# Uncovering Optimal Solar Site Locations in India using Unsupervised Learning Approaches

Smitesh Nitin Patil

June 19, 2023

## Contents



## List of Figures



## List of Tables

# 1 Introduction

The process of transitioning the global energy supply from fossil fuel-based sources to sustainable energy sources like wind and solar will be crucial for mankind to make in the 21st century. Solar energy is generated from photovoltaic cells which need a high amount of solar irradiance throughout the year to be profitable. Countries in tropical regions like parts of India tend to receive abundant sunlight throughout the year. However, many aspects need to be studied before identifying promising regions where solar farms could be built to tap into that region's solar potential, like the slope gradient of the terrain, proximity to urban centers, and nature, wildlife preserve areas.

In order to identify sites for setting up renewable energy plants various studies have been carried out at various location. Most of the studies use Analytical Hierarchy Process (AHP) a Multi-Criteria Decision-Making (MCDM) technique. This study proposes a novel technique of using Unsupervised approaches to identfy suitable locations.

# 2 Literature Review

There have been extensive studies carried out by researchers for identification of geo-locations to setup solar farms using various approaches.

## 2.1 Optimal site selection for solar photovoltaic (PV) power plants using GIS and AHP A case study of Malatya Province, Turkey

In[5] Colak et al carried out a study to identify suitable locations for setting up of photovoltaic power plants in Malatya province in Turkey. The authors use 11 layers of GIS data for identifying suitable sites. These layers included factors that affect the decision making process like:

1. Solar energy potential: Identification the solar potential of a region is of utmost importance as it determines the amount of energy producted by the region if a photovoltaic power-plant is set there.

2. Slope: The slope of the terrain is an important factor as a solar pv plant requires a even terrain for setting up the PV panels.

3. Transformer centers and energy transmission: Carrying electricity over huge distances without energy infrastructure results in loss of energy output through leakage thus there should be a power transmission system inplace for developing a new plant.

4. Land Cover: The land designated for nature reserves, tribal population, and other uses cannot be used for energy generation by law thus this factor must be considered beforehand.

5. Residential areas: Building a solar plant near a urban center can be detrimental for the region as urban sprawl tends to expand with time, On the other hand if a PV solar plant is close by a urban center it would lead to less transmission loss this tradeoff needs to considered.

For data preprocessing, various hardset conditions were set to restrict certain areas like slope elevation of land cannot be more than 20 percent, distance to road, rail network should be more than 0.1 km, no residential areas nearby and proximity to energy transmission network. The classification problem was solved using AHP one of the MCDM methods developed in 1977. In AHP each factor, a data layer in this case is assigned a weighted value bases on a table value that indicates the suitable value of the layer. The resulting suitable area was foundby multiplying the weighted average of each layer to find the optimal location.

As a result of using AHP, 34 suitable locations were found in the region corresponding to the districts of Yarimcahan, Karakasciftligikoyu, Karakoy and

Goller. Presence of a pre-existing PV plant in the region strengthen the methodology proposed in this Paper.

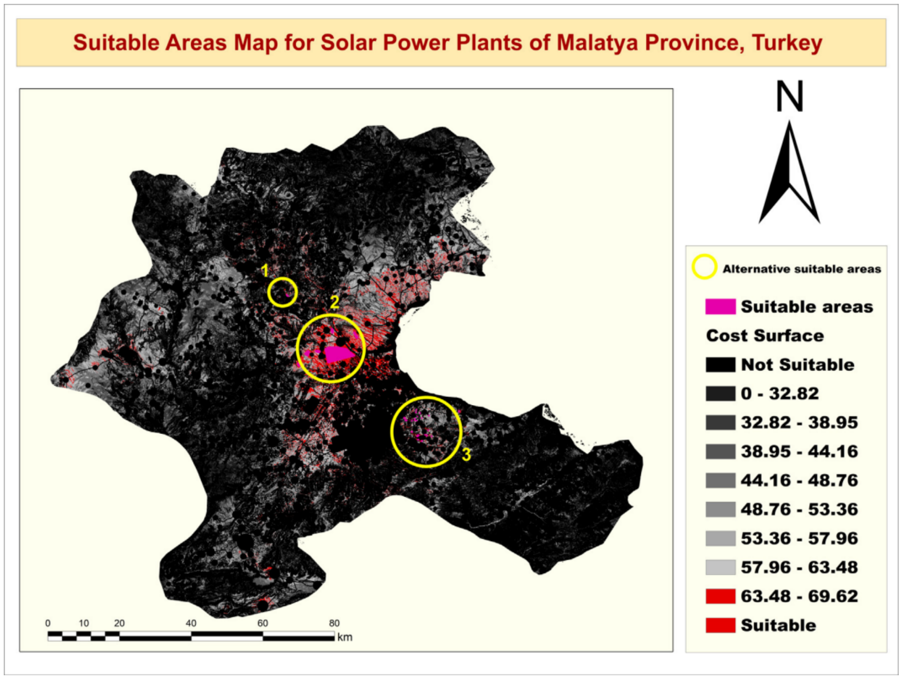


Figure 1: Suitable sites for Solar Powerplant with cost Factor

## 2.2 Solar PV power plant site selection using a GIS-AHP based approach with application in Saudi Arabia

In[1] Al Garni et al try to identify sites for solar plant selection by categorising the available land in five categories, i.e. least suitable, marginally suitable, moderately suitable, highly suitable, most suitable. The decision-making process for site selection was done in three phases.

