YZV202E/Optimization for Data Science Project Report Optimization-Based Placement of Solar Panel Installations in Turkey

Muhammed Abdullah Özdemir
AI and Data Engineering
Istanbul Technical University
ozdemirmuh22@itu.edu.tr
150220340

Muhammed Furkan Şıhanoğlu
AI and Data Engineering
Istanbul Technical University
sihanoglu22@itu.edu.tr
150220301

***Abstract*—This project focuses on selecting the best locations for solar panel installations in Turkey by using optimization methods. Since energy has become a strategic global issue, producing clean energy locally is especially important for developing countries like Turkey. We built a mixed-integer linear programming (MILP) model to decide where to install solar panels and how to connect them to the electricity grid. To handle larger datasets and speed up the process, we also used heuristic methods like Genetic Algorithms and Simulated Annealing. The model is based on real-world data such as solar irradiance, land use, installation costs, and grid capacities. Our goal is to maximize energy output while keeping the costs low. The results include optimal sitemaps and cost-performance comparisons, which can help guide solar energy planning in Turkey.**

I. Introduction

Renewable energy is becoming more important every year as countries aim to reduce their dependence on fossil fuels and lower energy costs [1]. Among these, solar energy stands out as one of the most accessible and clean energy sources. Turkey, being geographically well-positioned, has high solar potential compared to many European countries [2]. However, this potential is still far from being used efficiently on a national scale. While there are individual efforts to install solar panels, large-scale planning and optimization remain limited. We believe that with better planning, Turkey can lead in solar energy use and significantly cut down on energy-related expenses [4].

In this project, we had the chance to apply the optimization techniques we learned in class to a real-life problem that caught our interest. By combining technical and economic factors, we aimed to find the most suitable locations for solar panel installations across Turkey.

Our main goal was to reduce energy costs and support energy independence for our country. We designed a model that considers both installation expenses and energy generation capacity to select optimal locations. To do this, we used a Mixed-Integer Linear Programming (MILP) approach. We also tested heuristic methods like Genetic Algorithms and Simulated Annealing to compare performance on larger datasets.

II. Problem Formulation

The solar panel placement problem is modeled as a facility location optimization problem. We aim to select the best locations for photovoltaic (PV) installations and connect them to existing substations to maximize profit while meeting technical and geographical constraints.

A. Sets and Parameters

- I: Set of candidate locations for PV installations
- J: Set of existing grid connection points (substations)
- C_i^{inst}: Installation cost at location i
- C_{ij}^{conn}: Cost to connect location i to substation j
- E_i: Estimated annual energy output (kWh) from location i
- P^{rev}: Revenue per kWh of generated solar energy
- R_i: Annual solar irradiance at location i
- R_{min}: Minimum required solar irradiance
- C_i^{cap}: Capacity of a PV plant at location i (MW)
- C_j^{grid}: Maximum capacity of grid point j (MW)
- D_{ij}: Distance between location i and substation j
- D_{max}: Maximum allowable connection distance
- L_i: Land use restriction indicator (1 if restricted, 0 otherwise)

B. Decision Variables

- $x_i \in \{0, 1\}$: 1 if a PV plant is installed at location i; 0 otherwise
- $y_{ij} \in \{0, 1\}$: 1 if plant at location i is connected to grid point j; 0 otherwise

C. Objective Function

We aim to maximize the total annual profit:

$$\max Z = \sum_{i\in I}\sum_{j\in J}\left(P^{\text{rev}} \cdot E_i \cdot y_{ij}\right) - \sum_{i\in I}\left(C_i^{\text{inst}} \cdot x_i\right) - \sum_{i\in I}\sum_{j\in J}\left(C_{ij}^{\text{conn}} \cdot y_{ij}\right)$$

D. Constraints

- **Solar Irradiance Constraint:**

$$R_i \cdot x_i \geq R_{\text{min}} \cdot x_i \quad \forall i \in I$$

- **Land Use Restriction:**

$$L_i \cdot x_i = 0 \quad \forall i \in I$$

- **Connection Constraint:**

$$\sum_{j \in J} y_{ij} = x_i \quad \forall i \in I$$

- **Distance Constraint:**

$$D_{ij} \cdot y_{ij} \leq D_{\text{max}} \quad \forall i \in I, j \in J$$

- **Grid Capacity Constraint:**

$$\sum_{i \in I} C_i^{\text{cap}} \cdot y_{ij} \leq C_j^{\text{grid}} \quad \forall j \in J$$

- **Binary Constraints:**

$$x_i, y_{ij} \in \{0, 1\} \quad \forall i \in I, j \in J$$

This formulation ensures that selected locations are economically viable, technically feasible, and within allowable connection and land use limits.

