A PROJECT REPORT

ON

HYBRID SOLAR IRRADIANCE USING KALMAN FILTER

BY

UDAY MITTAL
2019B4A70662P

Under the Supervision of
DR. SUMANTA PASARI
DEPARTMENT OF MATHEMATICS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF

MATH F266: STUDY PROJECT

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI

(RAJASTHAN)

PILANI CAMPUS

CONTENTS

1)	Abstract	1
2)	Introduction	2
3)	System Model	3
4)	Methodology	5
5)	Dataset Review	9
6)	Conclusion and Future Work	11
7)	Appendix	12
8)	References	13

ABSTRACT

In this work, a hybrid solar irradiance now-casting mechanism is proposed. The proposed Hybrid Predictor fuses the Kalman Filter and Regressor Predictors results to benefit from both techniques. The Kalman filter's time-varying adaptive system function is designed to deal with the ramp-down events for more accurate predictions. Three fusion alternative mechanisms have been proposed and compared as well. They are based on the Local Root Mean Square Error (RMSE) computation.

INTRODUCTION

Governments and International Organizations have consistently promoted sustainable and renewable energy to alleviate the energy crisis dominated by fossil fuels and act against adverse effects indicted by the growing emission of carbon dioxide in the atmosphere. Compared to finite quantities of fossil fuels, solar light, winds, and waves are energy resulting from the activities of the Earth, Sun, and Moon and enjoy the advantage of being sustainable. In addition, the production and use of renewable energy involve negligible emissions compared to fossil fuels. Although these sustainable energy sources' characteristics are unstable and intermittent, the attention towards these renewable forms of energy motivates researchers to overcome the dangerous nature of energy sources.

Solar energy is one sustainable energy that has received growing attention recently. Photo-Voltaics (PV) are in great demand nowadays due to their clean and renewable electricity generation in residential, commercial, and industrial areas. There has been a noticeable installation of PVs even in the regions with moderate solar energy resources. One of the crucial factors for the performance of PV systems is the availability of solar energy on the ground surface that can be converted into electricity.

The PV grid operators require medium and short-term prediction schemes for efficient and accurate Solar Irradiance Prediction. The Prediction Horizon for short-term to medium-term forecast ranges from a few hours to a few days, while it ranges from three to fifteen minutes for the short term. If the exact amount of solar energy drop can be forecasted, it will help reduce the overhead of backup storage, making renewable energy generation even more economical.

SYSTEM MODEL

The System aims to obtain the Predicted Irradiance $\hat{I}_{k+\Delta P}$ using the information of the current and the previous ground truth irradiance $I_{k-\Delta T}, \ldots, I_{k-1}, I_{k}$, and all-sky images $A_{k-\Delta T}, \ldots, A_{k-1}, A_k$. The Regressor Predictor generates the prediction $\hat{I}^{(R)}_{k+\Delta P}$, while the Kalman filter generates the prediction $\hat{I}^{(KF)}_{k+\Delta P}$. The predictions are based on the current and previous ground truth irradiance and the ramp-down event forecasting results. The extracted features from the all-sky images include the mean and variance of several cloud pixels, the mean value of gradient magnitude, the mean value of intensity level, and the mean value of the accumulated intensity along the vertical line of the sun.

The forecast is achieved by the two-class support vector machine (SVM) classifier. All sky-image features are the history ground-truth irradiance is used to form feature vectors for classification purposes. The feature vectors and corresponding ramp-down events train the classifier $C_{RD}(K)$. $C_{RD}(K)$ outputs one if the ramp down event is forecasted at time $K+\Delta P$. Otherwise, it outputs zero.

The local root mean square errors (RMSE) of the two predictors within a time Δt are computed. The proposed hybrid predictor fuses the results from the two predictors according to the local RMSEs. The proposed system framework has been picturized below.

The dataset used was collected from the National Central University of Taiwan. The dataset contains all-sky images and irradiance readings from January 2014 to September 2014. The dataset records one all-sky image and one irradiance reading per minute. The device used to capture the all-sky image is the all-sky camera manufactured by the Santa Barbara Instrument Group. The instrument used to measure solar irradiance is Deltra OHMLP RYRA 03. It is a point sensor with a 10s sampling interval. The measured solar irradiance within each minute is averaged to generate the ground truth irradiance data for each minute.

