

The data covers a period from 2019 to 2022. There is an entry every 15 minutes. It contains more than 40 features and as the goal is forecasting wind and PV output, the focus lies on the following features:

Table 1: Selected Features for Forecasting

| Feature | Description |
|---|---|
| u100 | zonal wind component at 100m height |
| v100 | meridional wind component at 100m height |
| ssr | surface solar radiation |
| t2m | temperature in Kelvin at 2 meters above surface |

For the output of wind energy, windspeed obviously plays an important role. Through the parameters u100 and v100, the windspeed can be calculated as follows:

$$\text{windspeed} = \sqrt{u_{100}^2 + v_{100}^2} \tag{1}$$

Due to the fact that we have data from different locations at the same time, the average hourly and daily energy output for wind and photovoltaic is used, as the goal is later to forecast one and also 24 hours ahead.

In the following, the Kalman Filter and how it can be used within this scope will be further explained.

### 3.1 Kalman Filter

The Kalman Filter is a recursive algorithm, that is used for estimating the state of a system from a series of noisy measurements. The core of the Kalman Filter is the **state vector**, denoted as $\mathbf{x}_k$, which encapsulates all the information necessary to describe the system's current state at time step $k$. This is the true hidden state which we want to infere., in this case the energy output.

The **state transition model**, which is represented by the matrix $\mathbf{F}_k$, describes how the state evolves over time. This matrix is therefore responsible for predicting and is descibed as:

$$\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{w}_{\mathbf{K}} \tag{2}$$

where $\mathbf{w}_k \sim \mathcal{N}(0, \mathbf{W}_k)$ is the process noise , which accounts for uncertainties in the state transition.

Measurements (sometimes also referred as observations) of the system are incorporated into the filter using the **observation model**, described by the matrix $\mathbf{H}_k$. The relationship between the observed measurements and the state vector is given by:

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{q}_k \tag{3}$$

In this equation, $\mathbf{z}_k$ denotes the measurement vector, and $\mathbf{q}_k$ represents the observation noise, capturing the inaccuracies in the measurement process. As one can see, the true hidden state and measurements do not have to be the same, nor do they have to be the same dimensions.

The Kalman Filter operates in two main steps: the prediction step and the update step. During the **prediction step**, the filter estimates the state at the next time step, denoted as $\hat{\mathbf{x}}_{k+1|k}$, based on the current state estimate. The prediction equations are as follows:

$$\hat{\mathbf{x}}_{k+1|k} = \mathbf{F}_k \hat{\mathbf{x}}_{k|k} \tag{4}$$

$$\mathbf{P}_{k+1|k} = \mathbf{F}_k \mathbf{P}_{k|k} \mathbf{F}_k^\top + \mathbf{Q}_k \tag{5}$$

$\mathbf{P}_{k+1|k}$ is the predicted state covariance, which quantifies the uncertainty in the state estimate, and $\mathbf{Q}_k$ is the process noise covariance matrix.

Following the prediction step, the Kalman Filter refines the state estimate using the latest measurement in the **update step**. The updated state estimate is computed using the Kalman gain, $\mathbf{K}_k$, as follows:

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^\top (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^\top + \mathbf{R}_k)^{-1} \tag{6}$$

The updated state estimate and its covariance are then given by:

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}) \tag{7}$$

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1} \tag{8}$$

Here, $\mathbf{P}_{k|k}$ represents the updated state covariance, which incorporates the new information from the measurement, reducing the uncertainty in the state estimate.

A huge challenge is now how we define such a state transition for the environment, as this matrix is responsible for prediciting the energy output. In addition, we also have other parameters such as the variances, which must also be determined. One way to tackle this problem is going to be presented in the following chapter.

