

A Guide to Solar Power Forecasting using ARMA models

Bismark Singh

Discrete Mathematics & Optimization, Sandia National Laboratories

David Pozo

Center for Energy Systems, Skolkovo Institute of Science and Technology

Abstract

We describe a simple and succinct methodology to develop hourly auto-regressive moving average (ARMA) models to forecast power output from a photovoltaic generator. We illustrate how to use statistical tests to validate the model and construct hourly samples. These samples can be used as scenarios for energy planning or in stochastic optimization models.

Keywords: ARMA, solar power, photovoltaic, forecasting, scenario generation

1. Introduction

Increasing penetration of renewable energy sources, such as wind and solar, in the electricity grid requires good day-ahead power forecasts. Solar power differs from wind power due to its diurnal nature, and can have much greater ramps than wind [1]. In this article, we focus on forecasting
5 hourly solar power generation, in particular from photovoltaic (PV) technology.

Forecasting methods for solar power are broadly divided into two categories: (i) physics-based models—these models predict solar power from numerical weather predictions and solar irradiation data, and (ii) statistical models—these models forecast solar power directly from historical data. Comparisons of these two methods are also available; see, e.g., [2, 3]. There are other approaches
10 available as well which combine these two methods [4]. In this article we center on statistical methods alone, and specifically the use of auto-regressive moving average (ARMA) models to develop our forecasts. Despite their limitations [5], ARMA models are widely used to forecast wind power [6, 7], as well as solar power [8, 9], because of their ease of implementation and parameter selection. Yet, accurate and fast methods to generate solar power scenarios are often unavailable
15 or significantly complex, and normal approximations are frequently used; see, e.g., [10]. Here, we

describe a summary of the methodology to forecast solar power using ARMA models. The software codes and generated scenarios are available on request. The presented models can be applied either to a local PV generating plant or at the regional level.

The main contribution of this article is to provide a step-by-step approach and easy-to-implement ARMA model to forecast PV solar power generation. The proposed model is able to capture the important statistical features of the parameters, while maintaining simplicity. The model allows modelers to embed it into more complex decision-making structures, statisticians to have an all-in-one place ARMA model design for PV power generation, and policy makers and electrical engineers to have a scenario generation tool.

2. Methods

We take hourly year-long historical solar power output from a site described in [11]. This zone is at an altitude of 595m, has a nominal power of 1560 MW, a panel tilt of 36° , and a 38° clockwise panel orientation from the north. Further installation specifics are available in zone 1 from Table 1 of [11], while technical specifications are available in [12]. We use approximately nine months of data for training. The data does not have any solar power for the ten hours [20:00-5:00], and hence we restrict the forecasts in these hours to be zero as well. Equivalently, a criteria based on the solar zenith angle can be used; i.e., 0° at sunrise and 90° when the sun is directly overhead. For each of the remaining 14 hours of the day, we build an $\text{ARMA}(p, q)$ model. One limitation of this approach is that it does not capture the correlations between different hours; see, e.g., [13]. Further, for each hour, we verify the stationarity of the time series and test a number of $\text{ARMA}(p, q)$ models to find the best one. We use statistical tests on the residuals to validate the models. Finally, we use Monte Carlo sampling from the best ARMA model, for each hour, to create hourly scenarios. Below we provide more details.

2.1. Stationarity

An ARMA model may be suitable if a time-series is stationary. We test the hourly data for stationarity using the Augmented Dickey-Fuller (ADF) test [14]. The ADF test has a null hypothesis that the series includes a unit root (or, is non-stationary). We reject the null hypothesis at a level 0.05 if the test-statistic exceeds its 0.95 level quantile. For all the 14 hours of the day, the null

hypothesis is rejected suggesting the series may be stationary, and hence an ARMA model may be
45 suitable. If the series were not stationary, an ARIMA model may be suitable; see, e.g., [15].

