|  |  |
| --- | --- |
| Multi-Plant Photovoltaic Energy Forecasting Challenge |  |

### The challenge is closed. Thank you for your participation!

### The winners will be soon contacted via email with the instructions to redeem their prize.

### [Final leaderboard](http://193.204.187.201:8080/pv_challenge_website/leaderboard_final.php)

### [Temporary leaderboard](http://193.204.187.201:8080/pv_challenge_website/leaderboard_temp.php)

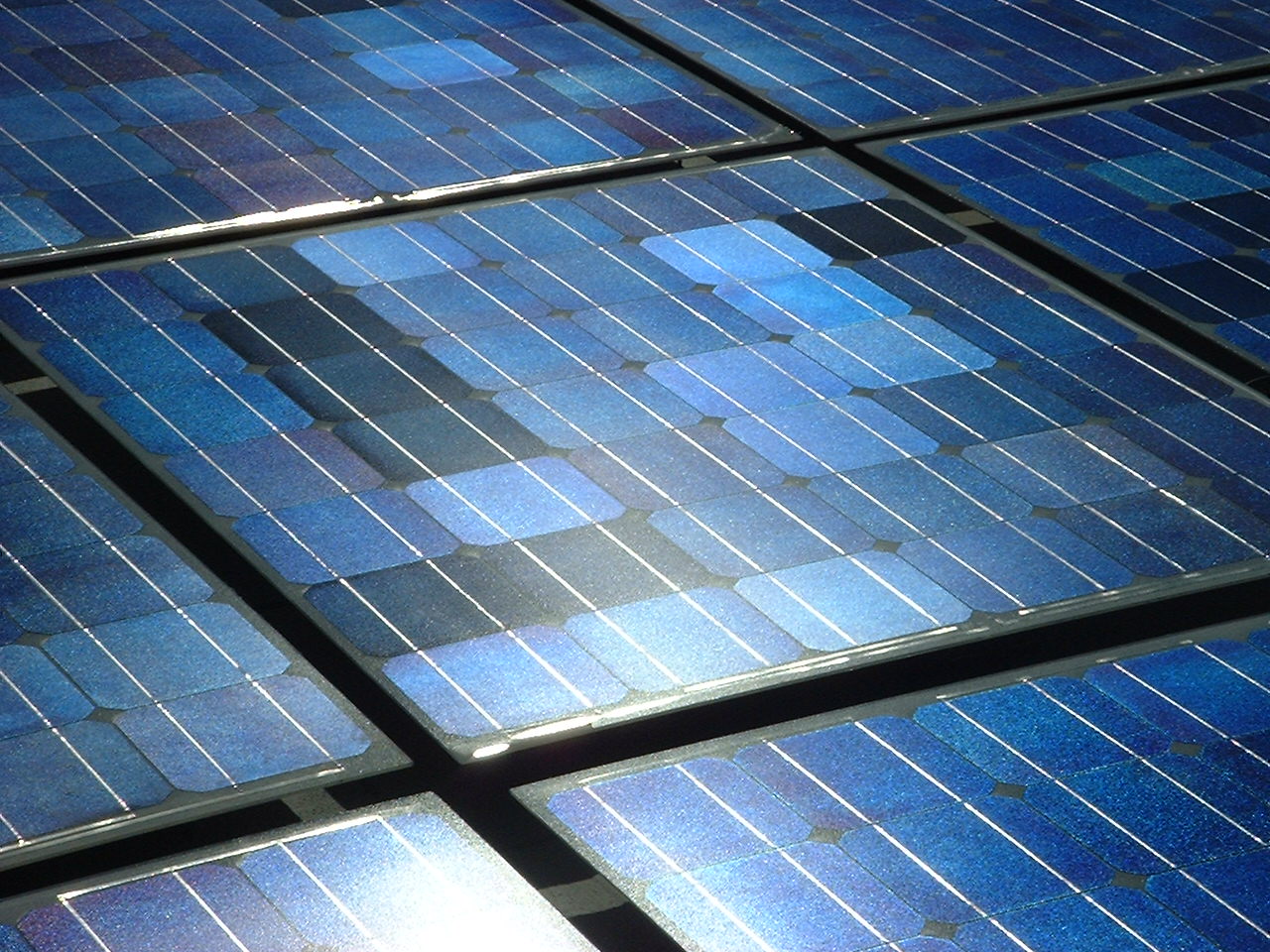
### Introduction

The urgent need to reduce pollution emission has made renewable energy a strategic European Union (EU) and international sector. This has resulted in an increasing presence of renewable energy sources and thus, significant distributed power generation. The main challenges faced by this new energy market are grid integration, load balancing and energy trading.

In order to face these challenges, it is of paramount importance to monitor the production and consumption of energy, both at the local and global level, to store historical data and to design new, reliable prediction tools.

In this challenge, we focus our attention on photovoltaic (PV) power plants, due to their wide distribution in Europe. During the last years, the forecast of PV energy production has received significant attention since photovoltaics are becoming a major source of renewable energy for the world.

Forecast may apply to a single renewable power generation system, or refer to an aggregation of large numbers of systems spread over an extended geographic area.

### Task & Dataset

The task proposed in this challenge is power forecasting for multiple photovoltaic (PV) plants spread over a defined geographical area and connected to a power grid.

The provided dataset consists of time series data regarding weather conditions and production collected by sensors on three closely located PV plants in Italy.

Each time point is the hourly aggregation obtained as the average of all the measures available in a specified hour.

For each plant, day, and variable (such as temperature, irradiance, cloud coverage, etc.), data consists of a time series of 19 values representing hourly aggregated observations (plants are active from 02:00 to 20:00).

Given multiple time series of meteorological conditions for a set of plants and for a specified day in the future, the goal is to predict the time series of production (power) for each plant and for the entire day, at hourly granularity.

Training data spans over a temporal period of 12 months (year 2012) including the daily target time series (power observed for each plant), whereas testing data consists of 3 months (January to March 2013) for which the target time series (power) is not provided.

Predictive models will be evaluated using the standard Root Mean Square Error (RMSE) measure.

Note that sensor data can contain reasonable missing values (zero valued measurements) or outliers.