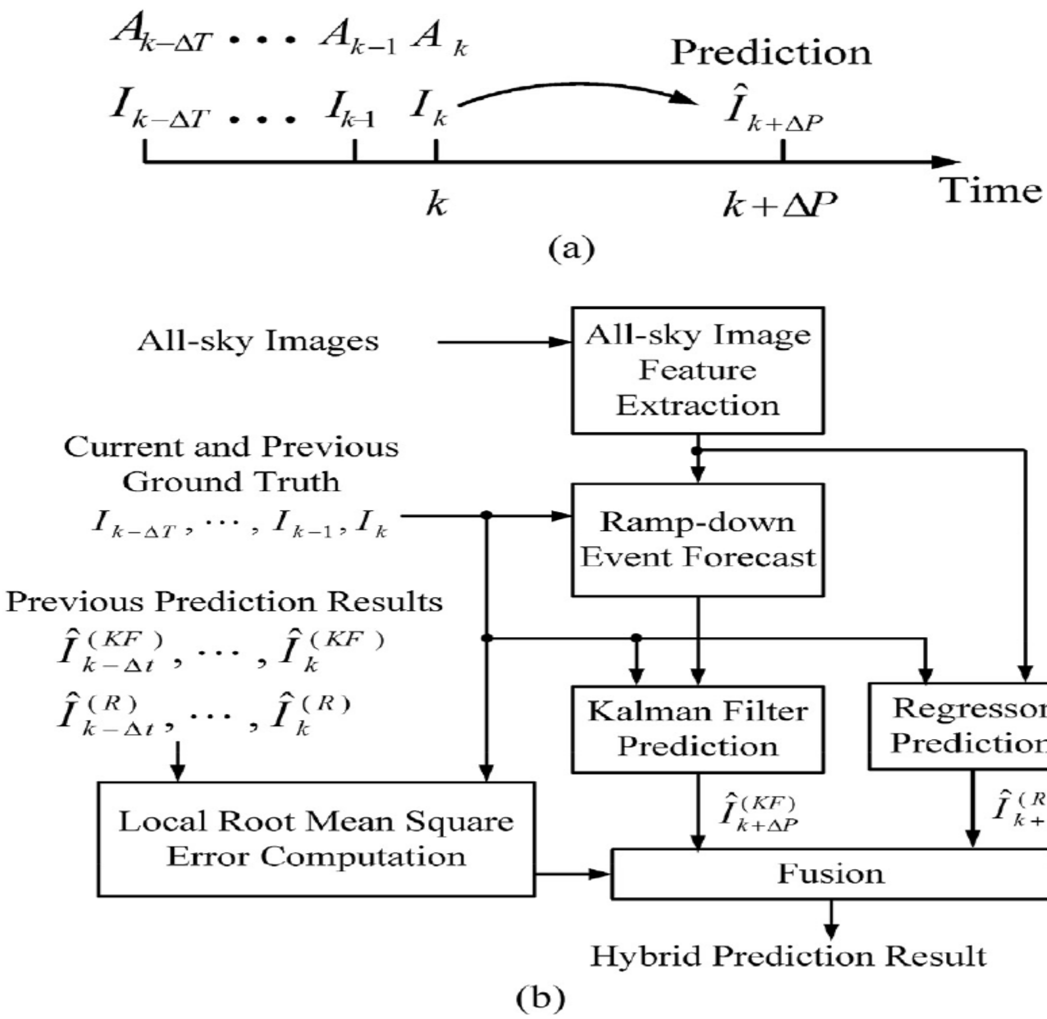

Fig. 1. Proposed system framework.

METHODOLOGY

Kalman Filter, also known as Linear Quadratic Estimation (LQE), is an algorithm that uses a series of measurements observed over time, including statistical Noise and other inaccuracies. It produces estimates of unknown variables that tend to be more accurate than those based on a single measurement alone by estimating a joint probability distribution over the variables for each timeframe. The irradiance prediction can be modeled as a dynamic system.

At time instance k, the current system state X_k can be expressed by the previous state $X_{k-\Delta P}$ going through a system function F_k perturbed by a process noise W_k. The measurement Y_k can be expressed using the system state X_k, a measurement function H_k, and a measurement noise R_k.

$$x_k = F_k x_{k-\varDelta P} + w_k$$

$$y_k = H_k x_k + r_k$$

Kalman filter assumes a linear system function and a linear measurement function. Hence the system function F_k and measurement functions H_k can be expressed in matrices. Kalman filter takes Gaussian process noises and measurement noises.

$$w_k \sim N(0, Q_k)$$

$$v_k \sim N(0, R_k)$$

The above equations denote normalization of Parameters

Here Q_k denotes the process noise covariance, and R_k denotes the measurement noise covariance.

