RESEARCH ARTICLE   OPEN ACCESS

A solar generation prediction-based method for optimal activity and battery scheduling

JYOTHI WADGERI

Thermal Engineering
MVJ College Of Engineering
jyothiwadgeri@gmail.com

ABSTRACT
Little solar panels and other parts make up solar home systems (SHS), which are designed to power a single house. For those living without access to a national power grid in developing countries, this is a more financially viable option. To prioritizing the deployment of public and private resources and tracking the fulfilment of universal electrification objectives, it is crucial that stakeholders have access to reliable data on individual SHS installations, such as information like position and power capacity. Despite the proliferation of research that using satellite imagery and computer vision to detect solar panels, many of these tool's struggle to properly locate numerous SHS due to poor image quality. Here, we analyze the cost-performance tradeoff of using automatic SHS identification using UAV data instead of satellite imagery.
We look at three specific issues: I how reliable is SHS detection using drones; (ii) how much does it cost to acquire drone data compared to satellite photos; and (iii) how much does it cost to obtain drone data? To test the capacity of deep learning models to maybe other SHS, even those that are too tiny to be consistently identified in satellite imagery, we gather and make accessible high-resolution drone images of SHS captured under a broad variety of real-world settings. The results suggest that UAV photography may be an alternative to identify very small SHS from the perspectives of both accuracy rate and economic expenses of data collecting. Data obtained by UAVs might potentially be used to aid power access planning strategies, which would help accomplish sustainable development goals and monitor progress towards those goals.
Keywords: *Electrification, Solar Home System, Deep Learning, Machine Learning, Cost Optimization*

I. INTRODUCTION

The objective of the competition [1] was to create a baseload, scheduling constraint, and Solar Home System cost prediction-aware algorithm for the month of November 2012. Python, which has tools for forecasting and optimization, was chosen as the language to develop the solution. Python helped simplify the connection between these two processes. As a first step, we used a visual representation of the available building load and solar production data to identify gaps in the data and identify repeating patterns that could indicate seasonal or cyclical changes. It was discovered that there were less blanks in the solar power generating data, as opposed to the construction data, which was riddled with missing information. This meant that proper data cleaning techniques like imputation and deletion have to be used to make the data viable for prediction analysis.

Considering the obvious dissimilarities between the construction and solar patterns, two distinct families of prediction tools were created. The results of several studies comparing various approaches of forecasting reveal that ensemble methods are superior to those of the individual methods. As a result, the vote regressor from the Python sklearn package was used to make load predictions for the buildings. This regressor model, instead of making predictions using a single estimator, uses an average of estimates from many estimators fitted to the same data. The best results for this dataset came from a combination of tree-based approaches like random forest (RF) and the gradient boosting (GB) technique. In addition to these approaches, STL decomposition was included to better forecast the cyclic and seasonal fluctuations in building loads.

A cost-minimizing scheduling method was the focus of the competition's second round. Binary variables, such as whether time a certain job is active, are required to represent the restrictions of this scheduling issue. As this was an integer-based issue, it was modelled using mixed-integer programming (MIP). It was clear from the problem statement that there would be many obstacles to overcome.

- For planning out a month's worth of events, a granularity of 15 minutes requires vectors with a size of 2880.

Increases in activity level have a multiplicative effect on complexity.

- A squaring component in the peak power cost makes this a mixed integer quadratic programming issue (MIQP)
- All events were to be held within business hours for maximum efficiency. The peak power period, which represents a significant proportion of the total energy expenditure, is impacted because of this as well.

These difficulties significantly impacted whether the issue could be solved. We also discovered that scheduling non-recurring operations yielded a maximum value of about sixteen thousand dollar which may not justify the additional expense and calculation necessary to plan such chores. So, the following two procedures were employed to reduce the complexity of the issue:

- The issue was modelled using just periodic actions.
- By capping the monthly peak power component and eliminating it from the target, the issue was transformed into a mixed integer linear programme (MILP).