2.2. Selecting parameters of the ARMA model

Next, we estimate the parameters of the ARMA model, p , the order of the autoregressive part
and, q , the order of the moving average part. For each hour, we construct 16 models with both
 p and q between one and four, and compute the log-likelihood objective function value. Next, for
50 each hour, we calculate the Bayesian information criteria (BIC) for the 16 models using $p + q + 1$
parameters. The BIC penalizes for models with more parameters, and the smallest value of the
BIC gives the best model, for each hour. Table 1 provides our estimated p and q values for the 14
hours of the day. We note that none of the hours have an order value exceeding two.

Table 1: Estimated p and q values for ARMA(p, q) models for 14 hours of the day

Hour	6:00	7:00	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00
p	1	1	1	1	2	1	1	1	1	1	1	1	1	2
q	1	1	1	1	1	1	2	1	1	1	1	1	2	1

2.3. Prediction

Figure 1 plots a day-ahead prediction using the above constructed ARMA models; i.e., one hour
55 ahead predictions from the 14 ARMA models. A number of metrics are available to evaluate the
prediction; see, e.g., [16]. We use a few of them here. The mean absolute error between the actual
and the predicted series is 39.6 MW, or 3.3% of the maximum actual value. The root mean square
error between the actual and the predicted series is 61.0 MW, or 5.1% of the maximum actual value.

60 We further verify autocorrelation in the series, for each hour, using the Ljung-Box test [17] on
the residuals for lags of 5, 10, and 15. The Ljung-Box test has a null hypothesis that the residuals
are uncorrelated up to a given lag. We reject the null hypothesis at a level 0.05 if its test-statistic
exceeds its 0.95 level quantile. For all the 14 hours of the day, the null hypothesis is not rejected
suggesting a zero autocorrelation in the series, or the model choice may be appropriate.

65 With increasing penetration of solar power in the electricity grid, a number of stochastic op-
timization models for bidding, storage, and generation have been developed; see, e.g., [18, 19].
Stochastic optimization models rely on the availability of a large number of scenarios. We can use
Monte Carlo sampling to generate hourly solar power scenarios. The output from an ARMA model

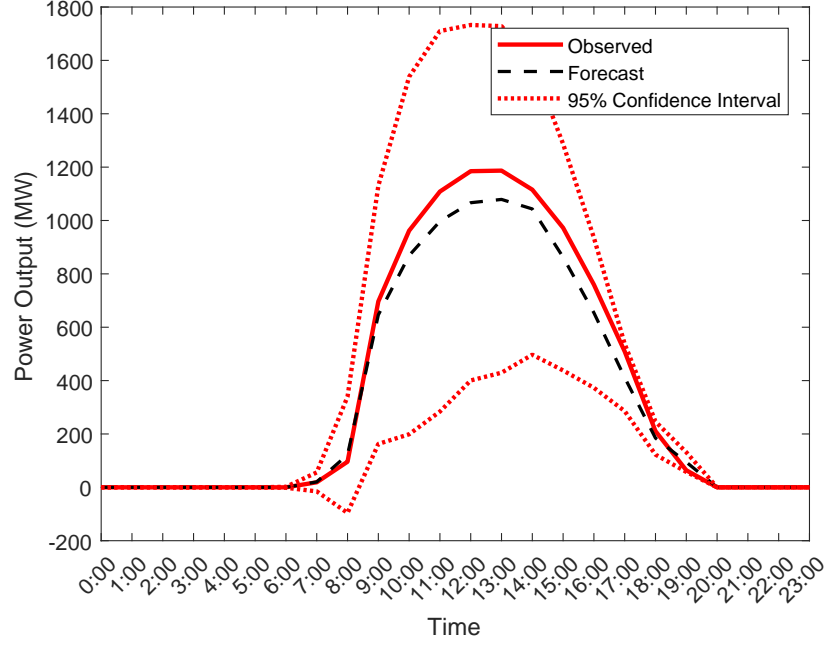


Figure 1: Day-ahead actual and predicted values using ARMA models from Table 1

is real valued, and hence can be negative. In our analysis, we truncate the negative powered outputs
70 to 0. For the 14 hours of the day, this sampling resulted in 1.6% of the outputs with estimated
power output below -5MW. Figure 2 plots 2000 day-ahead scenarios as well as the median and 10
percentile values.