1. Setting decision criteria and restriction for site selection study.
2. Prioritizing certain sites with high solar potential.
3. Analysis on the prioritised region for decision making

Like the previous study, authors here have taking GIS data formulated by National Renewable Energy Laboratory (NREL) and selecting attributes that determine criteria for site selection which include Digital Elevation Model (DEM), Solar irradiation, Air Temperature. These factors could be divided into roughy two categories technical (factors affectiong energy production) and economical (factors affecting economical viability of the project). For identification of suitable sites similar approach was used as in[5] by Colak et al. AHP was used to determine weights and each criteria was assinged a weight that were added together as a raster to form a map that identified suitable location in aforementioned five categories.

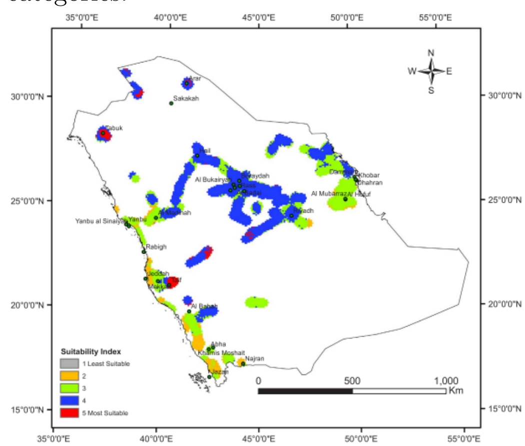

Figure 2: Suitability Index results using AHP weights in[1]

### 2.3 Optimization solar site selection by fuzzy logic model and weighted linear combination method in arid and semi-arid region: A case study Isfahan-IRAN

In[2] Zoghi et al use probabilistic models in AHP a MCDM to determine sites for PV plants in Isfahan province, Iran. Again, various factors were considered for identification of high potential locations like

1. Enviromental: Landuse, Protected Areas, Wetlands and Water Resource.
2. Geomorphological: Elevation, Slope, Aspect.
3. Location: Distance to City, Distance to Power line, Distance to Transport network.
4. Climatic: Sun shine, Cloudy Days, Dusty Days, Solar Radiation, Rainy and Snowy Days, Humidity.

Fuzzy logic was using for determining optimal location, each GIS layers consisting of the above listed factors was assigned to a memebership function that is bounded by a value between 0 and 1 i.e. a fuzzy set membership grade. Fuzzy Weights was assinged to the factors based on their relevance and Weighted Linear Combination (WLC) was calculated which provided the feasbility of location in terms of power generation.

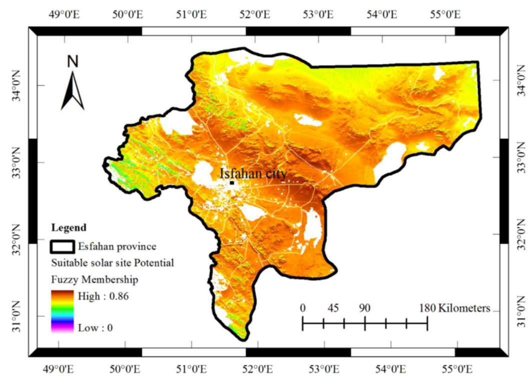


Figure 3: Suitability Index results using fuzzy logic system in[2]

### 2.4 MCDM and GIS based modelling technique for assessment of solar and wind farm locations in India

This study is carried out by[3] for whole of India in 2021. Again AHP was used for classification of suitable land for solar power plant setup. Data was sourced from various govermental bodies like NREL for solar radiation, DIVA-GIS for roads, inland water bodies, DEM model used was provided biblography United States Geological Survey (USGS). The factors were divided into three categories technical, socio-enviromental and economical.

1. Technical: Solar Radiation, Slope, Aspect, elevation.
2. Socio-Enviroment: Distance from coastline, Distance from waterbodies, airports, Landuse.
3. Economic: Distance from urban areas, roads, transmission lines, power plants.