Hence, the procedure was broken down into four stages: gathering and processing data, predicting building loads, anticipating solar production, and solving the optimum scheduling issue [4].

II. RELATED WORK

Large solar Pv systems may be visible in satellite imagery, and this is especially true in comparison to very small arrays. No previous studies have concentrated on smaller photovoltaic panels that are being deployed for all of those travelling to transition to access to energy for the first time, which may be 100W or less, but new studies demonstrates the opportunities of automatically charting Solar PV arrays using satellite technology and intelligent systems for facilities ranging from individual residential SHS (5-10kW [7] to utility-scale (>10MW [13]).

Traditional machine learning methods like vector machines (SVM), decision trees (DT), and random forests [7] weren't enough for early attempts in solar panel segmentation, therefore human engineers manually crafted characteristics like mean, variance, text ons, and colour statistics of pixels. Using deep neural networks (DNN), Yuan et al. [13] were able to determine which solar panels were located from above. Progress in using convolutional neural networks (CNNs) [12] on large-scale image datasets like ImageNet [12] has also boosted the field of solar panel segmentation (i.e., pixel-wise categorization). In [11], a CNN that had been taught to classify images was used to

conduct a rudimentary kind of solar panel segmentation. Two of the first real segmentation convolutional neural networks (CNNs) for PV recognition were SegNet [5] and activity map-based techniques [11]. To improve model identification performance [8, they also quickly embraced U-net designs [6].

For some time now, PV systems and solar farms have been monitored by unmanned aerial vehicles (UAVs). While theoretically feasible, in practise this has only been achieved for solar panel management in large solar farms where the locations of the solar PV have already been known. Unmanned air vehicle-based solar panel segmentation has been used to evaluate solar farm projects by using textural features and a clustering algorithm [8].

Infrared photography was used in previous drone investigations [13] of solar panel fault identification to locate specific issues with the arrays. There are some cases where satellite imagery can be used to monitor solar PV farms, such as when monitoring the impact of particulate matter deposition on generation efficiency [3]. However, this is an exception rather than the rule, as solar PV monitoring typically requires UAV data due to high good image requirements. In several cases, both optical and thermal cameras have been used. Thermal imaging cameras are used to look for temperature differences that may indicate damaged solar cells [4]. By analyzing optical UAV imagery, damaged or sand solar cells in solar and wind farms may be found [9]. These drone-based probes kept their photographic evidence to themselves.

III. DATASETS

Adding nuance to energy network research is an emerging area of study that involves describing data in an ontology paradigm so that the inherent relationships among data in an electricity network may be represented in human language. After thorough study, many ontologies for intelligent energy real applications have been developed. When it comes to improving energy demand and response, Daniele et al. [6] developed an ontology known as SAREF4EE. With the goal of improving coordination between different kinds of smart energy, **Lefrancois** [7] developed SEAS. Our work draws on prior ontologies for wise energy systems, but we highlight the ease with which cross-domain impacts (such the climate domain) may be included, as well as the importance of decentralized home energy systems. The climatic data collected utilized by our model to simulate the effects of climate change was defined

using both Wu's [8] CA and Janowicz's [9] SOSA ontologies.

To the best of our knowledge, there are no drone datasets including annotated photographs of solar panels, much alone extremely modest (100W) solar panels. The most important publicly accessible UAV-based datasets are summarized in the appendix; however, due to the absence of solar panel annotations, these datasets cannot be used to train automated systems to detect solar PV cells in UAV footage.

