1. Adrian PONIATOWSKI, 2. Jerzy BARANOWSKI

DOI: 10.15199/48.2025.04.13

Optimizing home energy balance with Bayesian models: predicting solar panel efficiency in variable weather conditions

Optymalizacja bilansu energetycznego domu za pomocą modeli bayesowskich: Prognozowanie efektywności paneli słonecznych w zmiennych warunkach pogodowych

Abstract: *Photovoltaic (PV) technology is revolutionizing renewable energy by providing a sustainable and inexhaustible source of electricity. This paper explores Bayesian models for predicting home energy balance when using solar panels, with a focus on optimizing energy production under varying weather conditions in Poland. The research emphasizes the importance of personalized energy management solutions due to the high variability in household consumption patterns. Two distinct Bayesian models are developed to predict energy production and consumption, leveraging factors such as temperature, insulation, cloudiness, and day length. Data collected over five months from a specific region in Poland was used to validate these models. The results demonstrate the models' effectiveness in capturing the variability in energy production and provide insights into optimizing the efficiency of PV systems. This study offers valuable guidelines for individuals considering investments in photovoltaic panels, highlighting potential profitability and efficiency improvements.*

Streszczenie: *Technologia fotowoltaiczna (PV) rewolucjonizuje odnawialne źródła energii, zapewniając zrównoważone i niewyczerpane źródło energii elektrycznej. W niniejszym artykule zbadano modele bayesowskie służące do przewidywania bilansu energetycznego w domu podczas korzystania z paneli słonecznych, ze szczególnym uwzględnieniem optymalizacji produkcji energii w zmiennych warunkach pogodowych w Polsce. Badania podkreślają znaczenie spersonalizowanych rozwiązań w zakresie zarządzania energią ze względu na dużą zmienność wzorców zużycia energii w gospodarstwach domowych. Opracowano dwa odrębne modele bayesowskie w celu przewidywania produkcji i zużycia energii, wykorzystując takie czynniki jak temperatura, izolacja, zachmurzenie i długość dnia. Dane zebrane w ciągu pięciu miesięcy z określonego regionu w Polsce zostały wykorzystane do walidacji tych modeli. Wyniki pokazują skuteczność modeli w uchwyceniu zmienności produkcji energii i dostarczają spostrzeżeń na temat optymalizacji wydajności systemów PV. Niniejsze badanie oferuje cenne wskazówki dla osób rozważających inwestycje w panele fotowoltaiczne, podkreślając potencjalną rentowność i poprawę wydajności.*

Keywords: PV, Bayesian models, energy balance, prediction
Słowa kluczowe: PV, modele Bayesowskie, równowaga energetyczna, predykcja

Introduction

Photovoltaic (PV) technology is transforming the landscape of renewable energy, offering a clean, inexhaustible source of electricity. Solar energy systems provide substantial benefits such as reducing carbon footprints, enhancing energy security, and mitigating the volatility of energy prices.

The adoption of PV installations enables households to stabilize their energy costs and generate their own electricity, leading to significant long-term savings [1]. In recent years, solar panels have been gaining increasing popularity, not only as photovoltaic farms but also among individual consumers investing in this type of energy production. A crucial aspect turns out to be the profitability of installing panels under the weather conditions prevailing in different areas of the world. The predictability of energy use in homes presents a complex challenge due to the high variability in consumption patterns across different households and types. Lachut, Banerjee, and Rollins [2] emphasize the necessity for personalized approaches in energy management systems. Their study highlights that generalized models are often inadequate due to significant variances in energy consumption behavior, which can be influenced by factors such as household size, occupancy patterns, and appliance usage. This variability underscores the need for tailored energy management solutions that can adapt to the unique characteristics of individual households.

Nfaoui and El-Hami [3] delve into methods for extracting maximum energy from solar panels, focusing particularly on the importance of calculating the optimal angle for solar panels. Their research addresses the need to consider various environmental and physical factors to optimize energy production. The study underscores the significance of precise angle adjustments to ensure maximum solar irradiance capture, which is critical for enhancing the overall efficiency of PV systems.

El Hammoumi et al. [4] provide a comprehensive overview of PV energy, from material selection to practical applications. Their work discusses advanced techniques such as solar trackers and floating solar panels, which are designed to maximize power output. This paper offers valuable insights into the latest innovations in PV technology aimed at optimizing energy production under various conditions, emphasizing the continuous evolution of methods to enhance PV system efficiency.

Sampaio and González [5] offer a broad conceptual framework for understanding photovoltaic solar energy. Their highly cited work covers fundamental aspects of PV systems, including the principles of solar energy conversion and the various technologies employed. This foundational text serves as a critical resource for understanding the basic mechanisms and theoretical underpinnings of PV systems, providing a solid base for further exploration of optimization strategies. Fouad, Shihata, and Morgan [6] conduct an integrated review of factors affecting the performance of PV panels. They identify and analyze various parameters such as temperature, irradiance, and panel orientation that significantly impact energy production. This comprehensive review provides a detailed understanding of how these factors interact and influence the overall efficiency of PV systems, offering valuable guidelines for improving panel performance.

Mattei et al. [7] propose a method for calculating the temperature of polycrystalline PV modules using an Energy balance approach. Their work, which has been widely cited, investigates the impact of meteorological parameters on PV module temperature and, consequently, on energy production. This study provides a practical method for estimating PV module performance under varying environmental conditions, contributing to more accurate predictions and better system designs.

Despite the extensive research, several limitations remain. Most studies emphasize technical optimization and

theoretical models without sufficiently addressing the practical integration of energy production with real-world consumption patterns. Additionally, while predictive models using various techniques have been proposed, there is a notable lack of research employing Bayesian models specifically for home energy balance predictions in the context of PV systems.

Bayesian models provide a powerful framework for managing uncertainty and incorporating prior knowledge, making them particularly suitable for predicting home energy balance. These models can synthesize diverse data sources, including historical energy production and consumption, weather forecasts, and other relevant variables, to deliver more precise and reliable predictions [8]. Despite their potential advantages, Bayesian methods remain underrepresented in studies focused on residential PV systems.

