

Research Paper Presentation

Ayush Jha - CS20BTECH11006

June 3, 2021

Gaussian Mixture Model for Estimating Solar Irradiance Probability Density

Abstract

- 1 The increasing need for photovoltaic generation resources demands power network designers to obtain accurate estimates of solar irradiance at a given site.
- 2 The current parametric Beta distribution can be problematic and may lead to model mis-specification.
- 3 The paper proposes Gaussian Mixture Model (GMM) as a more robust estimation of solar irradiance probability density at a certain site.
- 4 Multi-year solar data from eight locations in the United States is utilized to evaluate the accuracy of the GMM estimate and compare its performance with the popular Beta distribution.

Definitions

Probability Density Estimation

Using the observations in a random sample to estimate the general density of probabilities beyond just the sample of data is known as Probability Density Estimation.

Solar Irradiance

The solar irradiance is the output of light energy from the entire disk of the Sun, measured at the Earth.

Gaussian distribution

Gaussian distribution (also known as normal distribution) is a continuous probability distribution for a real-valued random variable with a bell-shaped curve.

Definitions

Gaussian PDF

$$f_X(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp -\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \quad (1)$$

with σ as standard deviation and μ as mean.

Gaussian Mixture Model

A Gaussian mixture model is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters.

The Beta Probability Density Function

Introduction

- 1 The Beta distribution is the most common parametric family of probability distributions used to estimate the probability density of Normalized Solar Irradiance (NSI) data.
- 2 However, the use of a simple parametric distribution (such as the Beta distribution) can possibly lead to the risk of mis-specification of the model and can result in wrong planning decisions.
- 3 It is also used for research works on network planning and design, distribution system partitioning into microgrids, and voltage stability for networks with distributed generation, etc.

The Beta Probability Density Function

PDF

$$f_{\text{Beta}}(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \quad (2)$$

where $x \in (0, 1)$, Γ is the gamma function, and $\alpha, \beta \in (0, \infty)$ are the first and second shape parameters respectively, given by,

$$\alpha = \frac{\gamma \times \beta}{1 - \gamma} \quad (3)$$

$$\beta = (1 - \gamma) \left(\frac{\gamma(1 - \gamma)}{\sigma^2} - 1 \right) \quad (4)$$

where σ and γ are the standard deviation and mean of the random variable, respectively.

Gaussian Mixture Model

Introduction

- ➊ To overcome the limitations by Beta distribution, recent works have suggested new models for estimating solar irradiance based on a non-parametric approach.
- ➋ In this paper, the GMM is proposed for estimating solar irradiance probability density as a more accurate alternative to the Beta distribution.
- ➌ The GMM is a finite convex linear combination of Gaussian densities with different parameters of each probability density.
- ➍ The careful selection of a correct number of mixture components is of top importance when fitting a finite mixture distribution.

Gaussian Mixture Model

Details of model

- 1 The Expectation–Maximization (EM) framework (advanced iterative clustering technique) is adopted in this paper to obtain an estimation of the parameters of the GMM PDF.
- 2 The Bayesian Information Criterion (BIC) approach is considered to determine the number of components in an optimal manner.
- 3 The probability density distribution of the GMM can approximate any arbitrarily shaped non-Gaussian density

Gaussian Mixture Model

PDF

Suppose that the observed solar irradiance values are denoted by $x = x_1, \dots, x_n$, and let the j^{th} entry of the random variable be modeled as a linear combination of Gaussian densities then,

$$f_X(x_j|\theta) = \sum_{i=1}^C w_i \phi(x_i, \theta_i), \quad x_j \geq 0, \quad j = 1, \dots, N \quad (5)$$

where C is the number of mixture components, and each i^{th} component is given by

$$\phi(x_i, \theta_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x_j - \mu_i)^2}{2\sigma_i^2}\right) \quad (6)$$

where its weight is $w_i > 0$, with $\sum_{i=1}^C w_i = 1$. The parameters σ_i^2 and μ_i correspond to the variance and mean of the i^{th} mixture component.

Results and Performance evaluations

Introduction

- 1 The robustness of the proposed GMM and common Beta distributions for characterizing the variability of solar data of eight cities is assessed in this section.
- 2 The solar data had the mean and variance of the solar irradiance values for all eight cities.
- 3 The proposed GMM routine was coded in MATLAB with the PDF parameters computed using the EM algorithm and the optimal number of mixture components found using the BIC approach.
- 4 The Beta distribution shape parameters and density estimates were computed in MATLAB using the functions betafit and betapdf.

Results and Performance evaluations

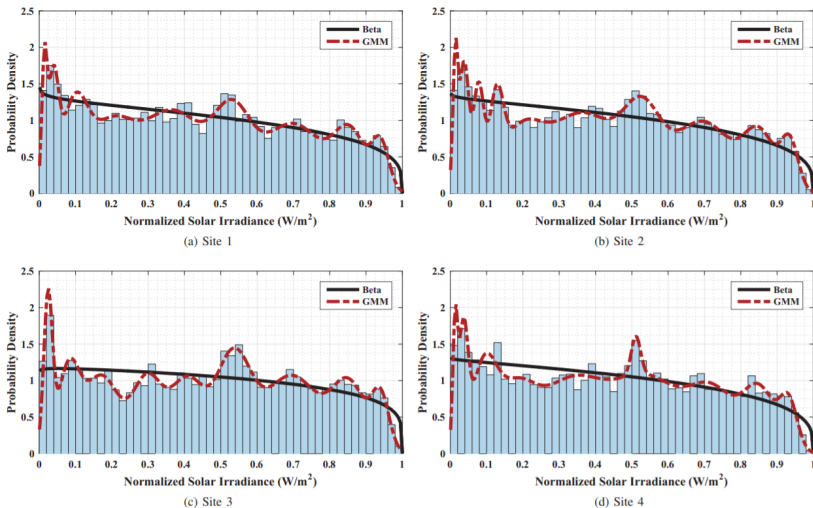


Figure: Probability Density plots and Histograms (sites 1-4)