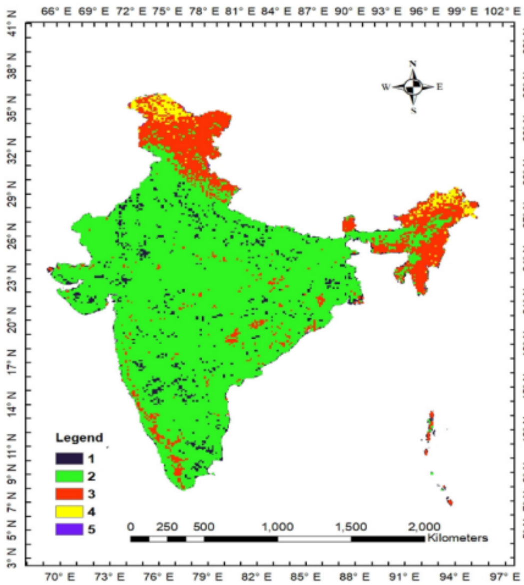


Figure 4: Solar farm sustainability map in[3]

A major limitation of the study lies in the data as the spatial resolution of data (around 1000m) is really less for DEM modelling and other attributes. As a result the solar farms suitablity map generated in this study is not intricate on a spatial level. There are various databases available with spatial resolution of 30m from USGS and NREL that can be utilised to make more precise prediction.

### 2.5 Landslide Susceptibility Prediction Based on Remote Sensing Images and GIS: Comparisons of Supervised and Unsupervised Machine Learning Models

This study differs from the previous papers discussed as it uses Unsupervised and Supervised learning techniques against the MCDM techniques used in previous papers. Also the problem statement that the authors [4] Zoghi et al try to solve is different than the previously referenced studies. Landslide susceptibility of a region is identified in this study using GIS data and applying Supervised learning techniques like Support Vector Machine (SVM) and CHi-squared Automatic Interaction Detection (CHAID) and Unsupervised learning techniques like K-Means and Kohonens Model. For Landslide prediciton various hydrological factors like Terrain Wetness Index (TWI), Slope, Relief Map, and Drainage density were considered.

For Unsupervised Learning, K-means clustering was performed by setting up the number of clusters to five for the five categories very high susceptibility, high, moderate, low, very low. By getting the euclidean distance for each datapoint we can assign them to clusters for a certain number of iterations by updating the centroids during each iteration.

Kohonens model is a type of feed-forward Artificial Neural Network (ANN) that specialises in Unsupervised learning. It can be used for reducing dimentionality and creating vector embeddings. These embeddings could be used for identification of susceptable areas by clustering.

The authors found the Supervised learning algorithm performed better that Unsupervised learning techniques. But the Supervised learning algorithms require labelled data, thus Unsupervised learning algorithms are cost effecting as it doesn't require annotation.

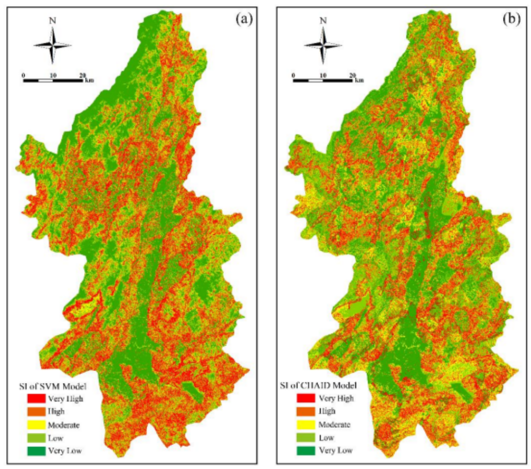


Figure 5: Landslide susceptibility in Ningdu county of Jiangxi Province[4]

# References

[1] H. Z. Al Garni and A. Awasthi, "Solar pv power plant site selection using a gis-ahp based approach with application in saudi arabia," *Applied Energy*, vol. 206, p. 1225–1240, 2017.

[2] M. Zoghi, A. H. Ehsani, M. Sadat, M. javad Amiri, and S. Karimi, "Optimization solar site selection by fuzzy logic model and weighted linear combination method in arid and semi-arid region: A case study isfahan-IRAN," *Renewable and Sustainable Energy Reviews*, vol. 68, pp. 986–996, Feb. 2017. [Online]. Available: https://doi.org/10.1016/j.rser.2015.07.014

[3] S. Saraswat, A. K. Digalwar, S. Yadav, and G. Kumar, "MCDM and GIS based modelling technique for assessment of solar and wind farm locations in india," *Renewable Energy*, vol. 169, pp. 865–884, May 2021. [Online]. Available: https://doi.org/10.1016/j.renene.2021.01.056

[4] Z. Chang, Z. Du, F. Zhang, F. Huang, J. Chen, W. Li, and Z. Guo, "Landslide susceptibility prediction based on remote sensing images

and GIS: Comparisons of supervised and unsupervised machine learning models," *Remote Sensing*, vol. 12, no. 3, p. 502, Feb. 2020. [Online]. Available: https://doi.org/10.3390/rs12030502

[5] H. E. Colak, T. Memisoglu, and Y. Gercek, "Optimal site selection for solar photovoltaic (pv) power plants using gis and ahp: A case study of malatya province, turkey," *Renewable Energy*, vol. 149, p. 565–576, 2020.