GEOLOCALIZED PHOTOVOLTAIC ENERGY PREDICTION METHODOLOGY USING MACHINE LEARNING

Nicolas Chianella
Institute of Informatics, HES-SO Valais Wallis, Switzerland
nicolas.chianella@hevs.ch

Dominique Genoud
Institute of Informatics, HES-SO Valais Wallis, Switzerland
dominique.genoud@hevs.ch

Jean-Marie Alder
Institute of Sustainable Energy, HES-SO Valais Wallis, Switzerland
jean-marie.alder@hevs.ch

Olivier Arbellay
Institute of Informatics, HES-SO Valais Wallis, Switzerland
olivier.arbellay@hevs.ch

Jérôme Treboux
Institute of Informatics, HES-SO Valais Wallis, Switzerland
jerome.treboux@hevs.ch

Jérémie Vianin
Institute of Sustainable Energy, HES-SO Valais Wallis, Switzerland
jeremie.vianin@hevs.ch

David Wannier
Institute of Sustainable Energy, HES-SO Valais Wallis, Switzerland
david.wannier@hevs.ch

ABSTRACT

Knowing the photovoltaic (PV) energy produced at a given time and a specific position is crucial to handle renewable energy. Nowadays, Machine Learning (ML) is broadly used to predict energy production and consumption, but existing datasets cover only some regions in each country, which impedes the deployment of such systems in production. This paper proposes a novel generalized geo-localized methodology to predict photovoltaic production for the next hour and the next day at a given geographical point. This result is part of the deliveries of a Swiss Federal office of Energy (SFOE) project called Micro Storage Intelligent and Distributed (MSID).

INTRODUCTION

In an evolving renewable energy context, solar energy prediction is necessary to manage microgrid systems, such as potential shortages in bad weather. Indeed, to exploit a photovoltaic (PV) power plant, knowing the weather is an essential information (e.g., solar radiation and cloud cover to predict the energy production during the day for the next hour and the next day. Weather information includes many variables, such as solar radiation, ambient temperatures, and wind speed [1].

Factors like bad weather, trees, mountains, and buildings' shadows will reduce energy production because fewer radiations will reach solar panels. Large PV sites can exploit locally installed weather stations and collect this important weather information close to the solar panels; unfortunately, many stations are only sometimes available, and gathering consistent local weather data as that can be expensive and complicated. Moreover, some microgrid systems using PV power plant as the primary power source and installed in isolated locations require an optimally sized battery to handle weather inconsistencies [2].

Nowadays, some PV power predictions are made using Machine Learning (ML) or deep-learning (DL) techniques such as Artificial Neural Networks (ANN) [3]. Often, these techniques are model-based, using weather information as input features and power production values as predicted output. Historical weather data and power production are mostly used to train and evaluate ML models. [5]

To overcome weather inconsistencies and acquisition methods issues, we chose a scalable and integrated solution that exploits (1) historical PV power production and (2) historical and predicted weather data provided by *meteoblue* (www.meteoblue.com) [6], a platform able to deliver weather data from many given geographical points. Then for some of our microgrids, we compared the PV prediction of solar power at **t+1h** and **t+24h** generated by our ML models to actual measurements and actual power production (ground truth).

We finally show how to exploit the ML models in the energy production system (MSID) using live geo-localized daily data like the data used to train the ML models.

DATASET AND METHODOLOGY

Dataset

We used two datasets generated by equipment installed during a four-year SFOE project, in two pilot areas located in the swiss alps, more precisely in Valais. In this paper, we refer to both datasets using the name of the villages: *Saillon* and *Nendaz.*
Saillon is a village on the south side of a mountain, which is more exposed to solar radiation. The PV system is placed on the roof of a local thermal station. *Nendaz's* PV system is located on our pilot site's roof. *Nendaz* is on the northern side of a mountain with less exposition to the sun. We installed a weather station of the company Davis Instrument to collect data that we use to compare to our

methodology results (see next Chapter). The dataset of this weather station in *Saillon* is composed of 45 days, where all days containing missing values have been dropped.

We had to deal with multiple missing values using data from production. Indeed, several time ranges were lost, and the sensors sometimes generated faulty values. For example, a sensor wrongly returned the cumulative power of the different microinverters instead of the three phases retrieved. Therefore, we decided to extract the most extended time range available without missing values for the first pilot site, *Saillon*, i.e., the data from mid-May 2020 to October 2020. The second site of *Nendaz* contains data from mid-August 2021 to November 2021.