Results and Performance evaluations

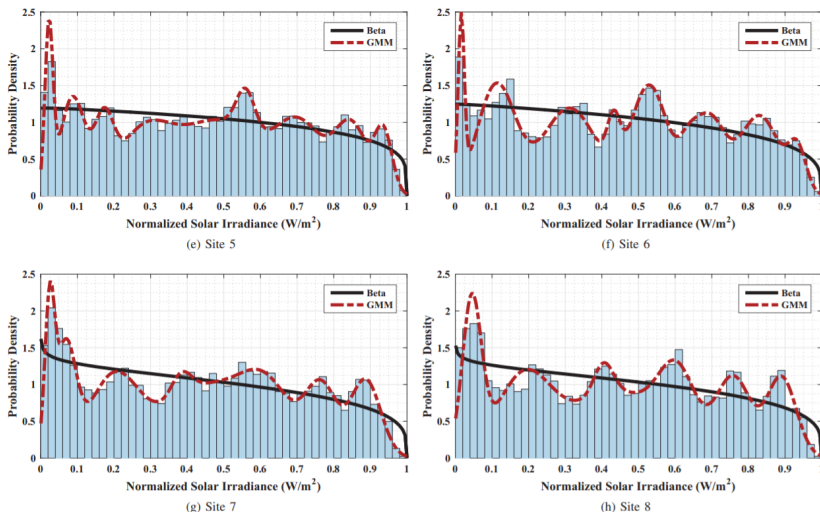


Figure: Probability Density plots and Histograms (sites 5-8)

Results and Performance evaluations

Analysing Figures

- 1 Figures 1 and 2 shows the normalized histograms (with density scaling) of NSI data at the eight selected sites, overlaid with the two density estimates.
- 2 Visually, these plots show substantial discrepancies between the data histogram and the Beta distribution.
- 3 As depicted, the proposed GMM provides a more accurate estimate of NSI probability density.

Theoretical Performance measures

- 1 Coefficient of determination, R^2 index
- 2 A goodness-of-fit test: the Kolmogorov–Smirnov (K–S) test
- 3 Three standard error measures: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE).

Results and Performance evaluations

K-S test

- 1 Each K-S test returns a p-value which, informally, is a measure of the evidence against the null hypothesis.
- 2 When the observed data set does not provide sufficient evidence validating the hypothesis that it is associated with the statistical model, the K-S test rejects the null hypothesis at the α_{K-S} significance level.
- 3 The proposed BIC-assisted GMM applying the EM algorithm produces p-values which indicate a failure to reject the null hypothesis at the α_{K-S} significance level for all sites.
- 4 This suggests that the proposed model is a more robust approach for estimating solar GHI data.

Results and Performance evaluations

Formulas

In addition to the goodness-of-fit test, the performance of the proposed GMM PDF is evaluated against f_{Beta} using the R^2 index, and three standard measures: MAE, MAPE and RMSE given by

$$R^2 = 1 - \frac{\sum_{i=1}^t (y_i - \hat{y}_i)^2}{\sum_{i=1}^t (y_i - \bar{y})^2} \quad (7)$$

$$MAE = \frac{1}{t} \sum_{i=1}^t |y_i - \hat{y}_i| \quad (8)$$

$$MAPE = \frac{1}{t} \sum_{i=1}^t \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (9)$$

$$RMSE = \sqrt{\frac{1}{t} \sum_{i=1}^t (y_i - \hat{y}_i)^2} \quad (10)$$

Results and Performance evaluations

Formulas contd.

where t is the number of data bins chosen using the square root rule, y_i is the probability of solar GHI data being within bin i calculated from the data set, \hat{y}_i is the probability within the same bin calculated from the estimated data set, $\bar{y} = \frac{1}{t} \sum_{i=1}^t y_i$ and $i = 1, 2, \dots, t$.

Result

- 1 The computed values of R^2 , MAE, MAPE and RMSE values with error percentage improvements computed with respect to the Beta distribution for all sites.
- 2 The proposed GMM produces the highest R^2 indices and the lowest error values for all eight sites.
- 3 So this shows that the GMM is a more robust estimate than the popular Beta distribution for estimating solar irradiance probability density.

Conclusion

Points

- ➊ A Gaussian Mixture Model (GMM) is proposed in this paper for estimating solar irradiance probability density at a given site.
- ➋ The proposed GMM parameters are obtained using the Expectation–Maximization (EM) algorithm
- ➌ The optimal number of mixture components is determined using the Bayesian Information Criterion (BIC).
- ➍ The proposed GMM produces the best performance metrics and the highest percentage improvements providing a more accurate estimation of solar irradiance compared with the popular Beta distribution.