### 3.1.1 Finding State Space Representation

As mentioned earlier, Wu et al. (2006) has dealt with a similar situation where the parameters are unknown, but all the data is available. Having access to the actual values and with the assumption that the parameters $A_k$,$H_k$,$W_k$ and $Q_k$ are independent of time $t_k$, we can estimate the parameters by maximizing the joint probabilities $p(\mathbf{X}_M, \mathbf{Z}_M)$

Using the linear Gaussian properties of $p(\mathbf{X}_M)$ and $p(\mathbf{Z}_M|\mathbf{X}_M)$, the above maximization has the following closed-form solutions:

$$A = \left(\sum_{k=2}^{M} \mathbf{x}_k \mathbf{x}_{k-1}^T\right) \left(\sum_{k=2}^{M} \mathbf{x}_{k-1} \mathbf{x}_{k-1}^T\right)^{-1} \tag{9}$$

$$W = \frac{1}{M-1} \left(\sum_{k=2}^{M} \mathbf{x}_k \mathbf{x}_k^T - A \sum_{k=2}^{M} \mathbf{x}_{k-1} \mathbf{x}_k^T\right) \tag{10}$$

$$H = \left(\sum_{k=1}^{M} \mathbf{z}_k \mathbf{x}_k^T\right) \left(\sum_{k=1}^{M} \mathbf{x}_k \mathbf{x}_k^T\right)^{-1} \tag{11}$$

$$Q = \frac{1}{M} \left(\sum_{k=1}^{M} \mathbf{z}_k \mathbf{z}_k^T - H \sum_{k=1}^{M} \mathbf{z}_k \mathbf{x}_k^T\right) \tag{12}$$

These solutions allow us to estimate the matrices $A$, $H$, $W$, and $Q$ directly from the observed sequences.

### 3.2 Metric

The metric used in this paper is the $R^2$ score. It is a statistical measure that indicates how well the regression model's predictions approximate the real data points.

Mathematically, the $R^2$ score is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{13}$$

where $y_i$ represents the actual value of the dependent variable for the $i$-th observation, $\hat{y}_i$ denotes the corresponding predicted value and $\bar{y}$ is the mean of all actual values $y_i$.

## 4 Results

To obtain the above matrices, the data from the entire period is used, such that every seasonal pattern can be captured and learned. After that, for each datapoint in 2022, the goal is to estimate the energy output by only using the state transition matrix and the above mentioned features in Table 1.

### 4.1 Daily prediction

For the 24 hour forecasting for photovoltaic, the model captures the seasonality. It is clear to see in Fig 4 that in the warmer months like june and july, where there is more sunlight, the time series has a peak. Furthermore, it is also clear to see that in the more colder months, where sunlight is more rare, the output becomes less aswell. However, even in winter there are obviously days when the sun comes out, resulting in these peaks in the original data that are barely captured by the model.

Interestingly, for predicting wind power, the filter still manages to model the basic structure well, even though wind is less seasonal than solar.

