Weather-based solar energy prediction

4 authors, including:

Marcin Detyniecki
AXA GROUP
212 PUBLICATIONS **2,095** CITATIONS

Ashwati Krishnan
Carnegie Mellon University
18 PUBLICATIONS **201** CITATIONS

Mel Siegel
Carnegie Mellon University
164 PUBLICATIONS **3,449** CITATIONS

Weather-Based Solar Energy Prediction

Marcin Detyniecki

Université Pierre et
Marie Curie
LIP6 – UPMC – CNRS
4 place Jussieu, 75005
Paris, France
Marcin.Detyniecki@lip6.fr

Christophe Marsala

Université Pierre et
Marie Curie
LIP6 – UPMC
4 place Jussieu, 75005
Paris, France
Christophe.Marsala@lip6.fr

Ashwati Krishnan

Dept of Elect. & Comp.
Engineering (ECE)
Carnegie Mellon
University
Pittsburgh PA USA
ashwatik@andrew.cmu.edu

Mel Siegel

The Robotics Institute
Carnegie Mellon
University
Pittsburgh PA USA
mws@cmu.edu

Abstract— **Photovoltaic solar panels are effective energy sources during periods of bright sunlight. Excess energy can be stored for later use at night or on cloudy days. The decision to use the stored energy now or later depends largely on being able to predict the weather on different timescales. Short term prediction of stored energy is challenging due to the non-trivial I-V characteristic of the solar cell. The erratic nature of the weather makes long term predictive energy management difficult. In this paper, we address these issues based on data collected from a solar panel, as well as its relationship to observations made of the weather. We observe that prediction, based on fuzzy decision trees, reduces the energy error by 22% compared to a constant prediction equal to the average on the studied period. Thus, exploiting the *fuzzy* classification provided by a fuzzy decision tree is a good improvement compared to the baseline.**

Keywords: solar energy; photovoltaic; power utilization planning; weather; energy prediction; fuzzy decision trees.

I. INTRODUCTION

The availability of solar energy is not guaranteed at any particular place or time: it depends, of course, on time-of-day, but also on the weather conditions that prevail and that prevailed recently. Since meteorological agencies provide detailed weather forecasts round-the-clock, we should be able to use their predictions to our advantage in planning activities that require solar energy. Three interesting questions are apparent: (1) given the standard *weather forecasts* available today, can we reliably predict the energy we will be able to capture tomorrow? (2) given our own *measured actual insolation and other local weather conditions right now*, to what extent can we make that prediction? and (3) what is the optimal approach to fusion of these two prediction sources? The answers to these questions are not only of academic interest but also of crucial practical importance [6]. Contemporary solar panels are series-arrays of silicon photovoltaic cells that are essentially large-area silicon *pn*-junctions. Incident optical photons promote electrons from the valence to the conduction band. The band-gap voltage across the junction capacitance thus has the potential to drive DC current through an external load. The cell's open-circuit voltage is essentially the band-gap. Its short-circuit current depends - not necessarily simply - on the incident optical power. Their ratio is the internal impedance to which an external load must be matched to achieve maximum energy transfer to the load.

Optimally extracting short-term power and long-term energy is thus a complicated business that requires active real-time control intelligently based on knowledge of present requirements and an ability to predict and plan for future requirements [7], [8].

Presenting how a fuzzy prediction method, and in particular the Fuzzy Decision Trees (FDTs) can improve energy prediction accuracy, prediction is this paper's main goal. We have chosen, for this early attempt in estimating the energy gain on real conditions, to use FDTs, in contrast to other approaches such as neuronal networks or other regression techniques, because FDTs produce human understandable rules that will allow us, in the future, to improve the system. In fact, not only relevant variables are automatically indentified, but also their interaction is identified. Moreover FDTs have the advantage to be able to handle simultaneously symbolic (here weather classes such as cloudy, sunny, thunderstorm) and numerical ones (such as temperature).

In order to achieve real conditions we used a standard solar panel for home use, described in Section II. We placed the panel in real conditions and collected I-V data with a dedicated electronic apparatus and weather conditions and forecast from the national service over the Internet, as presented in Section III. In the following section we briefly present the Fuzzy Decision Trees and how training and testing was performed. Section V and VI are dedicated, respectively, to data and results analysis.

II. SOLAR PANELS

Solar cells are connected in series to build solar *modules* or *panels*. Panels generally consist of 28 to 36 cells in series to produce 12VDC under defined illumination conditions. An ideal solar panel current-voltage (I-V) curve is shown in Figure 1.1. For any real panel there is a continuous family of these curves wherein open-circuit voltage increases with illumination level and current-droop increases with decreasing illumination. Thus optimum transfer of solar power to an external load requires *matching the load impedance to the illumination level*. Figure 1.2 shows a family of I-V curves for our solar panel collected during 5-hour period when insolation was changing. Notice especially the variations in curve scale and shape, and, based on the teaching of Figure 1.1, the consequent variation of available power and optimum load to extract it.
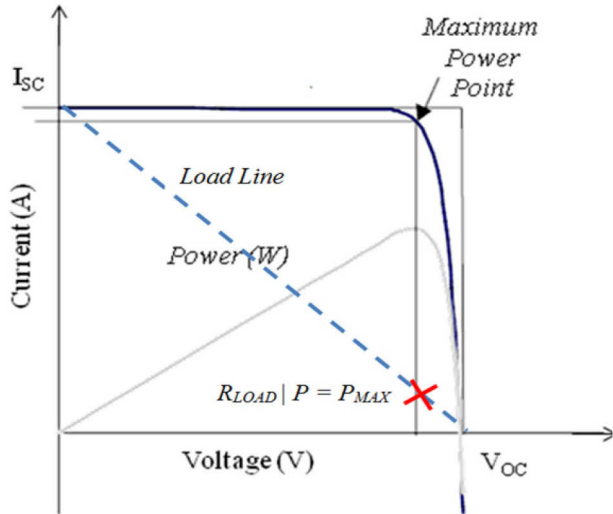
Figure 1.1: Ideal solar power panel. I_{sc} is the short circuit current (when load resistance R_L=0) and V_{OC} is the open circuit voltage (when $R_L = \infty$). The black curve is the I-V characteristic, the gray curve is the power available to an external load (the IV product), and the dashed-blue load line finds the operating point (V,I) - marked by the red cross - at which the load will extract maximum possible power from the panel.