Kalman filter can provide optimal solutions to the dynamic system under the linearity and Gaussian assumptions. If the model's premise holds, then the covariance matrix would accurately reflect the covariance of estimates.

$$\widehat{x}_{k|k-\varDelta P} = F_k \widehat{x}_{k-\varDelta P|k-\varDelta P}$$

$$P_{k|k-\varDelta P} = F_k P_{k-\varDelta P|k-\varDelta P} F_k^T + Q_k$$

The above equations denote Predicted State Estimate and Covariance Matrix

$$S_k = H_k P_{k|k-\Delta P} H_k^T + R_k$$

$$K_k = P_{k|k-\Delta P} H_k^T S_k^{-1}$$

The above equations denote Innovation Covariance and Kalman Gain

The measurement residual indicates how well the filter predicts.

$$\tilde{y}_k = y_k - H_k \hat{x}_{k|k-\Delta P}$$

The above equation denotes Residual of Kalman Filter

The error covariance can have updated using the Kalman gain.

$$P_{k|k} = (I - K_k H_k) P_{k|k-\Delta P}$$

The above equation denotes Predicted Estimate Covariance

The updated state estimate can be obtained using the prediction, the Kalman gain, and the measurement residual.

$$\hat{x}_{k|k} = \hat{x}_{k|k-\Delta P} + K_k \tilde{y}_k$$

The above equation denotes Predicted State Estimate

The adjusting factor (AF) depends on the ramp down event forecasting result. If there is no ramp-down event forecasted, then AF is set to zero.

$$F_k = \begin{bmatrix} 1 & (1 + \alpha_k) \\ 0 & 1 \end{bmatrix}$$

F_K is the adaptive system matrix

$$\alpha_k = \begin{cases} -\operatorname{sgn}\left(\dot{I}_k\right) \cdot M_{RD} / \left|\dot{I}_k\right| & \text{if } C_{RD}(k) = 1 \\ 0 & \text{if } C_{RD}(k) = 0 \end{cases}$$

The above relation denotes the Adjusting Factor

When the prediction results using Kalman Filter are compared with the results using regressors, it was observed that each method has superior performance within different time intervals. Therefore, a hybrid predictor has been designed to integrate the advantages of a Kalman filter predictor and a regressor predictor.

Local RMSE is computed within a local time window for the two predictors to fusion.

$$E_k^{(KF)} = \frac{\sqrt{\sum_{\tau=k-\Delta t}^{k}\left(I_\tau - \widehat{I}_\tau^{(KF)}\right)^2}}{\sqrt{\sum_{\tau=k-\Delta t}^{k}\left(I_\tau\right)^2}}$$

$E_K^{(KF)}$ denotes the Local RMSE for the Kalman Filter.

The three hybrid alternatives are proposed.

1) Fusion alternative 1 selects the predictor with the minimum local RMSE at the current time. The motivation is to choose the prediction scheme that performs better within the local time interval.

$$\widehat{I}_{k+\Delta P}^{(Hybrid\ \ A)} = \begin{cases} \widehat{I}_{k+\Delta P}^{(KF)} & if\ E_k^{(KF)} \leq E_k^{(R)} \\ \widehat{I}_{k+\Delta P}^{(R)} & Otherwise \end{cases}$$

The above expression denotes the Fusion Alternative 1.

2) Fusion alternative 2 selects the Kalman filter prediction result is Kalman filter predictor is less than a threshold, else selects the regressor prediction result., The motivation for this is to use the development of the Kalman filter predictor only when its local RMSE is low enough.

$$\widehat{I}_{k+\Delta P}^{(Hybrid\ \ B)} = \begin{cases} \widehat{I}_{k+\Delta P}^{(KF)} & if\ \ E_k^{(KF)} \leq \theta \\ \widehat{I}_{k+\Delta P}^{(R)} & Otherwise \end{cases}$$

The above expression denotes the Fusion Alternative 2.