We then created the first hourly output class based on the theoretical maximum photovoltaic production curve (***PV$_{max}$***) using the European PVGIS data for each pilot site in the dataset. [7]

PV$_{max}$ represents the maximum daily from sunrise to sunset, calculated on the values from 2005 to 2020. To generate the target for our training set, we used different information for each site:

- GPS position of the solar panels ($positionGPSsite$),
- Installed PV power ($puissancePV_{site}$),
- Panel inclination ($inclinaisonPV_{site}$),
- The azimuth of the panels ($azimutPV_{site}$).

With this information, we generated a PV production curve ($prodPVsite$) for each hour between 2005 and 2020 using the data from PVGIS structured as follows:

[positionGPS$_{site}$, puissancePV$_{site}$, inclinaisonPV$_{site}$, azimutPV$_{site}$].

For each day of the year between 2005 and 2020, we generated the theoretical maximum PV production curve for each pilot site using the following equation:

$$PVmax_{site} = \max_{hour}(prodPV_{site}).$$

For example, the maximum PV production for *Saillon* on the 30th of March at 13:00 is estimated by taking the maximum of the PV production curve (given by PVGIS data) of all available 30th of March at 13:00 between 2005 and 2020. Finally, our full dataset will consist of the following:

*[**t0**, station name, feature, target, **t0 + δ_T**],*

where ***t0*** is the prediction date, and ***δ_T*** is the time increment for which we want to predict the PV energy production.

We also integrated a new weather dataset that depends on the location where the prediction will be made. For this purpose, we retrieved simulation data and weather predictions provided by *meteoblue*.

Methodology

We propose a decision-making method for the battery charge cycle for each pilot site:

1. Generation of the PV$_{max}$ curve,
2. Retrieval of *meteoblue* simulation data,
3. Retrieval of *meteoblue* prediction data,
4. Generation of the target variable of our models (***t+1*** hour, and 24 steps of the following day) based on the *meteoblue* simulation data,
5. Training of models with *meteoblue* prediction data,
6. Incorporation of the models into an API for PV prediction and battery charge/discharge decision display (see Section Production implementation).

Generation of the PV$_{max}$ curve

The generation of the PV$_{max}$ curve is defined in the dataset section.

Retrieval of *meteoblue* simulation data

To generate our new target variable, we retrieved the simulation data at time ***t*** from *meteoblue*, composed of the sunshine, direct and diffuse radiation, and cloud cover.

Retrieval of *meteoblue* prediction data

We also retrieved ***t+n*** hour prediction data at time **t**. Indeed, the *meteoblue* models predict weather features every day at noon for the next day (one prediction per hour).

Generation of the target variable of our models based on the *meteoblue* simulation data

We retrieved from the *meteoblue* data a feature named *Cloud Cover Total* (with a value between 0 and 1, 0 being no clouds, 1 being a wholly covered sky). We calculated our new target variable as follows:

$$pv_{target}(t) = PVmax(t) * (1 - 0.75 * Cloud\ Cover\ Total(t)).$$

The goal is to compute a realistic simulation of PV production for a pilot site (pv_{target}). If we have 0% of clouds, we produce the complete PV$_{max}$. If we have 100% of clouds, we produce 25%*PV$_{max}$. Between these two limits, we consider that the evolution of production is linear. The 25% power production for a total cloud cover is extracted from our internal knowledge and observation on many solar production sites.

Training of models with *meteoblue* prediction data

We have one ML model to predict PV production at **t+1h** and **t+24h** using the simulation data from *meteoblue* available at **t**, **t-1h**, and **t-2h**, the prediction data from *meteoblue* available at **t+1h**, and the computed **PV$_{max}$** at **t+1h**.

EXPERIMENTS AND RESULTS

For both pilots' sites, we calculated the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Average Percentage Error (MAPE) metrics. This calculation is done for each day, from sunrise to sunset, using the actual PV measures as the ground-truth.

Table 1: RMSE, MAE, MAPE for Saillon and Nendaz pilots' sites for the PV prediction.

Metric/Site	***Saillon***	***Nendaz***
RMSE [Wh].	**470.4**	**820.9**
MAE [Wh]	**217.2**	**473.1**
MAPE [%]	**13.2**	**29.9**

Training of the ML models

For the first experiment, ML models were trained for each pilot site using 84530 hourly weather data between the 1st of January 2008 and the 22nd of August 2017 and validated using 41635 hourly weather data between the 23rd of August 2017 and the 23rd of May 2022. Over a whole day, the average error is 217.2 Wh for *Saillon* and 473.1 Wh for *Nendaz*, as shown in Table 1. Figures 1 and 2 show that the average error in Wh on a production day is at most one-third of the production. For *Saillon*, these results are encouraging and demonstrate a low model error.