Our research aims to fill these gaps by employing Bayesian models to forecast home energy balance for households equipped with solar panels in the Lublin Voivodeship of Poland. We collected comprehensive datasets, including daily energy production and consumption metrics from a residential PV installation, along with detailed local weather data. By integrating these datasets, we developed Bayesian predictive models that offer practical insights for households considering PV investments. This study not only advances the theoretical understanding of Bayesian applications in energy predictions but also provides actionable guidance for homeowners. The developed models enable potential investors to make informed decisions about the economic feasibility and effectiveness of PV installations tailored to their specific regional conditions. This localized approach ensures that our findings are relevant and beneficial for households in the Lublin Voivodeship and other regions with similar climatic characteristics.

Case study

Photovoltaics is an inexhaustible source of energy that doesn’t pollute the environment, is durable, and makes you independent from annual electricity price increases. This means that investing in a photovoltaic installation not only allows you to stabilize your energy expenses by generating your own electricity but also generates savings for decades.

However, the production of electrical energy is influenced by many factors, so we decided to analyze their individual impact. Additionally, to make it more than just purely statistical data that may not mean much to people unfamiliar with the topic, we added data and energy consumption predictions for the same household for which the energy is being produced using solar panels.

The whole problem initially involved gathering real data, which was possible thanks to having photovoltaic panels and a specialized application. The application showed how many kilowatt-hours were produced each day and how much was consumed by the household. Additionally, in specialized weather services, we found information regarding the parameters that have the greatest impact (according to sources) on solar energy production. In the case of consumption, these parameters are certainly not perfectly chosen, but as mentioned earlier, the main idea of the project was to predict production, while consumption is used to better illustrate the values.

The created models and obtained results can help people considering photovoltaic panels in making decisions regarding investments. The project provides the opportunity to obtain at least partial answers to whether such an investment can be profitable under the conditions prevailing in Poland. It is worth noting that the data comes from the Lublin Voivodeship, and there may be slight differences between regions within the country.

Dependencies between factors are visible on DAG diagram (see Fig. 1) for our project.

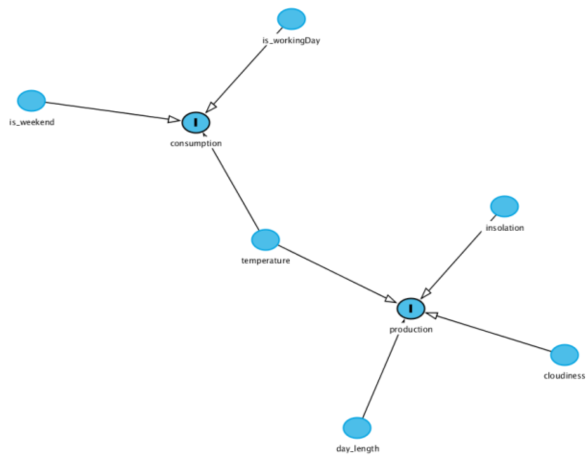

Fig. 1. Directed Acyclic Graph (DAG) showing the relationships between factors influencing the prediction of energy production and consumption

When it comes to confounders, the only case we have in our project, although difficult to categorize unambiguously (whether a fork or something), is the relationship between the variables "is_weekend" and "is_workingDay." Specifically, if one variable has a value of 1, the other is 0, and vice versa.

These variables convey information about the type of day of the week.

Another confounder that initially appeared in our project and can be seen in the DAG is "temperature". The type of this confounder is a "fork", which corresponds to omitted variable bias. Due to the high variability and inaccuracy of the data, we decided to split one model that predicted both energy consumption and production into two separate models, thus avoiding the problem.

Methods

Before using any methods, it is essential to have correct data tailored to the problem, thus data preprocessing was necessary. The data was obtained using the manufacturer’s photovoltaic panel application. The application tracked the production and consumption of electrical energy over a fivemonth period, from January to June, allowing for the consideration of different weather conditions in the data. Each row represents data from a single day and includes the following information:

- Day of the week
- Consumption (*kWh*)
- Production (*kWh*)
- Average temperature ($^{\circ}C$)
- Average insolation (W/m^2)
- Average cloudiness (*%*)
- Day length (*h*)

The collected data is fully complete (no missing values). Additionally, it did not require many preprocessing steps. The only necessary step was encoding the day of the week. Two new dummy columns were created: "Weekend" and "Working_week," each containing binary information indicating whether it is currently a weekend or not.

The next step in data processing was to divide them into separate DataFrames, where each DataFrame contains the data needed for a specific model.

The last step involved data normalization, as the data contained values with different units. Normalization was performed by subtracting the mean and dividing by the standard deviation.

Bayesian statistics

Bayesian statistics is an approach to data analysis based on Bayes' theorem, where available knowledge about parameters in a statistical model is updated with the information in observed data. The background knowledge is expressed as a prior distribution and combined with observational data in the form of a likelihood function to determine the posterior distribution. The posterior can also be used for making predictions about future events. [9]

One of the key advantages of Bayesian Statistics is its flexibility and robustness in dealing with complex models and small sample sizes. By incorporating prior information, Bayesian methods can yield more accurate and stable estimates, even when data are scarce. This is particularly useful in fields such as medicine, where prior clinical knowledge can be critical, or in environmental science, where observational data might be limited.

Moreover, Bayesian approaches are inherently probabilistic, providing a complete distribution of possible parameter values rather than single-point estimates. This allows researchers to quantify uncertainty and make probabilistic statements about the parameters, which is often more informative for decision-making processes. The computational advancements in recent decades, particularly in Markov Chain Monte Carlo (MCMC) methods and variational inference, have significantly enhanced the feasibility of Bayesian methods, enabling the analysis of high-dimensional data and complex hierarchical models.

Used models

Before proceeding to create the models, we decided to analyze the collected data.

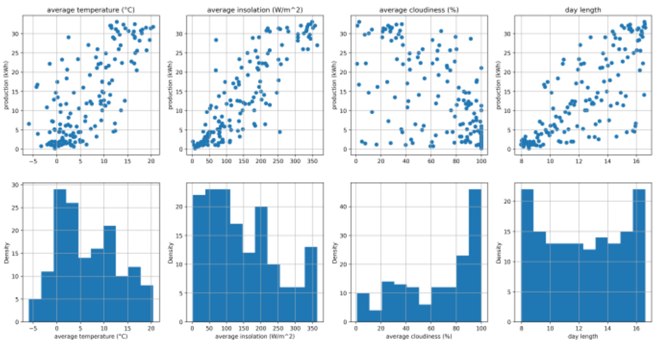

Fig. 2. Scatter plots and histograms illustrating the distribution of individual factors affecting energy production. The scatter plots show points depicting the relationships between production and the factor, while the histograms present the empirical distribution.