3) In Fusion alternative 3, the prediction values of the two predictors are combined using the weighted averages.

$$\widehat{I}_{k+\Delta P}^{(Hybrid\ \ C)} = \frac{\left(E_k^{(R)}\right)^2}{\left(E_k^{(KF)}\right)^2 + \left(E_k^{(R)}\right)^2}\ \widehat{I}_{k+\Delta P}^{(R)} + \frac{\left(E_k^{(KF)}\right)^2}{\left(E_k^{(KF)}\right)^2 + \left(E_k^{(R)}\right)^2}\ \widehat{I}_{k+\Delta P}^{(KF)}$$

The above expression denotes the Fusion Alternative 3.

DATASET REVIEW

The dataset being used is an Indian Dataset collected from the National Solar Radiation Database (NSRDB) website. The dataset is downloaded from a location in Uttar Pradesh, India, having the location ID as 37992, with the latitude being 28.57 and the longitude being 77.46. The time zone and the local time zone of the location are five, and the location's elevation is 207.

The dataset contains a lot of parameters and information, which need to be cleaned and processed before actually working.

The meaning of the various data parameters is as follows

1) DHI – Diffuse Horizontal Irradiance - The solar radiation that does not arrive on a direct path from the sun but has scattered by the clouds and particles in the atmosphere and comes equally from all directions.
2) DNI - Direct Normal Irradiance - The amount of light coming perpendicular to a surface. This type of irradiance belongs to rays that come in a straight line from the sun's direction at their current position in the sky.
3) GHI - Global Horizontal Irradiance - Total amount of shortwave radiation received from above by a horizontal surface (parallel) to the ground.
4) GHI = DNI * cos(Solar Zenith Angle) + DHI
5) Dew Point – The temperature the air needs to be cooled to (at constant pressure) to achieve a relative humidity (RH) of 100%.
6) Solar Zenith Angle – The angle between the sun's rays and the vertical direction
7) Precipitable Water – The water depth in a column of the atmosphere if all the water in that column were precipitated as rain.
8) Surface Albedo – Quantifies the fraction of the sunlight reflected by the surface of the Earth.

The unit of measurement for the parameters of the dataset has been mentioned below:

1) Clearsky DHI Units - W/m^2
2) Clearsky GHI Units - W/m^2
3) Dew Point Units – c
4) DHI Units - W/m^2
5) GHI Units - W/m^2
6) Solar Zenith Angle Units – Degree
7) Temperature Units – c
8) Pressure Units – mbar
9) Relative Humidity Units - %
10) Perciptable Water Units – cm
11) Wind Direction Units – Degrees
12) Wind Speed Units – m/s
13) Cloud Type – 15 – N/A
14) Cloud Type 0 – Clear
15) Cloud Type 1 – Probably Clear

16) Cloud Type 2 – Fog
17) Cloud Type 3 – Water
18) Cloud Type 4 – Super-Cooled Water
19) Cloud Type 5 – Mixed
20) Cloud Type 6 – Opaque Ice
21) Cloud Type 7 – Cirrus
22) Cloud Type 8 – Overlapping
23) Cloud Type 9 – Overshooting
24) Cloud Type 10 – Unknown
25) Cloud Type 11 – Dust
26) Cloud Type 12 – Smoke
27) Fill Flag 0 – N/A
28) Fill Flag 1 – Missing Image
29) Fill Flag 2 – Low Irradiance
30) Fill Flag 3 – Exceeds Clearsky
31) Fill Flag 4 – Missing Cloud Properties
32) Fill Flag 5 – Rayleigh Violation
33) Surface Albedo Units – N/A

The behavior of the data has also been considered. After sorting the data, the data was run through python code, after some slight modifications, and some exciting things were observed. They are:

1) Ozone concentration in the sky remains the same throughout the day, but it varies on different days in the month.
2) Solar Zenith Angle does not depend on the day; it depends only on the hour it is calculated. It is different for different hours of the day.
3) Precipitable Water depends on the day, and it remains constant during the day, at almost all hours.
4) Dew Point also shows the same trends as Perceptible Water.
5) Surface Albedo remains constant for the whole day and month. Surface Albedo remains constant for extensive intervals.

For conducting experiments, specific reference values must be defined to check the results' correctness. Surface Albedo fits this criterion and can be used to check the results. It does not even depend on the hours.