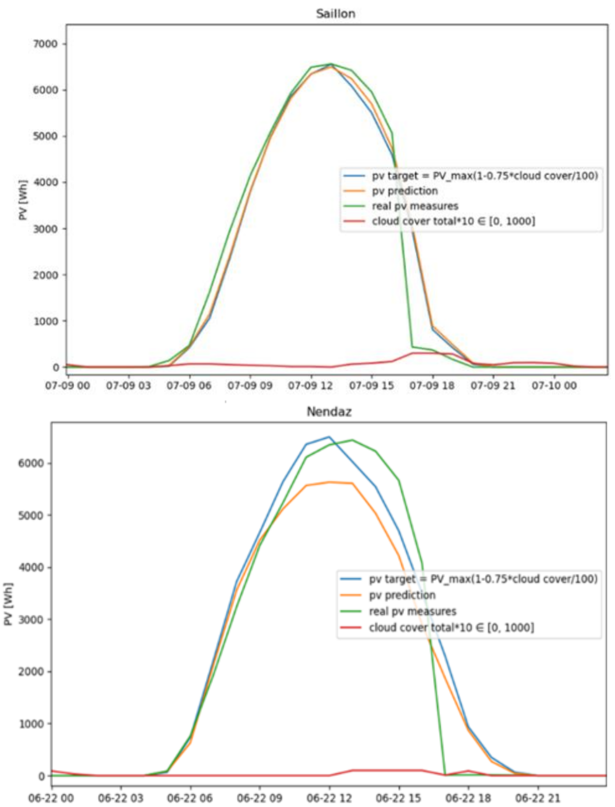

Figure 1: This Figure shows the pv_{target} variable, the PV prediction, the PV measurement, and the Cloud Cover Total curves calculated for sunny days for each pilot site.

For *Nendaz*, the computed error for our ML model is higher and might be improved. Figure 1 shows our calculated target variable for each pilot site, the ML model predictions, and the actual PV measurements for sunny days.

Our target curve closely follows the actual PV measurements for sunny days, as explained after that. For *Saillon*, the average pv_{target} for a sunny day is 3235.6 Wh, and the MAE between the pv_{target} and the measurement is 409.9. For *Nendez,* the average pv_{target} for a sunny day is 3154.9 Wh, and the MAE between the pv_{target} and the measurement is 798.7 Wh.

Our pv_{target} calculation is also validated for cloudy days: the average pv_{target} for a cloudy day is 1008.4 Wh, and the MAE between the pv_{target} and measurement is 306.8 Wh for *Saillon*. Similarly, the average cloudy day of the pv_{target} is 1569.7 Wh, and the MAE between the pv_{target} and the measurement is 724.8 Wh for *Nendaz*.

Figure 2 shows for each pilot site our calculated pv_{target}, ML model PV predictions, and the actual PV measurements for cloudy days.

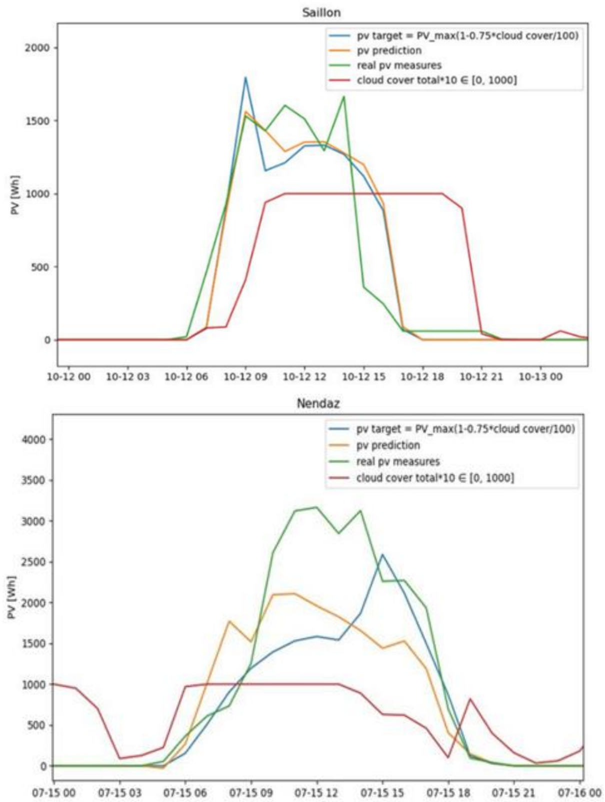

Figure 2: This Figure shows the pv_{target} variable, the PV prediction, the PV measurement, and the Cloud Cover Total curves calculated for cloudy days for each pilot site.