III. Dataset Description

To support the optimization model with realistic data, we collected and preprocessed multiple open-access spatial and infrastructure datasets from reliable sources.

- **GADM:** Provided Turkey's national and provincial administrative boundaries (shapefiles) to define the study area and align all geospatial layers.
- **OpenStreetMap (OSM):** Supplied geospatial point data for transformer substations across Turkey. This dataset was further refined using official substation counts from TEİAŞ.
- **TEİAŞ:** A PDF document listing the number and capacity of transformer substations by province was used to validate and complete missing grid connection data.
- **Copernicus CORINE Land Cover (CLC):** Delivered high-resolution land cover classification data. Only scrublands, sparsely vegetated areas, and natural grasslands were considered suitable. Urban areas, agricultural land, forests, wetlands, and water bodies were masked out.
- **Global Solar Atlas / Solargis:** Provided Global Horizontal Irradiation (GHI) raster files, which were aggregated into daily and annual mean values to estimate solar energy yield for each grid cell.

Data Preprocessing: All datasets were reprojected to EPSG:3857 and clipped to Turkey's national boundaries. Grid geometries were generated at 1 km² resolution. Raster layers were aggregated and standardized. Attribute tables were cleaned to remove irrelevant or missing data.

Land Use Filtering: CORINE-based filtering was applied using specific CLC codes. Only non-urban, non-agricultural, and non-protected lands with suitable terrain were retained.

Distance and Capacity Calculations: The geometric centroid of each candidate grid cell was computed using GeoPandas. The nearest substation was identified via a KDTree search (scikit-learn), and the Euclidean distance was calculated. Connection cost was derived from this distance using a fixed per-meter rate. Capacity per grid cell was estimated using:

$$\text{Capacity} = \text{Area} \times \text{Panel Density} \times \text{Efficiency}$$

IV. Methodology

The project methodology is composed of two main stages: (1) geospatial data preprocessing and (2) optimization model implementation using Mixed-Integer Linear Programming (MILP).

A. Data Preprocessing

All geospatial and technical datasets were cleaned, clipped to Turkey's national boundary, and reprojected to EPSG:3857. A 1 km² uniform grid was generated, where each cell represents an area of approximately 1000m x 1000m. Raster and vector data were overlaid to extract candidate zones based on solar irradiance, land use classification, and slope constraints.

The centroid of each candidate cell was computed using GeoPandas. The nearest transformer substation was identified using a KDTree search (scikit-learn), and Euclidean distance was calculated. These distances were used to estimate connection costs at a rate of \$30 per km.

B. MILP Model Implementation

We formulated the solar placement problem as a facility location optimization model and implemented it using the Gurobi Optimizer through the `gurobipy` Python interface. Default solver parameters were used (e.g., presolve, heuristics, and thread count).

The objective was to maximize total annual profit while ensuring feasibility of site selection and grid connections. The model incorporated the following key parameters:

- Maximum connection distance: $D_{\text{max}} = 20$ km
- Installation cost: $C_{\text{inst}} = \$300{,}000$/MW
- Grid expansion cost: $C_{\text{exp}} = \$180{,}000$/MW
- Land cost: \$3,000,000 per km²
- O&M cost: \$12,000 per MW per year
- Panel efficiency: 18%
- Performance ratio (PR): 80%
- Degradation rate: 0.5% per year (over 25 years)

The installed PV capacity for each cell was estimated based on land area (0.5 km²), panel coverage ratio (0.5), and efficiency. Energy yield was computed using average Global Horizontal Irradiance (GHI), system performance ratio, and degradation assumptions.

C. Runtime and Scalability

The model was tested on a local machine (MacOs M4 CPU, 16GB RAM and Google Colab). Solution times varied depending on the number of candidate cells:

- Test Cases: 1–3 minutes

This confirmed that the MILP model is scalable for medium to large datasets within reasonable time limits.

V. Experimental Evaluation

The MILP model was tested on a national scale to determine optimal locations for solar panel installation. The analysis focused on energy potential, cost structure, infrastructure compatibility, and long-term financial viability.

A. Grid Selection and Regional Distribution

A total of 17,206 grid cells were selected across Turkey, each corresponding to 1 km² of suitable land. Visual inspection of the selected areas reveals that most optimal sites are concentrated in the southeastern and central Anatolian regions, which offer high solar irradiance and low slope gradients. Regions such as Şanlıurfa, Konya, and Niğde displayed high selection density, confirming their solar suitability.