3. Conclusions

We present a simple step-by-step scheme for fitting an ARMA model to historical solar power
75 data, and use it to forecast future hourly scenarios. We present statistical tests to check the
applicability of various models, identify model parameters, and to finally forecast scenarios. If
significant statistical evidence is not present in support of a test, the chosen model is not expected
to perform well. This can lead to erroneous conclusions; see, e.g., [20, 21]. The methodology in
this article can be directly applied to historical data, both for a single PV source and a PV site, to
80 create future scenarios for use in a stochastic model for power system operation and planning.

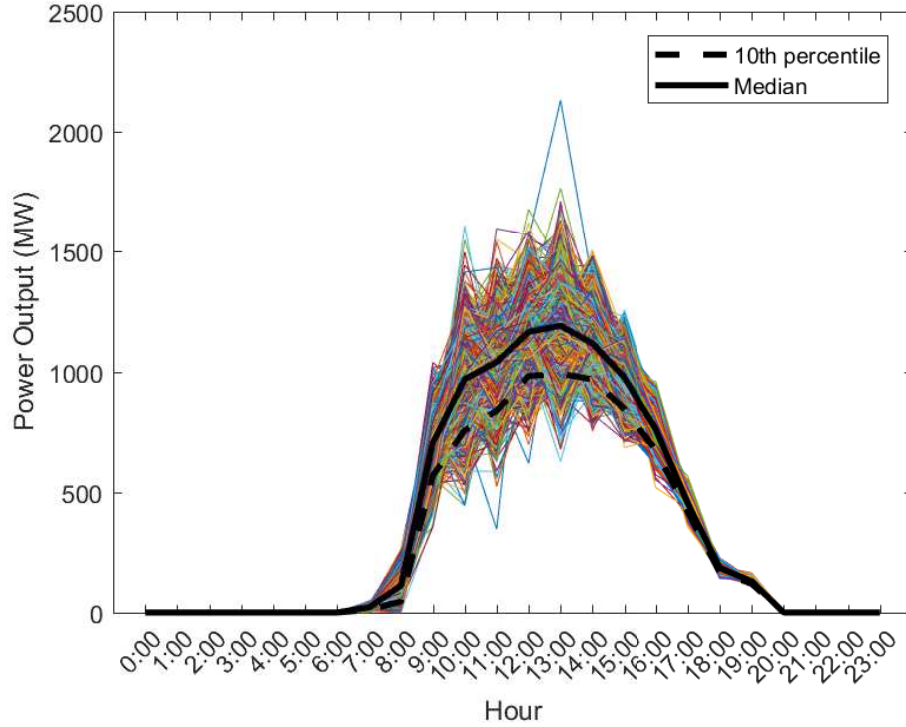


Figure 2: 2000 hourly scenarios for solar power generated using the ARMA models from Table 1. The dashed black line is the median hourly value, and the solid black line is the 10 percentile solar power value.

Acknowledgments

B. Singh thanks Jean-Paul Watson and Andrea Staid at Sandia National Laboratories for helpful discussions and for sharing data. Sandia National Laboratories is a multitechnology laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. The work of D. Pozo was supported by Skoltech NGP Program (Skoltech-MIT joint project).

References

- [1] I. Graabak, M. Korpås, Variability characteristics of European wind and solar power resources—A review, *Energies* 9 (6) (2016) 449.