The charts (Fig. 2 and Fig. 3) allowed us to observe certain linear relationships between predictors and production as well as consumption. Based on these observations, we deemed a linear model potentially suitable for this problem.

The first model is a linear model with the mean calculated based on beta coefficients, actual predictor values, and the intercept alpha. Then, a normal distribution is used with the obtained mean and standard deviation in the form of a value drawn from an exponential distribution. The first production model includes the following predictors: insolation, cloudiness, temperature, and day length. These predictors were selected based on articles related to photovoltaics found on the internet. Model can be described as follows (1) - (8):

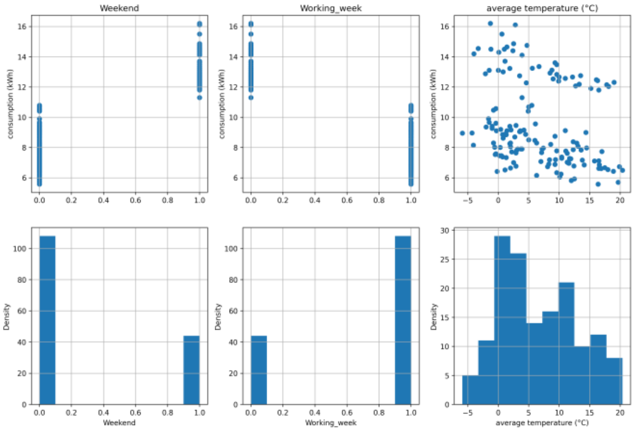

Fig. 3. Scatter plots and histograms illustrating the distribution of individual factors affecting energy consumption. The scatter plots show points depicting the relationships between consumption and the factor, while the histograms present the empirical distribution.

(1) $\alpha \sim normal(16, 5)$
(2) $\beta_1 \sim normal(3, 0.3)$
(3) $\beta_2 \sim normal(5, 0.5)$
(4) $\beta_3 \sim normal(-1.75, 0.3)$
(5) $\beta_4 \sim normal(3, 0.4)$
(6) $\sigma \sim exp(15)$
(7) $\mu_i = \alpha + X_{i1} * \beta_1 + X_{i2} * \beta_2 + X_{i3} * \beta_3 + X_{i4} * \beta_4$
(8) $y_i \sim normal(\mu_i, \sigma)$,

where α is the intercept, X is the matrix with input data (temperature, insolation, cloudiness and day length), β is the vector containing factor coefficients and σ is the standard deviation.

The second model, similar to the first one, differs in the number of predictors used, specifically considering insolation, cloudiness, and day length but not temperature. Due to that, equation (2) isn't included and equation (7) is changing to (9).

(9) $\mu_i = \alpha + X_{i2} * \beta_2 + X_{i3} * \beta_3 + X_{i4} * \beta_4$

Based on observations, we deemed it worthwhile to check if temperature has an impact on energy production. Its equation is the same as (8), but the difference here is only in the vector β, because it is smaller and does not include temperature as a factor.

In an effort to present the overall energy data in a more accessible way and having data that allows us to somewhat predict energy consumption, we decided to create an experimental energy consumption model, which is described by equations (10) - (15) and (8).

(10) $\alpha \sim normal(11, 3)$
(11) $\beta_1 \sim normal(1, 0.2)$
(12) $\beta_2 \sim normal(-1, 0.2)$
(13) $\beta_3 \sim normal(-1, 0.2)$
(14) $\sigma \sim exp(0.75)$
(15) $\mu_i = \alpha + X_{i1} * \beta_1 + X_{i2} * \beta_2 + X_{i3} * \beta_3$

This time matrixX contains other input data (weekend, working_week, temperature). First and second columns are dummies, relationship is shown at Fig. 1.

Expectations for this model were not high since the only available data that could be utilized were the day of the week and temperature. During data collection, we noticed that significantly higher consumption occurs on weekends (likely due to a higher presence of people in the household and cleaning activities often performed on those days). Additionally, energy consumption is higher at lower temperatures (heating requirements).

The selected priors for the models are the parameters of slope in the linear function - beta, the intercept - alpha, and a prior for the standard deviation - sigma. For the coefficients of the linear function, a normal distribution was chosen as the prior because we do not have any knowledge suggesting the use of a different distribution. As for the standard deviation parameter - sigma, we assumed an exponential distribution for it.

For all models we selected specific mean and standard deviation values for the priors of the beta, alpha, and sigma parameters. These values are based on our current knowledge of the discussed domain and determine the impact of each predictor on the final outcome.

STAN

STAN is a platform for statistical modeling and highperformance statistical computations. Thousands of users utilize it for statistical modeling, data analysis, and forecasting in the social, biological, and physical sciences, as well as in engineering and business.

Stan interfaces with the most popular data analysis languages (R, Python, shell, MATLAB, Julia, Stata) and runs on all major platforms (Linux, Mac, Windows).

Users specify log density functions in the Stan probabilistic programming language and then fit the models to data using:

- full Bayesian statistical inference with MCMC sampling: NUTS-HMC
- approximate Bayesian inference with variational inference: Pathfinder and ADVI
- penalized maximum likelihood estimation with optimization

Results

The actions taken to obtain the results are identical for each of the models used and are consistent with Bayesian statistics. We began generating results by compiling models based on our expert knowledge in the discussed topic, known as Prior Predictive Check. If positive feedback was received, meaning the quality of the obtained data was sufficient to describe real data, the next step was performed. The analysis referred to as posterior analysis is based on the available data and prior knowledge, and as a result, the parameters regarding the influence of individual factors on energy production from photovoltaic panels are updated and modified according to the model.

Model results

Model 1

According to the assumptions, the first step is to generate data from the prior model.

From the Fig. 4, it can be inferred that the chosen prior values allow for generating values that describe the real data.