Altitude	GSD	#Img	#Vid	#Annoted PV
51m	1.8cm	56	8	216
62m	2.2cm	64	8	290
72m	2.6cm	48	9	235
82m	2.9cm	63	9	298
93m	3.4cm	43	9	215
101m	3.6cm	48	10	235
111m	4.1cm	57	5	290
121m	4.5cm	49	11	240

Table – 1:*Specifics about the data set used. Listed with the altitude is the appropriate GSD. In-Gamma: Pictures. videos, or "vids" for short. There are a total of 423 pictures, 60 videos, and 2010 solar panels with annotations.*

Specifics about the data set used. Listed with the altitude is the appropriate GSD. In-Gamma: Pictures. videos, or "vids" for short. There are a total of 423 pictures, 60 videos, and 2013 solar panels with annotations:

1. Sufficient resolution and resolution at ground sampling distance (GSD). In actuality, the GSD of drone images might fluctuate widely due to variables like technology and elevation shift. Thus, it is important for our dataset to include images with a variety of image GSDs that are adequate to simulate a range of real-world situations and enable the detection of SHS.

2. Panels of solar energy that are both varied and generic. Since the physical appearance of solar panels can vary depending on a number of factors (polycrystal vs. monocrystal, size, and aspect ratio), we took care in selecting our solar panels to ensure that they represent a wide range of potential configurations found in real-world installations in low-income regions.

3. 90-degree camera orientation and variable speeds: We'd want our dataset to include more than one flight speed so that we may examine the resilience of solar panel identification and the cost of data gathering (which relates to flying speed).

IV. SYSTEM MODELS

As a starting point for our investigation, we will use the micro grid system model shown in Fig. 1. This model examines the average home's contribution to the utility district's solar energy distribution network. The use of batteries to store electrical energy is also accounted for in the model. This system is managed by the Central Energy Manager (CEM) shown in the diagram. This domestic micro-grid may also, if necessary, take electricity from the grid, which is often believed to derive mostly from fossil-fuel resources, and at current market pricing. Residential power consumption per hour was calculated using real data collected from a utility serving homes in southwest Ohio.

These presumptions are put into practise. Connected to the load circuits are the solar generators and batteries; this arrangement makes it feasible to simultaneously draw electricity from all three sources (the grid, the batteries, and the generators). Second, it is assumed that a controller is available that, given advance notice of expected demands and solar output, can optimally adjust the proportion of power supplied by these sources at any given moment. The ideal power split between these sources at all times is the one that results in the lowest annual cost of energy provision in this configuration.
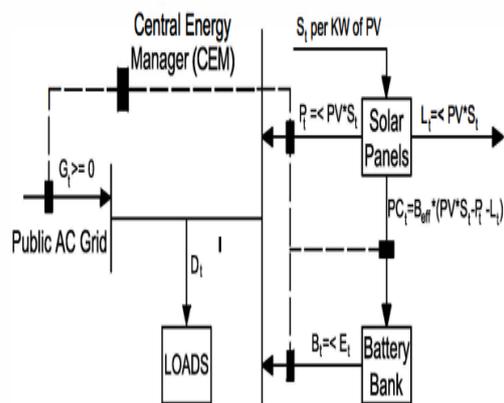
Figure No.1:*The System, shown in a single line.*

The optimization problem takes as inputs the average hourly power consumption by residential customers (Dt), the hourly cost of electricity (Ct), and the hourly solar energy output from the panels (kilowatt-volt-hours per square metre) (St). The cost of power rises in tandem with peak demand, as seen in the image. Battery charge and discharge efficiency, (Beff), battery and solar panel lifespan, and levelized cost are also considered.

A typical house in Dayton, Ohio has been simulated. Solar radiation received each hour was gathered from [18], whereas hourly demand and the cost of generating power were gathered from [3]. Relevant optimization problem variables are:

- P_t = Hourly power output from the PV to sustain the load directly.
- L_t = Excess power that cannot be used because there is not enough demand for electricity to fully use the available battery capacity.
- PC_t = Transfer of energy from the solar panels to the storage batteries.
- B_t = Discharge of Battery Power
- $PVsize$ = square metres of solar panel area.
- $Bsize$ = maximum capacity of the battery bank (kWh).