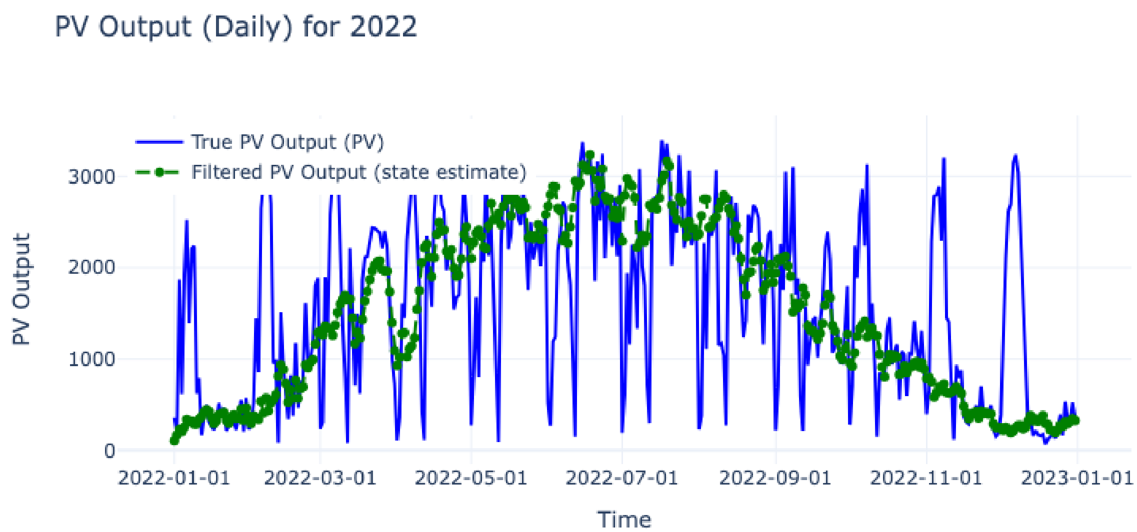


Figure 2: PV output daily.

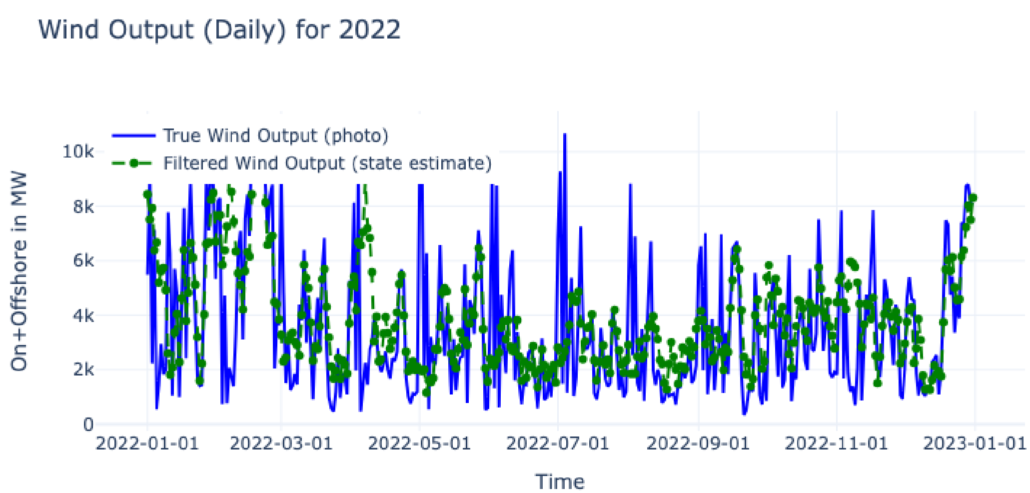


Figure 3: Wind output daily.

In Fig 3 can be seen that the original data shows less of a seasonal pattern. Even when just looking at the raw ssr values, a clear bell shaped curve can be seen. Still, the model manages to capture the rising and falling of the energy output to a good extent.

### 4.2 Hourly Predictions

Here, it can be seen that the hourly prediction for solar output capture the true values quite well. This is mainly due to the fact that in this case an additional feature has been chosen, namely the 24-hour energy output value before, so a lagged version has been added. This is due to the fact that otherwise, if we use the same features as for the other predictions, the model is hardly able to capture anything, resulting in an $R^2$ value that is close to 0. Therefore, this additional feature has been included, which improved the model by far.