ML model' prediction compared to weather station measurements

We validated our pv_{target} using the solar radiation by computing the MAE.

Table 2: MAE computed using Saillon pv_{target} and the weather station measurements for sunny and cloudy days.

MAE [Wh]/Day	Sunny	Cloudy
Best Day	389.08	268.80
Worst day	1786.69	833.82

Following our experiments, we determined a good weather day from a cloudy day by fixing the threshold of the mean *Cloud Cover Total* at 20 %. As shown in Figure 3, the best PV prediction for a sunny day has an MAE of 389.08 Wh and the worst PV prediction for a sunny day has an MAE of 1786.69 Wh. We notice that for the worst day of good weather, the *Cloud Cover Total* of this day is 0.0 % which means that this day should match the maximum PV production. However, this day has half the usual radiations of a sunny day, which indicates that the weather prediction is too optimistic for this day.

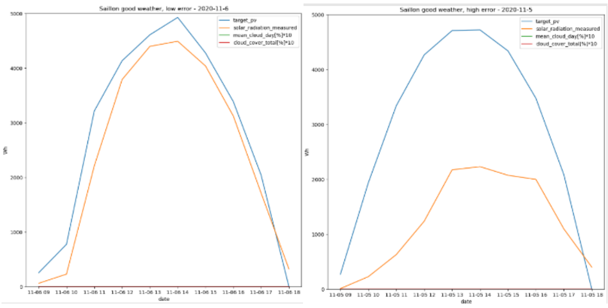

Figure 3: This Figure shows the pv_{target} and the solar radiation curves for a good weather day. Left is the best day and right is the worst day.

For a bad weather day, the best MAE is 268.80 Wh which is better than the best sunny day, and the worst pv_{target} generated is 833.82 Wh (Figure 4).

Overall, the MAE of the pv_{target} compared to the weather station measurement of solar radiation is 727.2 Wh. The results obtained considering the weather station are still relatively good and do not exceed the production tier.

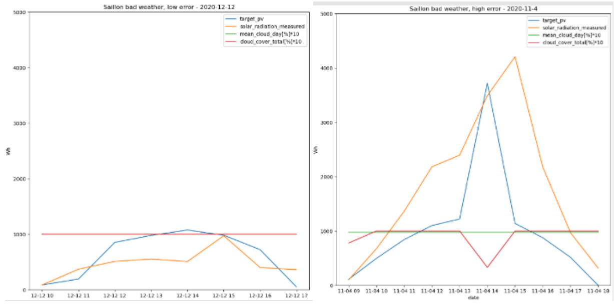

Figure 1: This Figure shows the pv_{target} and the solar radiation curves for a bad weather day. Left is the best day and right is the worst day.

Production implementation

We integrated the calculation of the algorithm previously explained in our Virtual Power Plant (VPP) information system (MSID) to visualize and interact with each distributed microstorage system and optimize the charging and discharging of batteries according to the prediction of photovoltaic production. This platform allows us to anticipate the production of the equipment and decide to change and discharge the battery at the optimal time.
Figure 5 shows the result 'ON [W] 3548', which means 3548 watts of PV should be produced in the next hour.

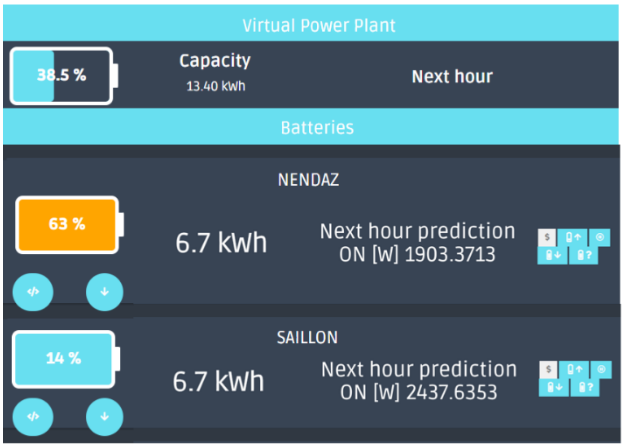

Figure 5: This Figure is an extract from our Virtual Power Plant (VPP) visualization and integration platform, MSID.