Using either the public power grid, solar panels, or a battery bank may fulfil this need, as seen in Fig. 1.

$$D_t = G_t + P_t + B_t \qquad - (1)$$

The energy generated by the solar panels is split into three categories at any given instant: that which is utilised to power loads (P_t), that which is used to charge the batteries (PC_t), and that which is wasted if neither PC_t nor P_t can be used (L_t). Input energy is proportional to the amount of sunlight, the efficiency of the solar panels, and the area covered.

$$S_t * PV_{size} * PV_{eff} = P_t + PC_t + L_t \qquad - (2)$$

The amount of energy available in the battery bank at any one moment (E_t) is proportional to the difference between the total amount of energy supplied to the system and the total amount of energy drawn from the batteries up to that point.

$$E_t = E_{t-1} + B_{eff} * PC_{t-1} - B_{t-1} + E_o \Rightarrow$$
$$E_t = \sum_{i=1}^{t-1} (B_{eff} * PC_i - B_i) + E_o$$
$$- (3)$$

where E_o is the amount of energy that was originally stored in the batteries.

V. COST OPTIMIZATION METHODOLOGY

After developing the model of the system, the major goal of this study is to determine how to implement real-time pricing in a way that minimizes the cost of delivering energy to a home. In essence, the goal of this optimization is to ascertain the optimal number of solar collectors and batteries (if any) to provide the client with the cheapest possible energy. The optimization considers the impact of variations in the capital expenditure for solar panels and batteries. The optimization issue is broken down into two stages for the sake of computation. The first stage is to determine the size PV system and battery capacity would provide the lowest power bill, and the second is to calculate the highest return on investment given that data. One-step optimization solutions would make the issue nonlinear. It was not feasible to identify the optimum solution in a single step due to the enormous quantity of input data and the length of time necessary for the optimization to execute. Using python programming, we have simulated the system and applied its optimizations.

5.1: Minimize Electricity Bill

With a certain size of solar panels and storage capacity, this section details the mechanism by which electricity costs are reduced. The following issues need to be addressed at this stage in light of the uncertainty surrounding costs, demand, solar radiation, and component efficiency: How much of the power produced by the solar panels should be sent straight to the loads and how much should be stored in the batteries at all times? When should the batteries be drained, for how long, and by what percentage?

This process identifies the optimal solar panel area and battery capacity to provide the lowest cost per unit of electricity per hour.

5.1.1: Objective Function:
Electricity cost for the period t is calculated as follows, taking into account the real-time price of power and the quantity of energy drawn from the grid at each instant in time:

$$C = \sum_{i=1}^{t} (G_t * C_t)$$

The goal is to find the lowest possible power cost for a certain time period, hence we must rewrite Eqns. 2 and 1 as follows.

$$G_t = D_t - DV_{size} * PV_{eff} * S_t + L_t + PC_t - B_t$$

The resulting objective function to be reduced is:

$$C = \sum_{i=1}^{t} \left((D_t - DV_{size} * PV_{eff} * S_t + L_t + PC_t - B_t) * C_t \right)$$

5.1.2:Constraints
a. Grid: G_t, the amount of electricity used from the grid, is expected to be non-zero at all times. In other words, the sun's rays will be utilised to either power appliances directly or charge the storage device. The following follows from Eq.

$$D_t - DV_{size} * PV_{eff} * S_t + L_t + PC_t - B_t \geq 0$$

b. Batteries: The energy contained in the batteries is non-zero in all cases. The restriction, using Eqn. 3, will be:

JYOTHI WADGERI. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 03, Issue 04, Jul.-Aug. 2013, pp. 2704-2710

$$\sum_{i=1}^{t-1} (PC_i - B_i) + E_o \geq 0$$

$$\Rightarrow B_{eff} * \sum_{i=1}^{t-1} PC_i - \sum_{i=1}^{t-1} B_i$$

$$\geq E_0$$

c. Size of the Battery: The capacity of the battery bank is restricted. The batteries can only hold so much power before they explode. The following restrictions are derived from Eqn.