CONCLUSION AND FUTURE WORK

The notes on the dataset are beneficial, as sorting, cleaning, and processing data are time-consuming tasks. Defining a base reference for the experiment also helps cut the time, as, without the base parameters, we can't proceed with any investigation.

Implementing Kalman Filter on the data set is a challenging task as the data is enormous, and hence defining the parameters such as Noise, etc., is tricky. Thus, future work involves implementing the Kalman filter in some programming languages and estimating the future values of DHI, DNI, and GHI, as they are the parameters of most importance.

The codes for the data analyses and failure obtained in implementing Kalman Filter are uploaded and made publicly available to anyone who wants to edit it and continue work can.

Link of Github: Uday-Mi/Study-Project (github.com)

APPENDIX

Nomenclature

$\hat{I}^{(KF)}_{k+\Delta P}$	Predicted solar irradiance by Kalman filter for time instance $k+\Delta P$
$\hat{I}^{(R)}_{k+\Delta P}$	Predicted solar irradiance by Regressor for time instance $k+\Delta P$
ΔP	Prediction horizon
I_k	Ground truth solar irradiance at time instance k
Δt	Time window for local root mean square error computation
A_k	All-sky image at time instance k
ΔT	Time interval for feature extraction
$MEAN_{NC}$	Mean of number of cloud pixels in the all-sky images within time interval ΔT
VAR_{NC}	Variance of number of cloud pixels in the all-sky images within time interval ΔT
$MEAN_{GM}$	Mean value of gradient magnitude in the all-sky images within time interval ΔT
$MEAN_{IL}$	Mean value of intensity level in the all-sky images within time interval ΔT
$MEAN_{SI}$	Mean value of the accumulated intensity along the vertical line of sun in the all-sky images within time interval ΔT
$C_{RD}(k)$	Support vector machine classifier for ramp-down event forecasting
x_k	System state of Kalman filter at time instance k
y_k	Measurement state of Kalman filter at time instance k
F_k	System function of Kalman filter at time instance k
H_k	Measurement function of Kalman filter at time instance k
w_k	System noise of Kalman filter at time instance k
r_k	Measurement noise of Kalman filter at time instance k
Q_k	Process noise covariance of Kalman filter at time instance k
R_k	Measurement noise covariance of Kalman filter at time instance k
$\hat{x}_{k\|k-\Delta P}$	Predicted state estimate of time k given the information of time $k-\Delta P$
$P_{k\|k-\Delta P}$	Predicted estimate covariance of time k given the information of time $k-\Delta P$
$\dot{I}_k$	The difference between I_k and $I_{k-\Delta P}$
$\tilde{y}_k$	Measurement residual of Kalman filter at time instance k
α_k	Adjusting factor for prediction refinement on ramp-down event
M_{RD}	Mean ramp-down value learned from training samples
$E^{(KF)}_k$	Local RMSE of Kalman filter predictor
$E^{(R)}_k$	Local RMSE of regressor predictor
θ	Threshold for the proposed hybrid predictor alternative 2
$[f_1,\cdots,f_n]$	The n dimensional input feature vector of the multiple regressor
$b_0,\cdots,b_n$	The coefficients of the multiple regressor

REFERENCES

1) https://www.weather.gov/arx/why_dewpoint_vs_humidity#:~:text=The%20dew%20point%20is%20the,water%20in%20the%20gas%20form.
2) https://en.wikipedia.org/wiki/Precipitable_water
3) https://www.yellowhaze.in/solar-irradiance/
4) NSRDB Data Viewer (nrel.gov)
5) Surface Albedo | Copernicus Global Land Service
6) Solar zenith angle – Wikipedia
7) https://github.com/rlabbe/Kalman-and-Bayesian-Filters-in-Python/blob/master/07-Kalman-Filter-Math.ipynb
8) https://arxiv.org/ftp/arxiv/papers/1204/1204.0375.pdf
9) https://simdkalman.readthedocs.io/en/latest/
10) https://www.kaggle.com/code/emaerthin/demonstration-of-the-kalman-filter/notebook
11) Hybrid Solar Irradiance now-casting by fusing Kalman filter and regressor by Dr. Hsu-Yung-Cheng, Department of Computer Science and Information Engineering, National Central University, Taiwan. The work was published in ScienceDirect, under the section of Renewable Energy. Hybrid solar irradiance now-casting by fusing Kalman filter and regressor - ScienceDirect
12)