B. Energy Output and Capacity Distribution

The estimated total annual energy generation from all selected locations is 2.97 TWh. Over 25 years, accounting for panel degradation (0.5% /year) and system losses (PR: 80%), the projected yield is approximately 74.3 TWh. This demonstrates that large-scale solar planning can contribute significantly to national energy needs. Figure shows the distribution of total PV capacity assigned to each substation.

C. Cost Analysis and Key Drivers

The total estimated cost of the project is $5.10 billion, composed of installation ($1.39B), connection ($1.26B), land ($1.55B), and O&M costs ($464M). Notably, land cost emerged as the single largest driver, especially in western Turkey where suitable areas are limited and expensive. Figure summarizes this distribution. These insights suggest that optimizing land selection in tandem with solar potential is crucial.

D. Financial Viability

The model results indicate a 68% Return on Investment (ROI), with a Levelized Cost of Energy (LCOE) of 69.1 USD/MWh. Compared to IEA benchmarks (60–80 USD/MWh for utility-scale solar), this result is competitive. The payback period is estimated at 14.8 years, aligning with standard PV system lifespans and typical Power Purchase Agreement (PPA) terms. As shown in Figure , the project generates a net profit of $3.49B over 25 years.

E. Infrastructure Constraints and Realism

Due to limitations in transformer location data, substation coordinates were synthetically inferred from TEİAŞ province-level reports. This approximation may introduce spatial inaccuracies. Figure shows the final site-to-substation mapping. In future versions, access to official transformer geolocations would enable a more precise connection model and potentially reduce total connection costs by avoiding misaligned pairings.

F. Heuristic Performance Comparison

To evaluate scalability and alternative solution quality, Simulated Annealing (SA) and Genetic Algorithm (GA) methods were benchmarked against the MILP solution. SA yielded a total net gain of $2.23B with 3,161 selected grid cells, while GA achieved $1.85T net gain with 23,997 selected cells. Though MILP remained the most precise method for small-to-medium datasets, GA showed strong potential in large-scale scenarios. SA offered moderate gains with faster convergence.

These findings indicate that heuristic algorithms can approximate optimal solutions when MILP becomes computationally infeasible, offering a viable trade-off between solution quality and runtime efficiency.

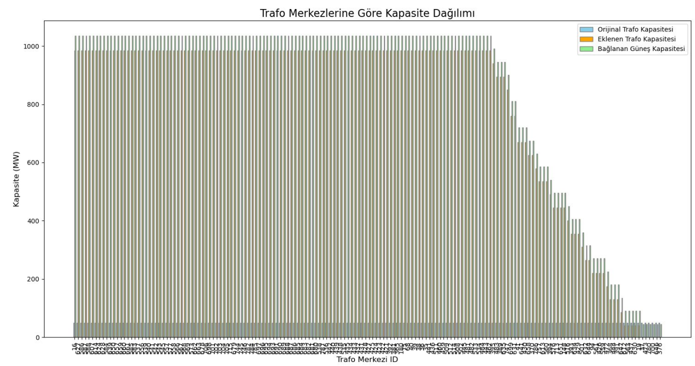

Fig. 1. Capacity distribution across transformer stations.

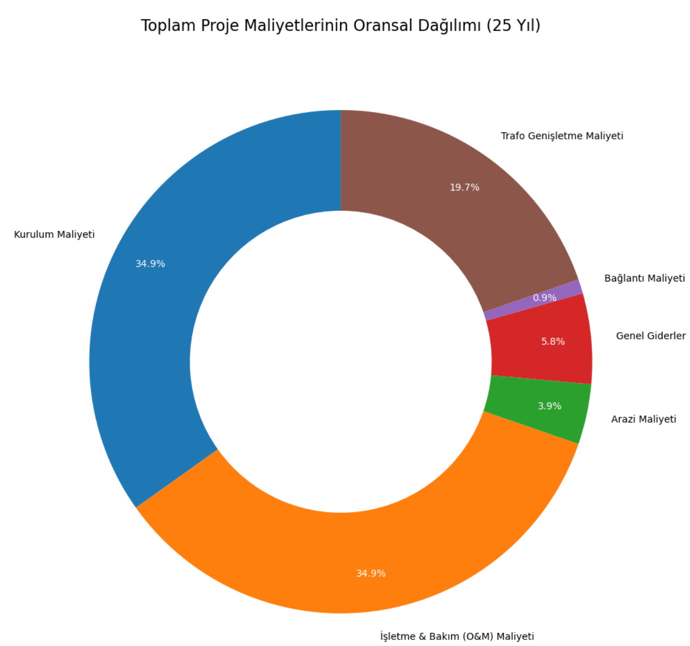

Fig. 2. Project cost breakdown over 25 years.