On our MSID platform, it is possible to use the value generated by the photovoltaic prediction algorithm to control the charge and discharge of the batteries. As shown in Figure 6, the user having a high level of permission, the *distribution system operator (DSO)*, might create one or many algorithms to control batteries through the inverter and other flexible loads.

Figure 6: This figure illustrates the algorithm creation using our VPP visualization and integration platform, MSID.

DISCUSSION

During our experiments, we needed to develop a method to validate our dataset because of the bad quality of production data for many days due to the chosen inverter creating random outliers and sometimes dropping values to 0 during multiple hours. Therefore, we compared the data from the inverter with the data provided by the nearby weather station and *meteoblue*. The values were significantly different because the sensors created random errors by miscalculating the energy production. Based on these two errors, we concluded that the sensor might fail during data gathering.

A fire made the weather station in *Saillon* inaccessible to compare the real-time data from the sensor with the station's data, therefore we computed the target value using weather data as an alternative to this poor-quality data. As we needed an easily reproducible method for a new site, we proceeded with a scalable and integrated solution based on *meteoblue* weather prediction to produce the $\mathrm{pv}_{\mathrm{target}}$.

Our proposed method had the lowest MAE compared to the on-site weather station's historical data and validated our ML model for *the pilot sites*. The error obtained error is only a quarter of the daily production, which is encouraging according to our previously conducted experiences.

CONCLUSION

Our experiments achieved an MAE for *Saillon* of 217.1 Wh and 473.1 Wh for *Nendaz*. The prediction models were trained using computed weather data provided by *meteoblue*. During this work, we created a generalized method to create the target value for each geolocalized point. For each point, we needed to create a specific ML model that needed to be adapted to match the features related to their solar panels for the new PV production site. The simulated weather combined with live weather forecast enables PV production prediction without using an on-site weather station. Current models are deployed and running on our VPP platform MSID, providing energy production prediction and helping the decision to charge the batteries, improving overall efficiency. The batteries are charged and discharged according to the predictions and parameters defined by the DSO on the visualization platform MSID.
Following our experiments, future research endeavors should focus on the efficiency behavior of the batteries using the deployed ML models. This line of inquiry has the potential to significantly enhance the efficiency of battery charging and, ultimately, contribute to the overall advancement of sustainable energy systems.

REFERENCES

[1] Weibel A., Jordan N., and Wannier D., "A data augmentation methodology for machine learning modelling of distribution power grid: Application on optimal storage sizing and control", The 26th International Conference and Exhibition on Electricity Distribution (CIRED 2021), 20-23 September 2021, Online Conference, pp. 2564-2568, https://doi.org/10.1049/icp.2021.2130.

[2] Huang X., Shi J., Gao B., Tai Y., Chen Z., and Zhang J., "Forecasting Hourly Solar Irradiance Using Hybrid Wavelet Transformation and Elman Model in Smart Grid", *IEEE Access*, 2019, 7: pp. 139909-139923, https://doi.org/ 10.1109/ACCESS.2019.2943886.

[3] Al-Dahidi S., Louzazni M., and Omran N., "A Local Training Strategy-Based Artificial Neural Network for Predicting the Power Production of Solar Photovoltaic Systems", *IEEE Access*, 2020, 8: pp. 150262-150281, https://doi.org/10.1109/ACCESS.2020.3016165.

[4] Al-Dahidi S., Ayadi O., Adeeb J., Alrbai M., and Qawasmeh B. R., "Extreme Learning Machines for Solar Photovoltaic Power Predictions", *Energies*, 2018, 11(10): article 2725, https://doi.org/10.3390/en11102725

[5] Monteiro C., Fernandez-Jimenez L.A., Ramirez-Rosado I.J., Muñoz-Jimenez A., and Lara-Santillan P.M., "Short-Term Forecasting Models for Photovoltaic Plants: Analytical versus Soft-Computing Techniques", *Mathematical Problems in Engineering*, 2013, 2013: article 767284, https://doi.org/10.1155/2013/767284.

[6] Meteoblue, "Meteoblue Weather History", accessed on December 2022, https://www.meteoblue.com/en/historyplus.

[7] European Commission, "PVGIS Online Tool. (n.d.)", accessed on February 2022, https://joint-research-centre.ec.europa.eu/pvgis-online-tool_en.