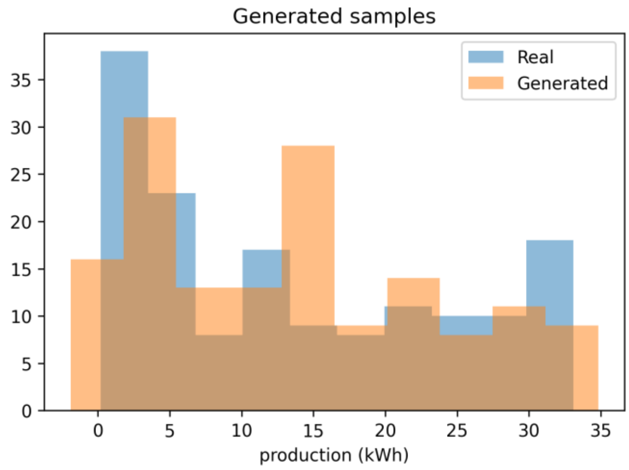

Fig. 4. Chart for the first model comparing histograms of real data and data generated from the prior model for energy production. The model shows high consistency, the histograms are similar, and the structure of the actual data has been reproduced quite good.

The samples generated from the prior in the model make sense in terms of the distribution values. To confirm this, we created scatter plots to better examine the generated values for each factor.

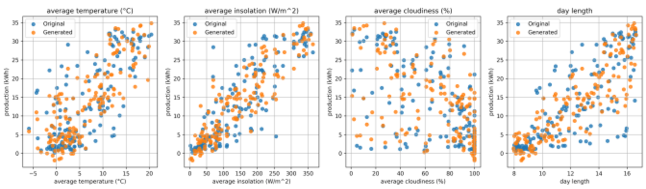

Fig. 5. Scatter plots for the first model comparing real data and data generated from the prior model separately for each factor affecting the amount of energy produced from solar panels. The values of both types are consistent, especially in terms of data distribution, which is the most important aspect.

Based on Fig. 4 and Fig. 5, we can state with great confidence that the chosen first prior model is correct due to the distribution of values. Therefore, using this model along with actual data, we can proceed to posterior analysis.

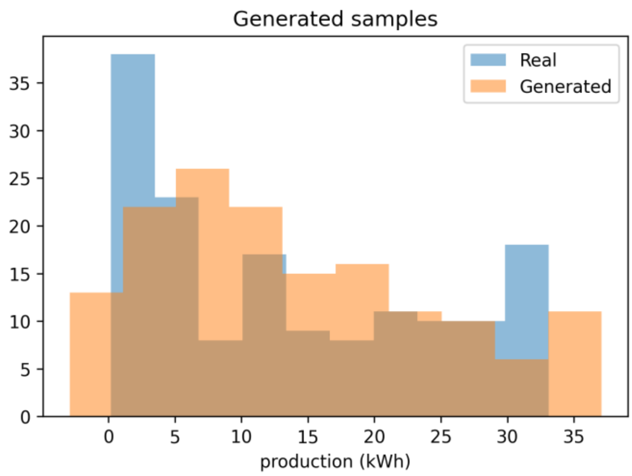

Fig. 6. Chart for the first model comparing histograms of real data and data generated from the posterior model for energy production. In terms of achieved values, the model is quite consistent; however, the distribution of generated data slightly differs from the irregular, hard-to-replicate distribution of real-world data.

From the histogram (Fig. 6), it can be concluded that the posterior model is able to describe the observed data. However, the model is not perfect as it may generate values that are too large or negative, which is physically unrealistic. The model did not have any constraints applied, which allowed for smooth sampling, but it affected the results (generated data). It is important to note that the model's ability to describe the data should be interpreted with caution, considering the limitations and potential unrealistic values it may generate.

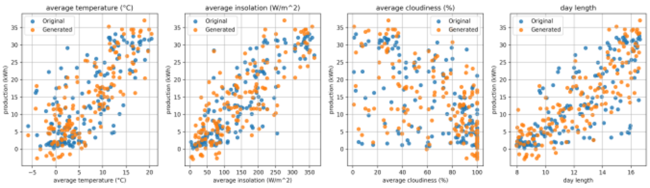

Fig. 7. Scatter plots for the first model comparing real data and data generated from the posterior model separately for each factor affecting the amount of energy produced from solar panels. The values of both types are consistent, especially in terms of data distribution, which is the most important aspect. Compared to priors, the charts also show values that deviate more from the distribution, known as outliers.

From the scatter plots (Fig. 7), we can observe that the model is capable of generating realistic samples. In comparison to the prior model, it is able to generate values that deviate more from the norm. This suggests that the posterior model has learned from the data and can generate samples that capture the variability and outliers present in the observed data.

The histograms (Fig. 8) show the distribution of values generated by the posterior model for α as the intercept and for the individual β components, illustrating their impact on the amount of energy produced.

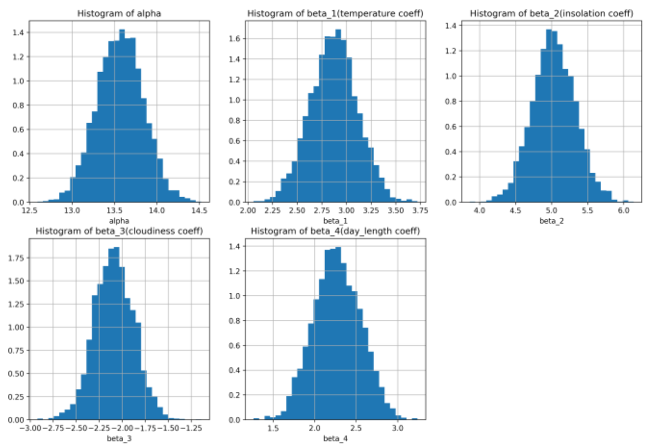

Fig. 8. Histograms for the first model showing the distribution of individual coefficients of the equation 8. The histograms created from the generated points confirm that the posterior distributions align with the initial assumptions. Posterior model parameter values are slightly shifted in contrast to our prior assumptions, so we can conclude that model learned on input data and corrected its parameter values.

Model 2

The steps performed are identical to Model 1.

Similar to the first model, from the Fig. 9, it can be inferred that the chosen prior values also allow for generating values that describe the real data. Once again, to confirm this, we created scatter plots to better examine the generated values for each factor.