- [2] Y. Huang, J. Lu, C. Liu, X. Xu, W. Wang, X. Zhou, Comparative study of power forecasting methods for PV stations, in: Power System Technology (POWERCON), 2010 International Conference on, IEEE, 2010, pp. 1–6.
- [3] R. H. Inman, H. T. Pedro, C. F. Coimbra, Solar forecasting methods for renewable energy integration, Progress in Energy and Combustion Science 39 (6) (2013) 535–576.
- [4] Z. Dong, D. Yang, T. Reindl, W. M. Walsh, A novel hybrid approach based on self-organizing maps, support vector regression and particle swarm optimization to forecast solar irradiance, Energy 82 (2015) 570–577.
- [5] D. Yang, J. Kleissl, C. A. Gueymard, H. T. Pedro, C. F. Coimbra, History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining, Solar Energy.
- [6] B. G. Brown, R. W. Katz, A. H. Murphy, Time series models to simulate and forecast wind speed and wind power, Journal of Climate and Applied Meteorology 23 (8) (1984) 1184–1195.
- [7] M. J. Duran, D. Cros, J. Riquelme, Short-term wind power forecast based on ARX models, Journal of Energy Engineering 133 (3) (2007) 172–180.
- [8] L. Mora-Lopez, M. Sidrach-de Cardona, Multiplicative ARMA models to generate hourly series of global irradiation, Solar Energy 63 (5) (1998) 283–291.
- [9] D. Yang, P. Jirutitijaroen, W. M. Walsh, Hourly solar irradiance time series forecasting using cloud cover index, Solar Energy 86 (12) (2012) 3531–3543.
- [10] W. Su, J. Wang, J. Roh, Stochastic energy scheduling in microgrids with intermittent renewable energy resources, IEEE Transactions on Smart Grid 5 (4) (2014) 1876–1883.
- [11] F. Golestaneh, H. B. Gooi, P. Pinson, Generation and evaluation of space–time trajectories of photovoltaic power, Applied Energy 176 (2016) 80–91.
- [12] Solarfun Power, SF 160-24-M, [https://www.energymatters.com.au/images/solarfun/SF160-24-M\(190W\).pdf](https://www.energymatters.com.au/images/solarfun/SF160-24-M(190W).pdf), Accessed: March 20, 2018 (2009).
- [13] H. T. Pedro, C. F. Coimbra, Assessment of forecasting techniques for solar power production with no exogenous inputs, Solar Energy 86 (7) (2012) 2017–2028.

- [14] D. A. Dickey, W. A. Fuller, Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association* 74 (366a) (1979) 427–431.
- 120 [15] J. Contreras, R. Espinola, F. J. Nogales, A. J. Conejo, ARIMA models to predict next-day electricity prices, *IEEE transactions on power systems* 18 (3) (2003) 1014–1020.
- [16] C. F. Coimbra, J. Kleissl, R. Marquez, Overview of solar-forecasting methods and a metric for accuracy evaluation, *Solar Energy Forecasting and Resource Assessment* (2013) 171–194.
- [17] G. M. Ljung, G. E. Box, On a measure of lack of fit in time series models, *Biometrika* 65 (2) 125 (1978) 297–303.
- [18] R. Banos, F. Manzano-Agugliaro, F. Montoya, C. Gil, A. Alcayde, J. Gómez, Optimization methods applied to renewable and sustainable energy: A review, *Renewable and Sustainable Energy Reviews* 15 (4) (2011) 1753–1766.
- [19] N. Sharma, Stochastic techniques used for optimization in solar systems: A review, *Renewable and Sustainable Energy Reviews* 16 (3) (2012) 1399–1411. 130
- [20] D. Kwiatkowski, P. C. Phillips, P. Schmidt, Y. Shin, Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?, *Journal of Econometrics* 54 (1-3) (1992) 159–178.
- [21] P. C. Phillips, P. Perron, Testing for a unit root in time series regression, *Biometrika* 75 (2) 135 (1988) 335–346.
- [22] P. Lauret, E. Lorenz, M. David, Solar forecasting in a challenging insular context, *Atmosphere* 7 (2) (2016) 18.

Appendix

We compare the developed model against two other forecasting methods. First, we fit a single ARMA model, as opposed to hour-by-hour, on the entire time series. Second, we use the Smart-Persistence model of [22], which assumes the solar power at hour t is the mean of the previous h hours. We use $h = 2$ here. Table 2 compares the three models; the proposed model performs better in both of the chosen metrics. 140

Table 2: Comparison of three forecasting models. MAE denotes the mean absolute error and RMSE denotes the root mean squared error.

	Model		
	Hourly ARMA	Single ARMA	Smart-Pers
MAE (MW)	39.6	48.2	45.7
RMSE (MW)	61	112.51	102.6