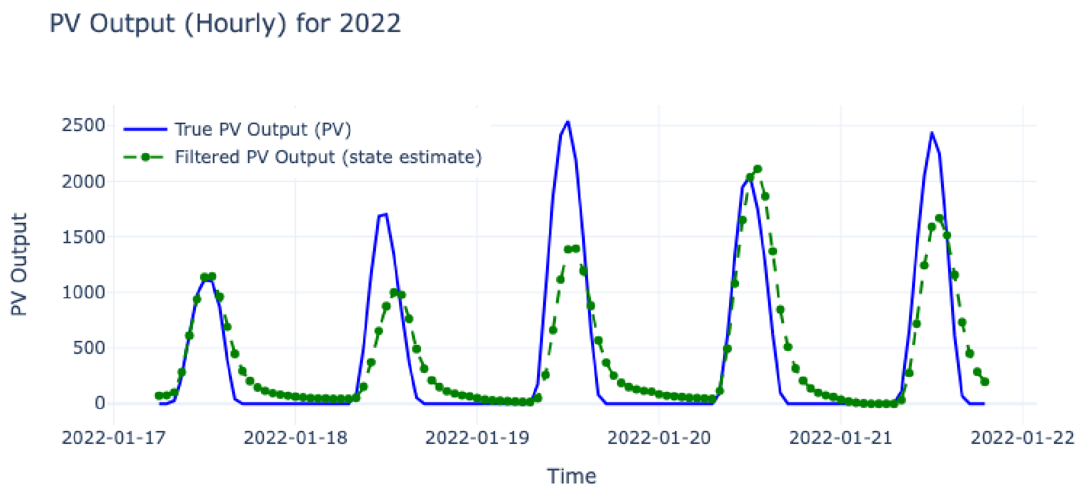


Figure 4: PV output hourly with **lagged** energy feature.

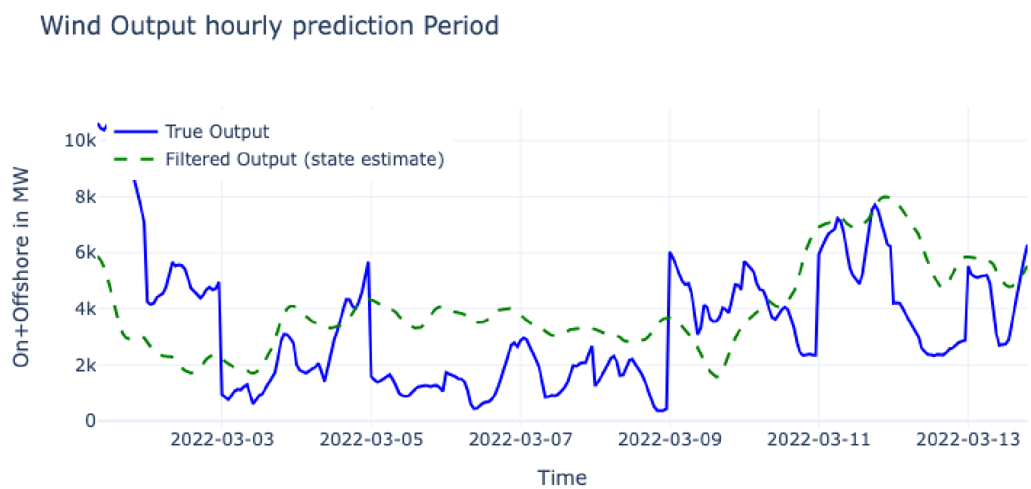


Figure 5: Wind output hourly.

For wind, such a modification is not neccessary, as the model is able to capture the slopes without any information about the output before. The estimate tends to fall and decay with the original output, however capturing the real values or having overlaps is very rare, resulting in a worse prediction than the daily output.

Overall, with the exception of the hourly PV forecast, the model is able to follow along the course of the energy output based on the features listed in Table 1. The intention at the beginning has been such that the model is able to find the relationship between the features listed above and the energy output without incorporating any energy output related value.

The R2 scores for each model are:

Table 2: R² Scores for Different Forecasting Models

| Forecast | PV Daily | Wind Daily | PV Hourly (with **lagged** Input) | Wind Hourly |
|---|---|---|---|---|
| R² Score | 0.17 | 0.27 | 0.65 | 0.10 |

It is clear to see that the model with the additional energy output information performs very well. However, for the daily forecasts, especially wind, the model is also able to fit the curve well without such modification, resulting in acceptable R2 scores. The hourly prediction for wind is quite low, so you can't really say that the model fits the series well here.