Based this time on Fig. 9 and Fig. 10, we can state with great confidence that the chosen second prior model also is correct due to the distribution of values. Therefore, using this model along with actual data, we can proceed to posterior analysis.

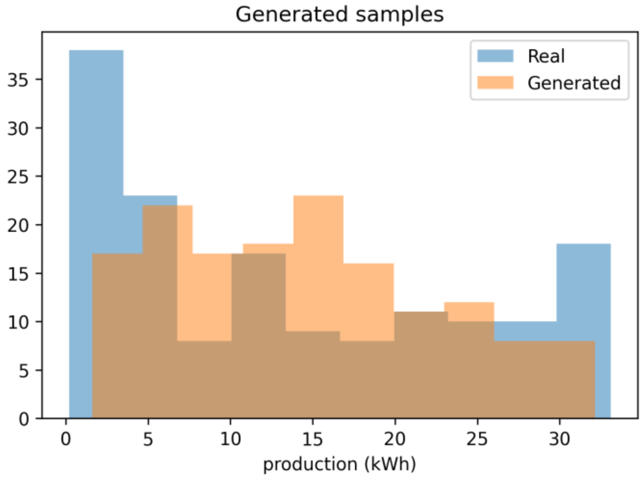

Fig. 9. Chart for the second model comparing histograms of real data and data generated from the prior model for energy production. The data distribution allows us to consider the model as correct; however,it reflects real-world data less accurately than the prior of model 1.

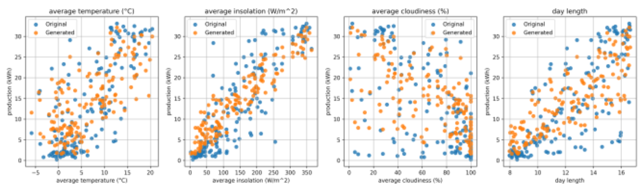

Fig. 10. Scatter plots for the second model comparing real data and data generated from the prior model separately for each factor affecting the amount of energy produced from solar panels. The values of both types are consistent, especially in terms of data distribution, which is the most important aspect.

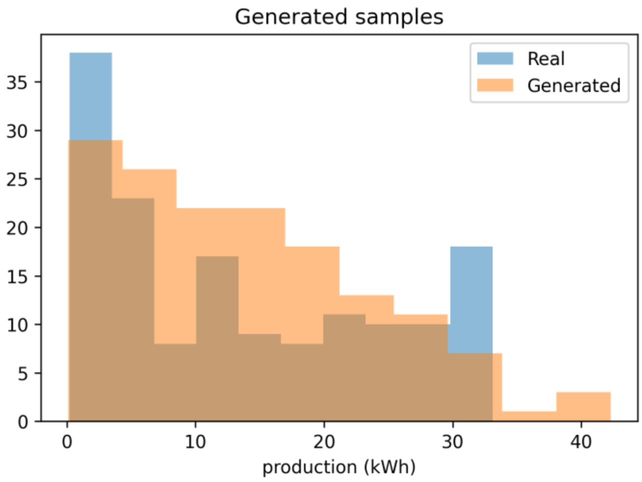

Fig. 11. Chart for the second model comparing histograms of real data and data generated from the posterior model for energy production. The range of achieved values generated is larger (due to maximum values) than in the case of real-world data.

Model 2, in comparison to model 1, still uses normal distributions. However, as seen in the plot (Fig. 11), the minimum values reached are close to zero. This is due to the application of the fmax function in the generated quantities. Unfortunately, there are also noticeable values that are larger than the actual data, which may suggest that the temperature used in model 1 had an possitive influence on the prediction.

This model generates data quite well, and it can be said that the values deviating from the norm are even better sampled than in the previous model. However, there are also many samples with values larger than expected, which is visible on Fig. 12.

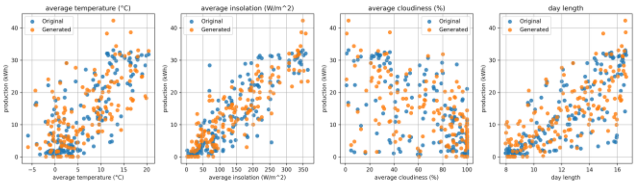

Fig. 12. Scatter plots for the second model comparing real data and data generated from the posterior model separately for each factor affecting the amount of energy produced from solar panels. The values of both types are consistent, especially in terms of data distribution, which is the most important aspect. Compared to priors, the charts also show values that deviate more from the distribution, known as outliers.

While the model captures the variability and outliers in the data, the issue of generating values larger than expected remains a concern. It is important to further evaluate and refine the model to address this discrepancy and ensure that the generated samples align more closely with the expected values.

The histograms (Fig. 13) show the distribution of values generated by the posterior model for α as the intercept and for the individual β components, illustrating their impact on the amount of energy produced.

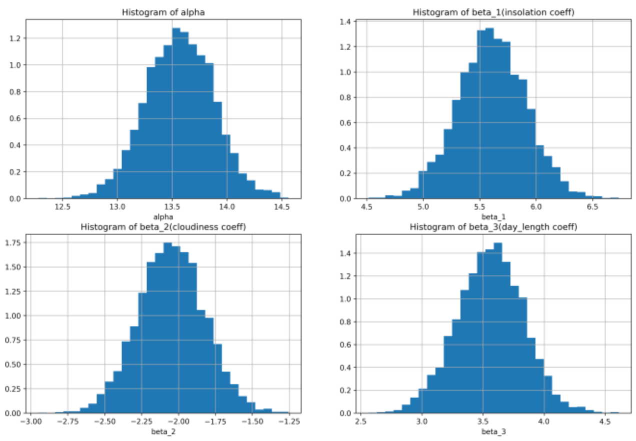

Fig. 13. Histograms for the second model showing the distribution of individual coefficients of the equation 8. The histograms created from the generated points confirm that the posterior distributions align with the initial assumptions. Posterior model parameter values are slightly shifted in contrast to our prior assumptions, so we can conclude that model learned on input data and corrected its parameter values.

Experimental model

As described in section , we also decided to prepare an experimental model with low expected quality, but allowing everyone to get acquainted with the approximate scale between energy consumed and produced. Despite the different model, the actions performed will be the same as in previous models.