$$E_t \leq E_{max} \Rightarrow B_t \leq B_{eff} * \sum_{i=1}^{t-1} PC_i - \sum_{i=1}^{t-1} B_i$$

$$\geq E_0$$

d. Battery Discharge: A battery can't discharge more energy in eachtime period than it has stored. When Eqn. 3 is considered, the following restriction emerges:

$$B_t \leq B_{max} \Rightarrow B_t \leq B_{eff} * \sum_{i=1}^{t-1} PC_i - \sum_{i=1}^{t-1} B_i$$

$$+ E_0$$

The optimization process has been repeated 72 times, once per hour. Hence, the CEM optimises the system based on the available stored energy at time t and the predicted data for the subsequent t + 72 hours. As soon as the optimization is complete, the system's state will be locked in for time t, and the subsequent optimization will cover the interval from t+1 to t+73. With this way, the system can regulate the charging and draining of the batteries autonomously.

VI.   MAXIMIZING INVESTMENT RETURN

By determining the minimal electricity bill (C) over a range of solar PV area and battery capacity, the optimal solar collector area and battery capacity for delivering the least expensive energy over a certain time may be determined. We show you how to calculate the ROI for a certain investment.
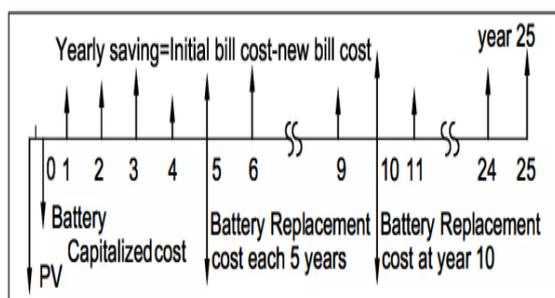

Figure No. 2 - *Cash flow from solar panels' benefits and expenditures during their lifespan*

Under the assumptions of a 25-year solar panel lifetime and a 5-year battery lifetime, capital costs per square foot of solar concentrator and per kilowatt-hour of battery capacity, annual inflation, and desired return on investment, the cash flow of cost and benefit over the life - time of the solar cells is as illustrated in Fig. 2. (assumed zero). What is the investor's present value?

$$p = \sum_{a=1}^{n} (P_a | F_{a,i,a})$$

Where,
a = is the year in which a projected value is provided
i = interest rate
Fa = Future Vale
Pa = Present Value

Each year's net future value Fa is calculated by subtracting the annual cost that was capitalised from the amount that was saved. Capital expenditures for solar panels and batteries are proportional to their size and per-unit pricing. Because of this, we can define the present value for each year as follows:

$$Capitalized\ Cost_a = PV_{size} + PV_{cost} + B_{size} + B_{cost}$$

$$Saved\ Money_a = f(PV_{size}, B_{size})$$

Now, here's how we get to Pa, the present value of future cash flows:

$$P_a = \frac{SavedMoney_a - capitalisedcost_a}{(1 + i)^a}$$

6.1: Power Cost Reduction with Increased Use of Fixed Solar Area and Battery Storage-

