Gaussian Process Regression and Active Learning for Solar Irradiance to Power Conversion

Andreas Holm Nielsen[1]

I. INTRODUCTION

Bayesian methods offer an interesting approach in regimes such as weather analysis and forecasting, where uncertainty quantification is essential. As stated in [1]; "Understanding what a model does not know is a critical part of many machine learning systems." Incorrect weather forecasts, especially in severe weather scenarios, can have dramatic consequences for the general public in the form of power outages and even natural disasters in some parts of the world. Therefore, being able to determine how confident the model is for any given estimate is deemed critical.

In this paper, we investigate the task of empirical irradiance to power conversion. Solar production has become an increasingly integral part of modern electrical grids, especially in countries such as Germany [2]. As a result, it has become increasingly important to have accurate forecasts of solar production for improved renewable energy planning. This also entails having quantifiable uncertainty estimates to avoid activating expensive short-term power reserves. Public solar production measurements are also limited, and data is generally hard to obtain. Therefore we experiment with an active learning strategy in this scenario, where we assume we only have access to data from one power plant. Then we try and make predictions and uncertainty estimates on another.

We start by introducing the relevant theory in Section II, followed by our experimental setting and cognitive discussion in Section III. Finally conclusions are drawn in Section IV.

II. METHOD

In this section we will start by describing our main method of choice, Gaussian Processes, followed by a description of our Active Learning strategy.

[1]A. H. Nielsen is with Department of Engineering, Signal Processing Group, Aarhus University, Denmark

A. Gaussian Process

The definition of a Gaussian Process (GP) is a probability distribution over functions [1]. More formally, it is a collection of random variables, any finite number of which have a joint Gaussian distribution. The method is non-parametric since the number of parameters grows with the number of data points. More explicitly, the GP defines a prior over functions, $p(f|x)$, which we can convert into a posterior over functions, $p(f|x, y)$, after we observe some function values for f. This is the idea behind Bayesian inference, i.e., we update our hypothesis as new information becomes available. For GP the new information is the training data, which we use to constrain the set of possible functions to pass through these points. Since we have assumed a joint multivariate Gaussian distribution, we can constrain our posterior due the fact that multivariate Gaussians are closed under conditioning [1].

The key feature of GP's lie in the formulation of the covariance function, which needs to be positive semi-definite. It is typically called the kernel function and is responsible for computing the covariance matrix stated above, Σ. One typical choice is the Squared-Exponential (SE) kernel [1].

Let us consider GP for regression, called Gaussian Process Regression. In contrast to linear regression, this is a probabilistic approach, meaning we can quantify uncertainty for our predictions, critical in weather forecasting regimes. Given weather phenomena typically exhibit a considerable degree of noise due to measurement inaccuracies, we consider the case of prediction using noisy observations. We assume the noise term is

additive and independent normally distributed with mean zero and variance σ_n^2.

If we assume our data has mean zero and standard deviation one, we can formulate the joint distribution of training observations y and test output y_* as the prior

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}_* \end{bmatrix} \sim \left(\mathbf{0}, \begin{bmatrix} K(\mathbf{X},\mathbf{X}) + \sigma_n^2\mathbf{I} & K(\mathbf{X},\mathbf{X}_*) \\ K(\mathbf{X}_*,\mathbf{X}) & K(\mathbf{X}_*,\mathbf{X}_*) \end{bmatrix}\right) \tag{1}$$

where X denotes the training points, and X_* denotes the test points. $K(X, X)$ is the $N \times N$ kernel (covariance) matrix for all training point pairs, $K(X, X_*)$ denotes the $N \times N_*$ kernel matrix for all combination of pairs from the training and test points. Finally, $K(X_*, X_*)$ denotes the $N_* \times N$ elements of the covariance matrix for all test point pairs.