From the Fig. 14, it can be inferred that the chosen prior values allow for generating values that describe the real data. The samples generated from the prior in the model perhaps do not reflect the data perfectly, but make sense in terms of the distribution values. To confirm this, we created scatter plots to better examine the generated values for each factor.

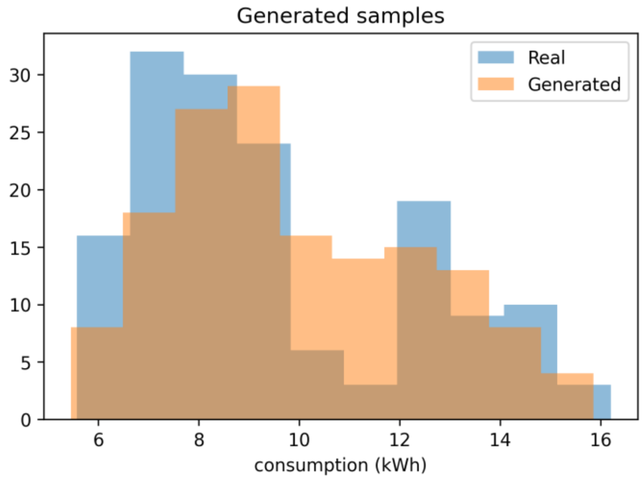

Fig. 14. Chart for the experimental model comparing histograms of real data and data generated from the prior model for energy consumption. The model shows quite a high degree of consistency, and the histograms are similar; however, the bimodal structure of the actual data is not particularly visible.

Based on Fig. 14 and Fig. 15, we can state with confidence that the chosen experimental prior model is quite correct due to the distribution of values. Therefore, using this model along with actual data, we can proceed to posterior analysis.

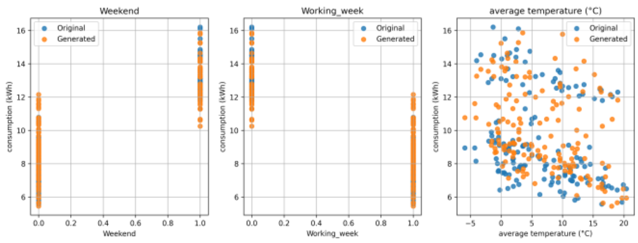

Fig. 15. Scatter plots for the experimental model comparing real data and data generated from the prior model separately for each factor affecting the amount of energy consumed. The values of both types are consistent, especially in terms of data distribution, which is the most important aspect.

From the histogram (Fig. 16), it can be concluded that, to our surprise, generated samples are clearly consistent with real data, so posterior model describes the data better than expected.

From the scatter plots (Fig. 17), again we can see that model learned on the data and describes it pretty well. Additionally there is clear improvement over model based only on our prior assumptions.

The histograms (Fig. 18) created from the generated points confirm that the posterior distributions align with the initial assumptions.

Against our expectations, this model behaves quite good despite having some primitive predictors.

Model comparison

To compare the models we will use LOO and WAIC criteria. Both methods were utilized from the arviz library using the ’compare’ method. To familiarize yourself with this library and the meaning of each column visible in the DataFrames below, please see [13].

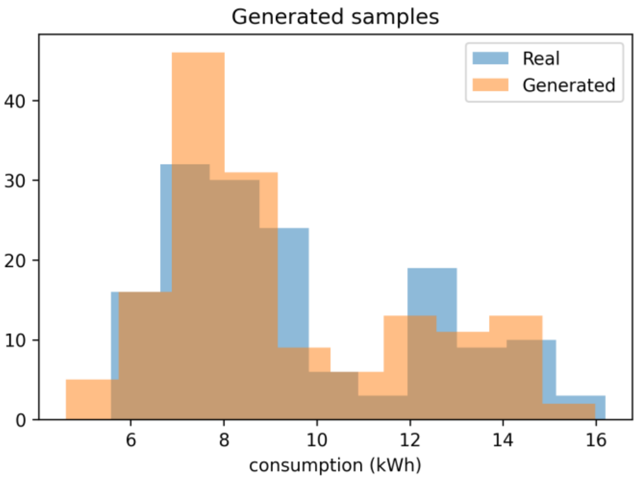

Fig. 16. Chart for the experimental model comparing histograms of real data and data generated from the posterior model for energy consumption. The model shows high consistency, the histograms are similar, and in comparison to the prior, the bimodal structure of the actual data has been reproduced.

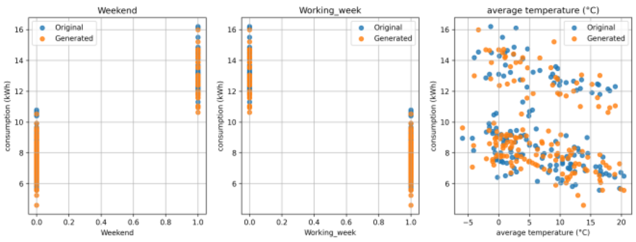

Fig. 17. Scatter plots for the experimental model comparing real data and data generated from the posterior model separately for each factor affecting the amount of energy consumed. The values of both types are consistent, especially in terms of data distribution, which is the most important aspect. Compared to priors, the charts also show values that deviate more from the distribution, known as outliers.

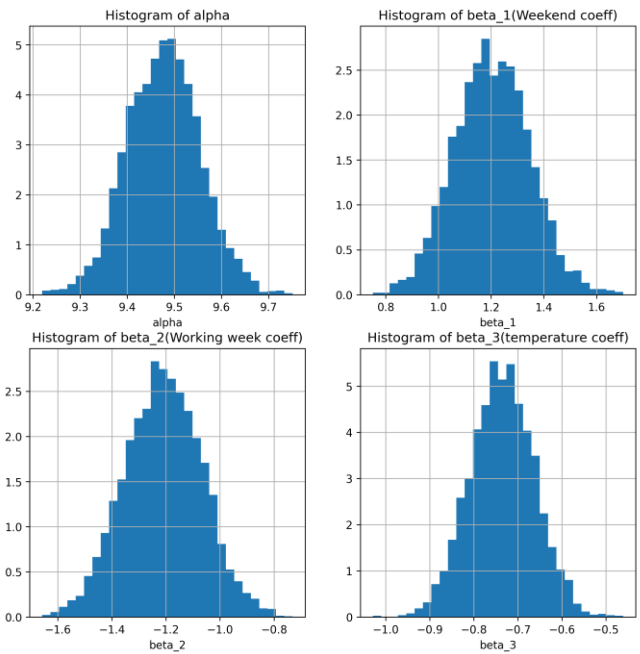

Fig. 18. Histograms for the experimental model showing the distribution of individual coefficients of the equation 8. The histograms created from the generated points confirm that the posterior distributions align with the initial assumptions. Posterior model parameter values are slightly shifted in contrast to our prior assumptions, so we can conclude that model learned on input data and corrected its parameter values.