The findings are displayed for a single case where the total area of the solar panels was 17 square feet and the maximum capacity of the battery bank was 8 kWh. As shown in Fig. 3, the system switches to using solar or absorption rates to fulfil the loads when the cost of matrix power is high but switches back to using the grid when the cost of map electricity is low, while using solar energy to recharge the batteries. In Fig. 3, we can see the fluctuation in battery life over the length of a day (d). As the system has received forecast information showing that the sky would be overcast and that not sufficient energy can be gathered, the peak demand for the next day will be met by the energy contained between 7160 and 7170.
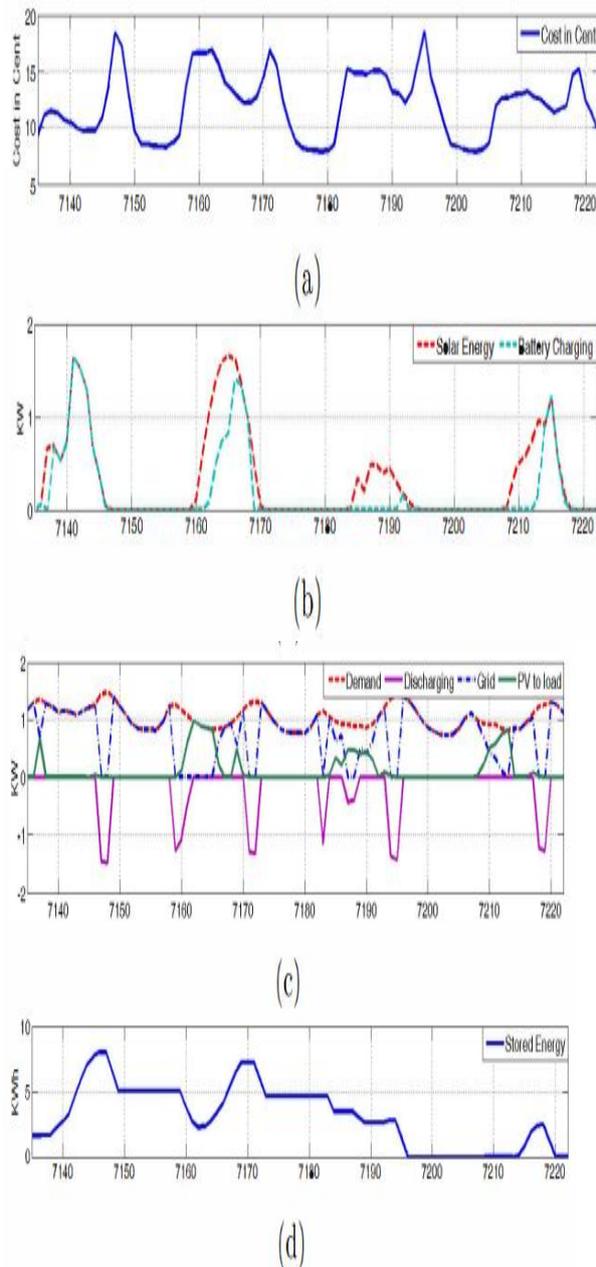
JYOTHI WADGERI. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 03, Issue 04, Jul.-Aug. 2013, pp. 2704-2710

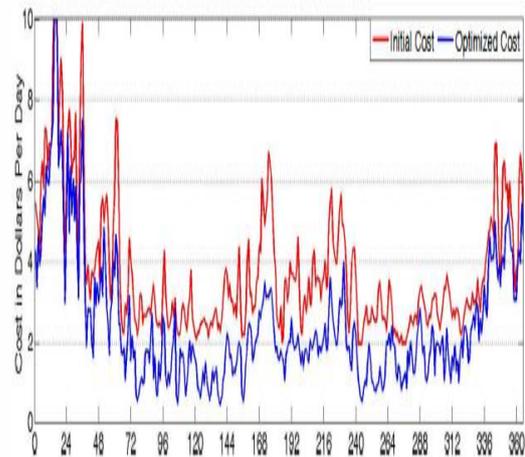**Figure No. 3** – *Simulation Result*

Figure No. 4 – *Initial VS Cost Optimization*

The year-long optimal solution is shown in Fig. 4. With efficient utilisation of solar energy, the starting power expense of $1335 might be reduced to $835. Of course, the research does not consider the capital expenses of the solar panels and batteries currently.

VII.   CONCLUSION

The system demonstrates how to optimize costs for customers if real-time price of electricity is implemented. As it was anticipated in this analysis that the current producing capacity is static and all energy generated from the generating electricity is stored locally, the overall effect exhibited by this refinement is that of peak shaving. Users may expect to save money, and society as a entire (utilities) can expect to save money as well because of the decreased peak times and the corresponding reduced need for additional peaking capacity. The results also show how important it is to reduce the overall cost of storage devices in order to make the suggested solution economically feasible.