Following this, we can condition the joint Gaussian prior distribution on our training observations, which gives us the posterior

$$\begin{aligned} p(\mathbf{y}_* \mid \mathbf{y}, \mathbf{X}, \mathbf{X}_*) &= \mathcal{N}\left(\boldsymbol{\mu}_{\mathbf{y}_*|\mathbf{y}}, \boldsymbol{\Sigma}_{\mathbf{y}_*|\mathbf{y}}\right) \\ \boldsymbol{\mu}_{\mathbf{y}_*|\mathbf{y}} &= K(\mathbf{X}_*,\mathbf{X})\,[K(\mathbf{X},\mathbf{X}) \\ &+\sigma_n^2\mathbf{I}]^{-1}\mathbf{y} \\ \boldsymbol{\Sigma}_{\mathbf{y}_*|\mathbf{y}} &= K(\mathbf{X}_*,\mathbf{X}_*) - K(\mathbf{X}_*,\mathbf{X}) \\ \text{K(X, X)}&+\sigma_n^2\mathbf{I}^{-1}K(\mathbf{X},\mathbf{X}_*) \end{aligned} \tag{2}$$

These are the key equations for Gaussian Process Regression, which we will implement in our experimental setting later.

B. Active Learning

In Active Learning we seek to reduce the "uncertainty" of our learning scheme as much as possible [3]. This implies that the learning process is fast as it only focuses on investigating important elements for the task at hand. There are two entities in an active learning strategy; a) a learner that can ask questions and b) a teacher that can answer questions posed by a). In the setting of machine learning, this is referred to as Active Machine Learning. It works by allowing the learner to decide which labelled data points it wants to use for learning, which it can only do by being able to quantify how confident it is in its predictions on some unlabelled data. This contrasts passive machine learning, where we keep the training dataset "frozen", which can be detrimental to learning in several scenarios [3].

There are many different sampling techniques in Active Learning, described in detail in [3]. Here, we use GP regression to estimate uncertainty for our sampling strategy. GPs naturally provide a way to estimate the uncertainty of a prediction via the covariance function, and as such, can be used directly in an active learning setting. The scheme we use then simply samples the point with the largest variance estimated by our GP model.

III. EXPERIMENTS AND HUMAN COGNITION COMPARISON

Solar irradiance to power conversion is a classical problem in the renewable energy literature. Roughly speaking, there are two methods to go from irradiance to power; a) physical calculation using a PV (photovoltaics) model or b) empirical conversion [4]. The physical calculation of a) requires all parameters related to PV production for each solar plant. The empirical conversion uses regression methods to estimate the empirical correlation between solar irradiance and solar production for any given power plant. It has been shown that empirical methods can be just as accurate as the physical methods, which reduces the amount of required information substantially [5]. Therefore, this is the approach we adopt for this paper.

We now turn to our experiment in Section III-B and then draw parallels to human cognition in Section III-C.

A. Experiment 1 - Power to Irradiance

The first dataset we collect is satellite-derived solar irradiance. It is acquired from the European Meteosat Second Generation (MSG) satellites operated by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT). This data has been made publicly available by [6] for the period 1983-01-01 - 2017-12-31 in 30 minute increments with a spatial resolution of $0.05°$. This serves as our independent variable.

Public solar production measurements are hard to come by in the literature. Instead, we use simulated power profiles from a recent paper [4] on a latitude-longitude basis for Europe. To create their simulated dataset, the authors used actual PV output measurements to correct their simulations.

These simulations were designed explicitly to capture the full distribution of different technologies and geography of solar plants in Europe.

For our experiment, we select two months of data from 2015. The specific training dataset we use is based on the latitude and longitude coordinates of the largest solar park in Germany, called Meuro Solar Park. We conduct several pre-processing steps to account for corrupt data, missing observations, etc., which are all described in the supplementary notebook `advanced_DTU_introduction_to_data.ipynb`.

Next, we instantiate a Gaussian Process Regression model using the SE kernel defined earlier. We define the noise term to be relatively high ($\sigma^2 = 4$), given that we expect noisy observations when using simulated power production estimates. Finally, we optimize the parameters of the SE kernel using Gradient Descent similar to [1]. The results of the GP model can be seen in Figure 1.