LOO Compare

Computes Pareto-smoothed importance sampling leave-one-out cross-validation (PSIS-LOO-CV).

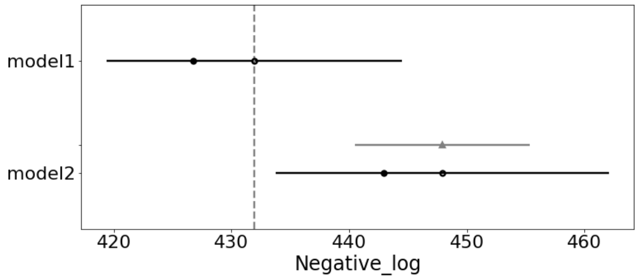

Fig. 19. A plot on a negatively logarithmic scale illustrating the comparison of models using the LOO criterion. According to [13], for this scale, smaller values indicate higher out-of-sample predictive fit ("better" model).

Estimates the expected log pointwise predictive density (elpd) using Pareto-smoothed importance sampling leave-one-out cross-validation (PSIS-LOO-CV) [10] [11].

	rank	loo	p_loo	d_loo	weight	se	dse
model1	0	431.95	5.22	0.00	0.83	12.57	0.00
model2	1	447.92	4.96	15.97	0.17	14.15	7.45

WAIC Compare

Computes the widely applicable information criterion.

Estimates the expected log pointwise predictive density (elpd) using WAIC. [10] [12]

	rank	waic	p_waic	d_waic	weight	se	dse
model1	0	431.94	5.22	0.00	0.83	12.57	0.00
model2	1	447.91	4.95	15.96	0.17	14.15	7.46

Taking into account the individual results and charts for Model 1 and Model 2, as well as considering the parameters for comparing these models, we can conclude that Model 1, which includes more predictors (including temperature), is better and more accurately reflects the real data. Thus, we agree with the ranking of models presented by the LOO and WAIC criteria.

We believe that the first model is superior because temperature does have an impact on predicting data related to the production of energy from photovoltaic panels. Our hypothesis was whether the temperatures reached in Poland are high enough to influence the panel's efficiency. As the comparison of the models has shown, the hypothesis turned out to be true.

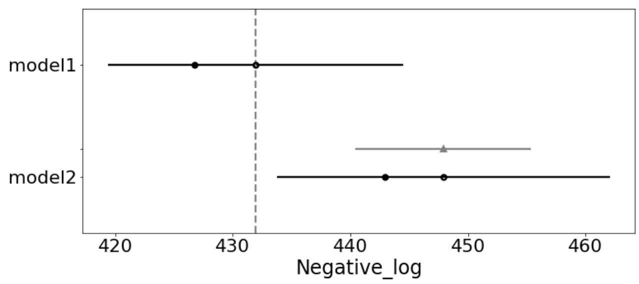

Fig. 20. A plot on a negatively logarithmic scale illustrating the comparison of models using the WAIC criterion. According to [13], for this scale, smaller values indicate higher out-of-sample predictive fit ("better" model).

Discussion & Conclusions

To sum up, both models reasonably capture the problem of energy production from photovoltaic panels. However, the main issue lies in the generation of negative values in the better-performing model. This issue can potentially be addressed by using different distributions or applying constraints to the parameters.

We created plots (Fig. 21, Fig. 22, Fig. 23) of generated values for specific input data. We plotted histograms of values for one day in every month of our data, so we could see the influence of predictors on output. As we can see generated data is overall consistent with real values.

In summary, with the help of the collected data and expert knowledge on the topic, prior and posterior models were created. Based on these models, an analysis of the prediction of energy production and consumption from solar panels, as well as the significance of the influencing factors, was conducted. The accuracy of the models was supported by the presented charts, and the obtained results largely align with the initial assumptions.

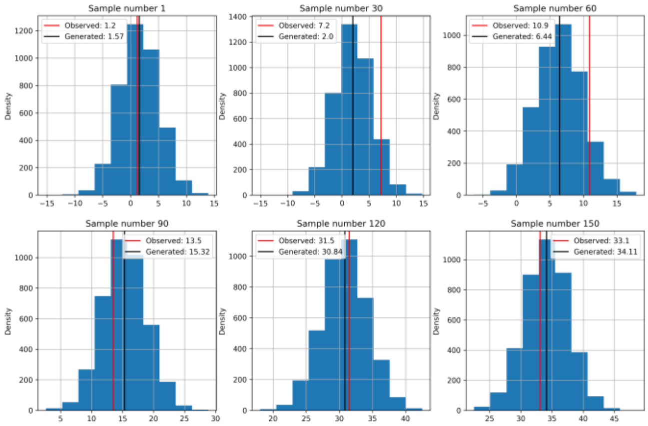

Fig. 21. Histograms for the first model showing the distribution of energy production for one day in every month along with overlaid real and analysis-generated values. These values generally overlap.