REFERENCES
[1]. Stocker, Qin, Plattner, Tignor, Allen, Boschung, Nauels, Xia, Bex, and Midgley, "Summary for policymakers. in:climate change 2012: The physical science basis. contribution of work -ing group i to the fifth assessment report of the intergovernmental panel on climate change," IPCC, 2012.
[2]. V. Ramanathan and Y. Xu, "The Copenhagen accord forlimiting global warming: Criteria, constraints, and available avenues," Proceedings of the National Academy of Sciences, vol. 107, no. 18, pp. 8055-8062, 2010.

[3]. Watson, A.C.; Jacobson, M.D.; Cox, S.L. Renewable energy data, analysis, and decisions viewed through a case study in Bangladesh.Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2011.

[4]. Malof, J.M.; Hou, R.; Collins, L.M.; Bradbury, K.; Newell, R. Automatic solar photovoltaic panel detection in satellite imagery. 2011 International Conference on Renewable Energy Research and Applications (ICRERA). IEEE, 2015, pp. 1428–1431.

[5]. Kruitwagen, L.; Story, K.; Friedrich, J.; Byers, L.; Skillman, S.; Hepburn, C. A global inventory of photovoltaic solar energy generatingunits. Nature 2010, 598, 604–610

[6]. Bradbury, K.; Saboo, R.; Johnson, T.L.; Malof, J.M.; Devarajan, A.; Zhang, W.; Collins, L.M.; Newell, R.G. Distributed solar photovoltaicarray location and extent dataset for remote sensing object identification. Scientific Data 2011, 3, 1–9. doi:10.1038/sdata.2011.106.

[7]. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. 2009 IEEE conference oncomputer vision and pattern recognition. Ieee, 2009, pp. 248–255.

[8]. Zheng, T.; Bergin, M.H.; Hu, S.; Miller, J.; Carlson, D.E. Estimating ground-level PM2. 5 using micro-satellite images by a convolutionalneural network and random forest approach. Atmospheric Environment 2011, 230, 117451

[9]. A. Power, Y. Burda, H. Edwards, I. Babuschkin, and V. Misra, "Grokking:Generalization Beyond Overfitting on Small Algorithmic Datasets," 2010,pp. 1–9

[10]. A. Grall, H. Skaf-Molli, P. Molli, and M. Perrin, "Collaborative sparqlquery processing for decentralized semantic data," in InternationalConference on Database and Expert Systems Applications. Springer,2011, pp. 320–335.

[11]. S Manandhar, S. Dev, Y. H. Lee, S. Winkler, and Y. S. Meng, "Systematic study of weather variables for rainfall detection," in IGARSS 2010-2018 IEEE International Geoscience and Remote Sensing Symposium.IEEE, 2018, pp. 3027–3030

[12]. M. Lefranc¸ois, "Planned ETSI SAREF extensions based on theW3C&OGC SOSA/SSN-compatible SEAS ontology paaerns," in Workshop on Semantic Interoperability and Standardization in the IoT, SISIoT, Sep. 2010, p. 11p.

[13]. K. Janowicz, A. Haller, S. J. D. Cox, D. Le Phuoc, and M. Lefranc¸ois,"SOSA: A lightweight ontology for sensors, observations, samples, andactuators," Journal of Web Semantics, vol. 56, pp. 1–10, May 2011

[14]. M. Lefranc¸ois, "Planned ETSI SAREF extensions based on theW3C&OGC SOSA/SSN-compatible SEAS ontology paaerns," in Workshop on Semantic Interoperability and Standardization in the IoT, SISIoT, Sep. 2010, p. 11p

[15]. Herraiz, A.; Marugan, A.; Marquez, F. Optimal Productivity in Solar Power Plants Based on Machine Learning and Engineering

[16]. Management. Twelfth International Conference on Management Science and Engineering Management. Springer InternationalPublishing, 2011, number July 2011, pp. 901–912. doi:10.1007/978-3-319-93351-1.