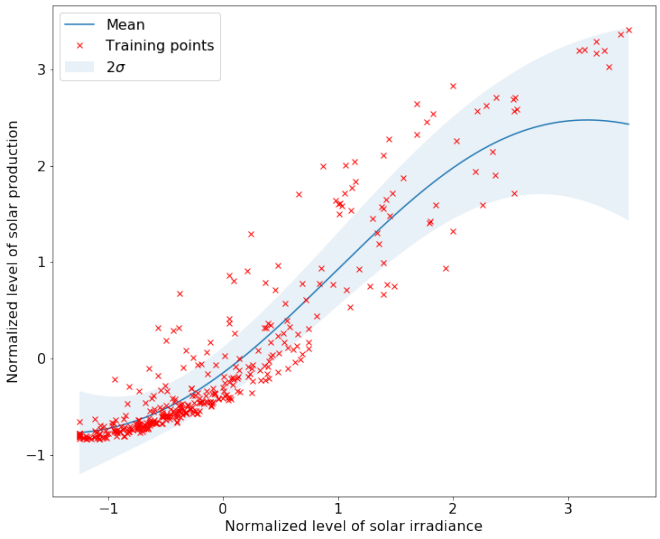

Fig. 1. Predicted solar production for Meuro Solar Park using an optimized Gaussian Process Regression Model with an Squared Exponential Kernel function.

We can observe from Figure 1, that a lot of training points are clustered for low levels of both irradiance and solar production. This results in the uncertainty estimates being lower for these values. The mean function also does a relatively good job at capturing the somewhat exponential relationship between irradiance and solar production.

As another experiment, we want to investigate how well the GP regression model performs on a held-out test set compared to one of the most popular deterministic supervised learning models, linear regression. We select another large solar park, called Neuhardenberg Solar Park, to serve as our test set. We do not perform any parameter calibration or similar on this dataset. The results of this comparison can be seen in Table I.

TABLE I

PERFORMANCE EVALUATION FOR THE GAUSSIAN PROCESS (GP) REGRESSION MODEL RELATIVE TO A LINEAR REGRESSION (LR) MODEL MEASURED USING ROOT-MEAN SQUARED ERROR AND STANDARD DEVIATION OF ERRORS.

Model	**Root Mean-Squared Error**	**Standard Deviation of Errors**
LR	0.160	0.230
GP	0.149	0.219

It should be quite evident from Table I that the Gaussian Process model is superior to regular linear regression. Additionally, we are able to quantify uncertainty estimates for our predictions, which makes it even more relevant for irradiance to power conversion as we do expect a considerable degree of uncertainty.

B. Experiment 2 - Active Learning Strategy

As described in Section II-B, we implement an active learning strategy to demonstrate how we can cope in settings with potentially limited data, such as the task of solar irradiance to power conversion. The specific active learning we use is as follows. We start by defining a labelled dataset as n, which is Meuro Solar Park, and the Neuhardenberg Solar Park as the test set. Next, we sample five initial points from n, which we use to train our GP model. After training, we can predict a mean value and uncertainty estimates on the test set, similar to the experiment in Section III-B. The results can be seen in Figure 2.

We can clearly see the uncertainty estimates are not capturing the distribution of possible outcomes very well, which is unsurprising given we are only training on five randomly chosen points.

Now, we will start using the active learning strategy. For 10 iterations, we will at each iteration ask the GP model to give us an uncertainty estimate for all points in our training set. Then, we select the largest uncertainty and ask the teacher to label