The advantages of the implemented models, as well as the overall topic being discussed, undoubtedly include the personally collected real-world data from our own source. Besides the obvious benefit of the uniqueness of this data, it is also very useful to have the ability to continuously add new values, which likely improves the quality of the models. Another advantage is the simple and clear way of presenting data on readable charts. In addition to the

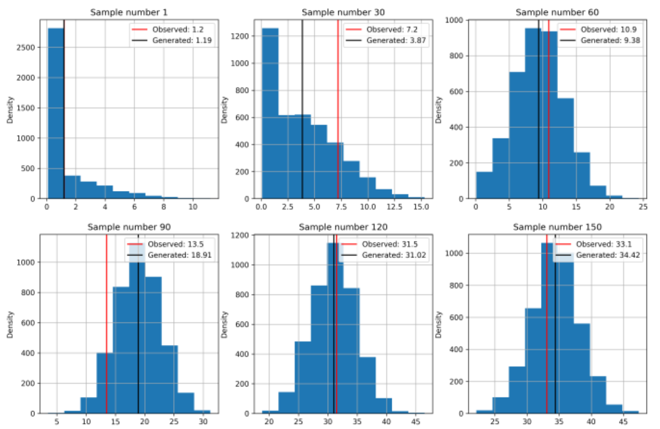

Fig. 22. Histograms for the second model showing the distribution of energy production for one day in every month along with overlaid real and analysis-generated values. These values generally overlap.

distributions and values of the predicted data, every possible factor affecting the results can be thoroughly examined.

On the other hand, the limitations we encountered include problems with finding comprehensive weather data presented in an accessible manner, which necessitated appropriate processing of such data for its subsequent use. This required a significant amount of work and can be somewhat discouraging, thus limiting further exploration of the topic. It is also worth mentioning that the collected data comes solely from a single set of solar panels, and therefore lacks diversity in terms of types and parameters of the devices.

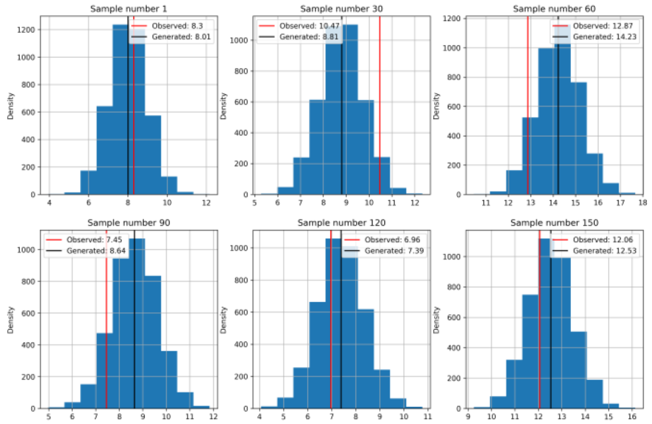

Fig. 23. Histograms for the experimental model showing the distribution of energy consumption for one day in every month along with overlaid real and analysis-generated values. These values generally overlap.

Improving the models to prevent the generation of negative values would be an important step in enhancing their accuracy and ensuring the generated samples align more closely with the expected values. Exploring alternative distributions or constraints for the parameters could potentially help mitigate this problem and further improve the models' performance.

Another potential direction for the development of the discussed topic is to obtain data from other solar panels, thereby adding additional factors such as panel power or tilt angle relative to the surface. This is an appealing prospect, which would allow for much more detailed advice for people interested in installing such devices in their households.

Authors contributions

Conceptualization, A.P. and J.B.; methodology, A.P. and J.B; software, A.P.; validation,A.P. and J.B; formal analysis, A.P.; investigation, A.P. ; resources, J.B.; data curation, A.P.; writing—original draft preparation, A.P.; writing—review and editing, A.P. and J.B; visualization,A.P.; supervision, J.B.; project administration, J.B; funding acquisition, J.B. All authors have read and agreed to the published version of the manuscript.

Acknowledgment

Second author's work was partially realised in the scope of project titled "Process Fault Prediction and Detection". Project was financed by The National Science Centre on the base of decision no. UMO-2021/41/B/ST7/03851. Part of work was funded by AGH's Research University Excellence Initiative under project "DUDU - Diagnostyka Uszkodzeń i Degradacji Urządzeń".

Authors: *Inż. Adrian Poniatowski, Prof. Jerzy Baranowski, Laboratory of Computer Science in Control and Management, Department of Automatic Control and Robotics, Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering , AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Kraków, Poland*

REFERENCES

[1] Breyer, C., Bogdanov, D., Gulagi, A., et al. (2017). On the role of solar photovoltaics in global energy transition scenarios. Progress in Photovoltaics: Research and Applications, 25(8), 727-745.

[2] Lachut, David, Nilanjan Banerjee, and Sami Rollins. "Predictability of energy use in homes." International green computing conference. IEEE, 2014.

[3] Nfaoui, Mohamed, and Khalil El-Hami. "Extracting the maximum energy from solar panels." Energy Reports 4 (2018): 536-545.

[4] El Hammoumi, Aboubakr, et al. "Solar PV energy: From material to use, and the most commonly used techniques to maximize the power output of PV systems: A focus on solar trackers and floating solar panels." Energy Reports 8 (2022): 11992-12010.

[5] Sampaio, Priscila Gonçalves Vasconcelos, and Mario Orestes Aguirre González. "Photovoltaic solar energy: Conceptual framework." Renewable and sustainable energy reviews 74 (2017): 590-601.

[6] Fouad, M. M., Lamia A. Shihata, and ElSayed I. Morgan. "An integrated review of factors influencing the performance of photovoltaic panels." Renewable and Sustainable Energy Reviews 80 (2017): 1499-1511.

[7] Mattei, Michel, et al. "Calculation of the polycrystalline PV module temperature using a simple method of energy balance." Renewable energy 31.4 (2006): 553-567.

[8] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian Data Analysis. CRC Press.

[9] van de Schoot, R., Depaoli, S., King, R. et al. Bayesian statistics and modelling. Nat Rev Methods Primers 1, 1 (2021).

[10] Aki Vehtari, Andrew Gelman, Jonah Gabry, Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC [web page] https://doi.org/10.48550/arXiv.1507.04544 [Accessed on 13 Jun. 2024.].

[11] Aki Vehtari, Daniel Simpson, Andrew Gelman, Yuling Yao, Jonah Gabry, Pareto Smoothed Importance Sampling

[12] [web page] https://doi.org/10.48550/arXiv.1507.02646 [Accessed on 13 Jun. 2024.].

[13] Sumio Watanabe, Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory [web page] https://doi.org/10.48550/arXiv.1004.2316 [Accessed on 13 Jun. 2024.].

[14] [web page] https://python.arviz.org/en/stable/api/generated/arviz.compare.html [Accessed on 17 Jun. 2024.].