## 5 Discussion

Despite having success with the daily forecast, the model fails to capture relevant parts for the hourly forecasts. This can have various reasons, one might be that the with higher granularity level, the model fails to exploit temporal correlations. Daily data smooths out short therm fluctuations and anomalies, which are more present in hourly data. Another important aspect to keep in mind is that we assume linearity. The less granular we look at this data, the smoother it gets and the more this assumptions works in our benefits. Trying to use a linear model that captures non linear events in shorter time intervals might lead to more inaccurate predictions. As mentioned in the earlier chapter, an extended kalman filter for the hourly prediction could have been one improvement which may enhance the model.

The solution of adding an output value that lies in the past is also a valid choice because it is easy to use in practice. If you want to predict the future, you can either take the closest value or the value from a week, a month, a year, etc., so that you can always find a reference value that acts as a feature. Of course, the further into the future you want to predict, the less accurate the prediction itself becomes, because the reference value is so far in the past.

Last but not least, it is important to mention that the derivation of the state transition matrix is crucial here, as its the only part that is responsible for prediction (and the added noise factor). A lot of time has been spent on trying to find a derivation for these matrices, as their is no accurate physical model that neither describes the relationship between solar radiation and photovoltaic nor wind speed and the wind energy output as it was provided in the dataset. There are formulas to that, however most of the time they also incorporate more information and have slightly different energy outputs, hence it has not been applicable for this usecase.

## 6 Conclusion

The Kalman Filter is used to forecast photovoltaic and wind energy supply in Germany. Moreover, it has been shown how the derivation of the matrices can be calculated when having access to the entire dataset. In addition, the results have been presented and discussed by evaluating the forecast but also highlighting the limitations of the used approach.

Due to its simple nature, we are able to understand why the results have potentially come about in this way and they thus provide a good baseline on which to build on, at least for the daily and the modified hourly forecast. For the future, more variables and features could be brought up into the prediction phase, making the matrices more complex but also more likely to be capable to model the series better. Furthermore, they are various algorithms for identifying a state space system. They could help finding a better fit for the matrices and therefore also enhance the forecasts and provide more reliable predictions that could help maintain and support the energy systems.

## References


Ang, T.-Z., M. Salem, M. Kamarol, H. S. Das, M. A. Nazari, and N. Prabaharan (2022). "A comprehensive study of renewable energy sources: Classifications, challenges and suggestions". In: *Energy Strategy Reviews* 43, p. 100939. DOI: `https://doi.org/10.1016/j.esr.2022.100939`.

Hur, S. ho (2021). "Short-term wind speed prediction using Extended Kalman filter and machine learning". In: *Energy Reports* 7, pp. 1046–1054. DOI: `https://doi.org/10.1016/j.egyr.2020.12.020`.

*Kalman Filter — darts documentation* (n.d.).

McGee, L. A. and S. F. Schmidt (n.d.). "Discovery of the Kalman Filter as a Practical Tool for Aerospace and". en. In: ().

Tian, Y., Q. Liu, Z. Hu, and Y. Liao (2014). "Wind speed forecasting based on Time series - Adaptive Kalman filtering algorithm". In: *2014 IEEE Far East Forum on Nondestructive Evaluation/Testing*, pp. 315–319. DOI: `10.1109/FENDT.2014.6928287`.

Van Overschee, P. and B. De Moor (1994). "N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems". In: *Automatica* 30.1. Special issue on statistical signal processing and control, pp. 75–93. DOI: `https://doi.org/10.1016/0005-1098(94)90230-5`.

Wu, W., Y. Gao, E. Bienenstock, J. P. Donoghue, and M. J. Black (Jan. 2006). "Bayesian Population Decoding of Motor Cortical Activity Using a Kalman Filter". en. In: *Neural Computation* 18.1, pp. 80–118. DOI: `10.1162/089976606774841585`.