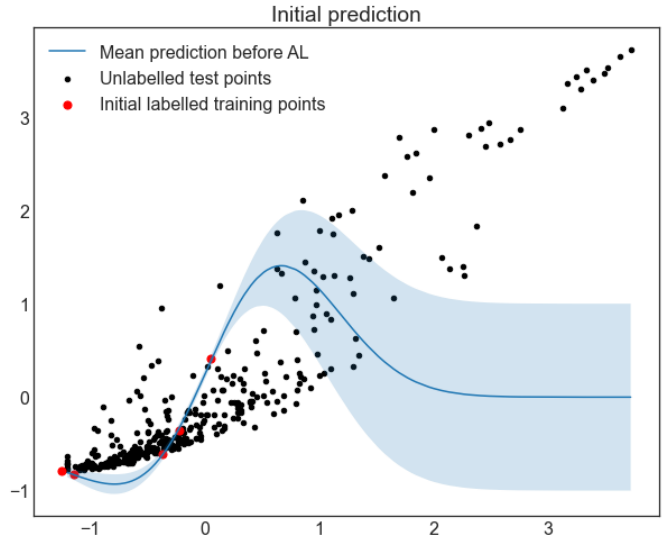

Fig. 2. Predicted solar production for Neuhardenberg Solar Park using Gaussian Process Regression with five randomly sampled points.

this observation for us. After that, we update the GP model with the new labelled observation. After doing this for only 10 observations, we achieve the predictions seen in Figure 3.

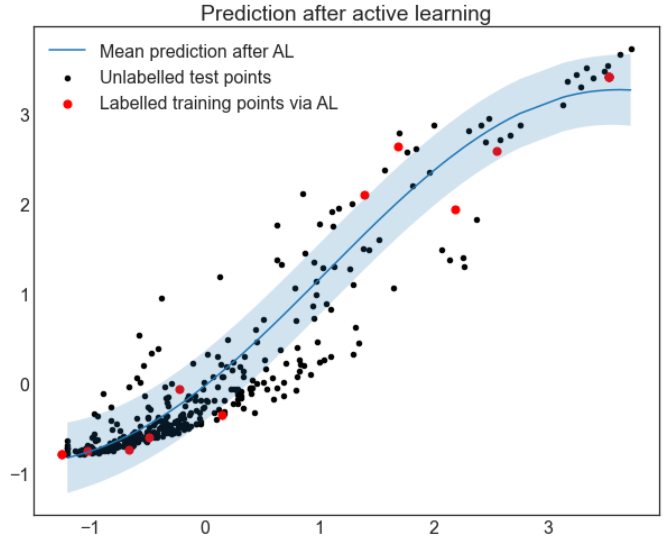

Fig. 3. Predicted solar production for Neuhardenberg Solar Park using Gaussian Process Regression with an Active Learning Strategy.

It is quite impressive that training using only 15 observations, we can get a GP fit that looks surprisingly similar to the GP from Figure 1. This tells us that an active learning strategy seems to be optimal for irradiance to power conversion in terms of being data-efficient and compute-efficient as we only need to train our model on 15 observations.

As a final experiment, we briefly compare the active learning strategy with a random sampling strategy for 15 data points. The results can be seen in Figure 4.

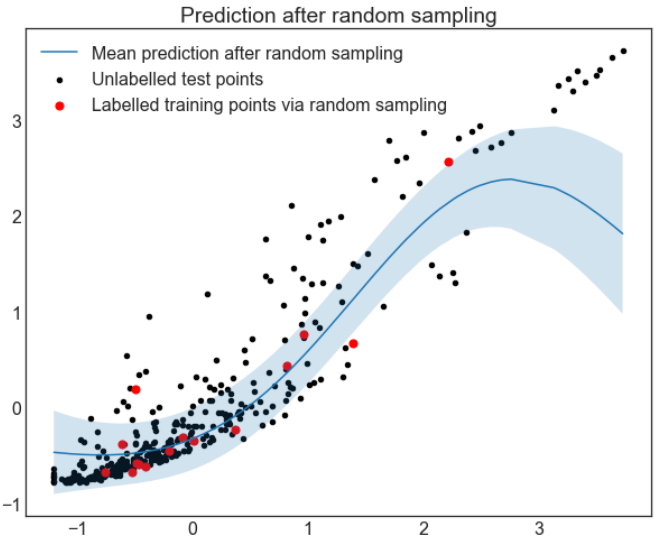

Fig. 4. Predicted solar production for Neuhardenberg Solar Park using Gaussian Process Regression with 15 randomly sampled points.