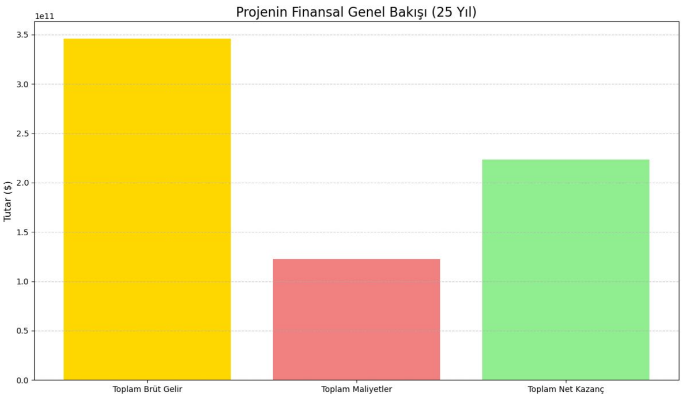

Fig. 3. Financial summary: gross revenue, total cost, net gain.

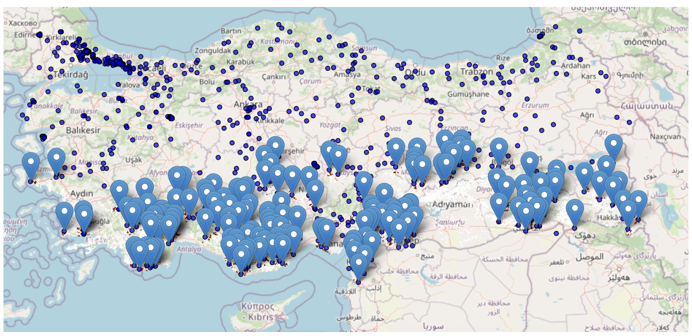

Fig. 4. Map of selected PV installation zones and substations.

VI. Conclusion

In this project, we tackled the problem of identifying optimal locations for large-scale solar panel installations across Turkey by combining geospatial analysis with mathematical optimization. Using real-world data and a MILP model implemented with Gurobi, we managed to generate a plan that maximizes energy output while minimizing overall costs. This experience taught us how critical data-driven decision-making is when dealing with national-level energy planning problems.

One of the most valuable takeaways was realizing that energy optimization is not just a technical issue, but also an economic and geographic challenge. Working with real datasets, adapting constraints to reflect land use and grid limitations, and interpreting results made the learning process both practical and exciting.

In future versions of this project, we hope to improve the accuracy of our model by accessing official transformer location data. Due to data restrictions, we had to rely on synthetic substation coordinates, which limits the realism of our results. With potential government support or formal data access, integrating verified grid infrastructure could make the model significantly more powerful and applicable in real-life planning.

References

[1] International Energy Agency, "Solar Power — Analysis," IEA, 2023. [Online]. Available: https://www.iea.org/reports/solar-power

[2] Global Solar Atlas, "Global Horizontal Irradiation (GHI) Data," World Bank Group, 2022. [Online]. Available: https://globalsolaratlas.info/

[3] Copernicus, "CORINE Land Cover 2018," European Environment Agency (EEA), 2018. [Online]. Available: https://land.copernicus.eu/pan-european/corine-land-cover/clc2018

[4] TEİAŞ, "Electricity Transmission Infrastructure Report," Turkish Electricity Transmission Company, 2023. [PDF]. Available from: https://www.teias.gov.tr

[5] OpenStreetMap, "Transformer Substation Data," 2023. [Online]. Available: https://www.openstreetmap.org

[6] Copernicus, "EU-DEM v1.1 — Digital Elevation Model," European Union, 2020. [Online]. Available: https://land.copernicus.eu/imagery-in-situ/eu-dem

[7] Gurobi Optimization, LLC, "Gurobi Optimizer Reference Manual," 2024. [Online]. Available: https://www.gurobi.com/documentation/

[8] International Energy Agency, "Projected Costs of Generating Electricity 2020," IEA/NEA, 2020. [Online]. Available: https://www.iea.org/reports/projected-costs-of-generating-electricity-2020