We can clearly see the uncertainty estimates do not capture the distribution equally well compared to the active learning strategy, especially evident in the upper right corner of Figure 4. The randomly chosen points typically lie close to the mean prediction and not where it would yield maximum benefit to our GP model fit. This contrasts with the active learning strategy, where points were chosen to maximize exploration, clearly seen in the previous Figure 3.

The code for all experiments can be found in the notebook `advanced_DTU_GaussianProcessRegression.ipynb`[1].

C. Human Cognition Perspectives

In [7] the authors present an interesting link between Gaussian Processes and *function learning*, which in the psychology literature is defined as how humans learn relationships between two continuous variables. Two popular theories exist

[1] `https://github.com/holmdk/gaussian_solar_project`

for explaining function learning. The first one is learning explicit functions from a given class such as polynomials of order d. The other approach is less deterministic and instead postulates that people learn by forming patterns or relationships between observed values of input and output variables, and then generalize using similarity measures between new inputs and old inputs. The latter theory bears a close resemblance to Gaussian Processes Regression with the SE kernel by having a simple similarity-based generalization underlying the smoothness assumption. That is, we predict similar y values for similar x values, or stated in another way; the closer the x's are, the higher the correlation between the y's, and vice versa. In this way, we can draw a direct parallel between GPs and one of the popular theories of function learning for humans. The authors also investigated their GP model for two experiments [7]. They demonstrate their GP model being competitive with existing process models in the literature and that the model can capture some of the basic phenomena of human function learning.

IV. CONCLUSIONS

We present a method for empirical irradiance to power conversion based on Gaussian Process Regression, which enables uncertainty quantification in addition to point estimates. This is critical in the renewable energy planning domain, as incorrect estimates of solar energy will lead to energy imbalances that need to be corrected with expensive short-term power reserves. Due to its non-parametric nature, GPs also offer an attractive way of dealing with small amounts of data without being overly prone to overfitting, which is not true for many parametric methods such as neural networks. We demonstrated that our solar irradiance to power conversion worked relatively well and could outperform linear regression on a held-out test set with no calibration. Finally, we took the limited data setting one step further by employing an active learning strategy to gradually grow our dataset and minimize the time it took to optimize the GP model from earlier. This model obtained good results by only training on 15 observations chosen via uncertainty estimates and was considerably more accurate than a random sampling strategy.

REFERENCES

[1] F. Bergamin and K. H. Madsen, *Gaussian Processes*. Advanced Topics in Machine Learning, Lecture Notes. Technical University of Denmark (DTU), 2020.

[2] Wehrmann, Benjamin. (April, 2020) Solar power in Germany – output, business perspectives. Available: https://www.cleanenergywire.org/factsheets/solar-power-germany-output-business-perspectives. Accessed: 2020-08-31. [Online].

[3] P. J. H. Jørgensen and K. H. Madsen, *Active Learning*. Advanced Topics in Machine Learning, Lecture Notes. Technical University of Denmark (DTU), 2020.

[4] S. Pfenninger and I. Staffell, "Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data," *Energy*, vol. 114, pp. 1251–1265, 2016.

[5] S. E. Haupt, B. Kosović, T. Jensen, J. K. Lazo, J. A. Lee, P. A. Jiménez, J. Cowie, G. Wiener, T. C. McCandless, M. Rogers *et al.*, "Building the sun4cast system: Improvements in solar power forecasting," *Bulletin of the American Meteorological Society*, vol. 99, no. 1, pp. 121–136, 2018.

[6] U. Pfeifroth, S. Kothe, R. Müller, J. Trentmann, R. Hollmann, P. Fuchs, and M. Werscheck, "Surface radiation data set–heliosat (sarah)–edition 2, satellite application facility on climate monitoring," 2019.

[7] T. L. Griffiths, C. Lucas, J. Williams, and M. L. Kalish, "Modeling human function learning with gaussian processes," in *Advances in Neural Information Processing Systems (NIPS)*, 2009, pp. 553–560.