A. Getting the maximum power out of a panel (MPPT)

Consider a system where the load is connected directly across the solar panel. Its maximum power point (MPP) is the point on the I-V curve where the area under the curve is maximum, as shown in Figure 1.1. For optimal simplicity and efficiency one should choose a solar panel that perfectly matches the intended load. But this is not possible: the I-V curve - hence the MPP - changes with illumination. It also changes with panel temperature, which also depends in part on illumination. Active measuring and switching power converters, called maximum power point trackers (MPPT), can switch the load so as to keep the operating point at the MPP. Several solutions, in particular based on fuzzy control, have been proposed [11] and are still under investigation [9]. A complete comparison can be found in [10].

On the one hand this is simple, on the other hand it is daunting. If all we want to do is, say, toast bread, then it is easy enough to switch the resistance of the heating element; the toasting time changes with illumination level, but within reasonable limits we still make toast. But for the vast majority of practical loads - "appliances" - it is impossible to flexibly and efficiently trade off voltage rating and current demand. We thus anticipate a critical near-future demand for active power converters that will accommodate a plausible range of fluctuating DC input voltages and deliver stable standard DC or AC output voltages without incurring unacceptable losses [12]. Note that the control algorithm required for MPPT is non-trivial. The MPP is not known *a priori*, and it moves with variations in illumination and temperature. In practice perturb-and-observe (P&O) algorithms are employed [10], despite the objection-in-principle that when the system is actually optimized any perturbation is guaranteed to reduce efficiency.

Clearly the scale of the integral term in the control algorithm is crucial, and should itself be dynamic, as the system needs on the one hand to respond rapidly to fast changes in illumination level, e.g., passing clouds, and on the other hand it must not spend too much efficiency hunting when conditions are changing only slowly, e.g., on cloudless days.
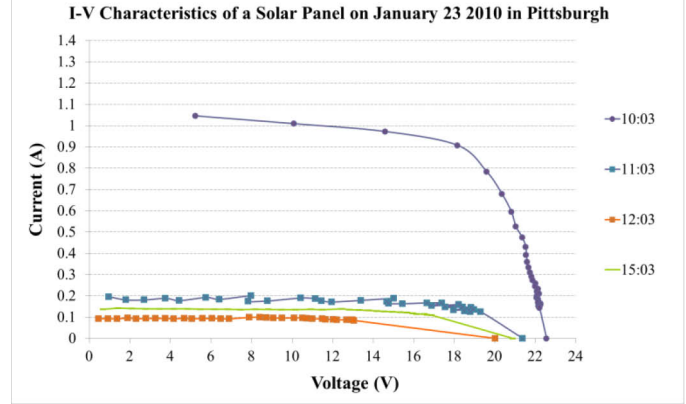

Figure 1.2: I-V of our solar panel at four times on a typical day. *Load resistors between 5 and 155 ohms in 5 ohm steps are switched in random order across the panel, voltage across the load is measured and current is calculated from voltage and resistance. Open-circuit voltage is also measured and recorded as the zero-current value. A diagrammatic representation of the setup is shown and explained in Figure 3.1.*

III.   SOLAR ENERGY UNDER REAL WEATHER CONDITIONS

The National Renewable Energy Laboratory recommends that solar panels be characterized under standard test conditions (STC): temperature 25 C and illumination 1000 W/m^2 (1.0 sun) with an air mass 1.5 (AM1.5) filtered solar spectrum. The idea is to match the illumination and spectrum of sunlight incident on a clear day on a sun-facing 37°-tilted surface with the sun at an angle of 41.81° above the horizon. This condition - with the panel aimed directly at the sun - geometrically approximates solar noon near the spring and autumn equinoxes in the continental United States. However insolation at the earth's surface is rarely as large as the prescribed 1000 W/m^2. And, as already noted, to realistically study electrical energy generation under realistic weather conditions, realistic fluctuations in lighting and temperature must be observed. Note also that a panel that is optimal in the NREL environment is almost certainly suboptimal in any natural environment. So to study solar energy production with practical goals under natural weather conditions it is advisable to combine the solar panel with an MPPT. But there are many such commercial devices, each one running some undisclosed proprietary algorithm, none of them arguably best or even in any sense standard. Thus we elect to organize our measurements in a way that allows us to simulate an ideal MPPT algorithm – that is collect all possible data first and compute after the fact the real optimal point.
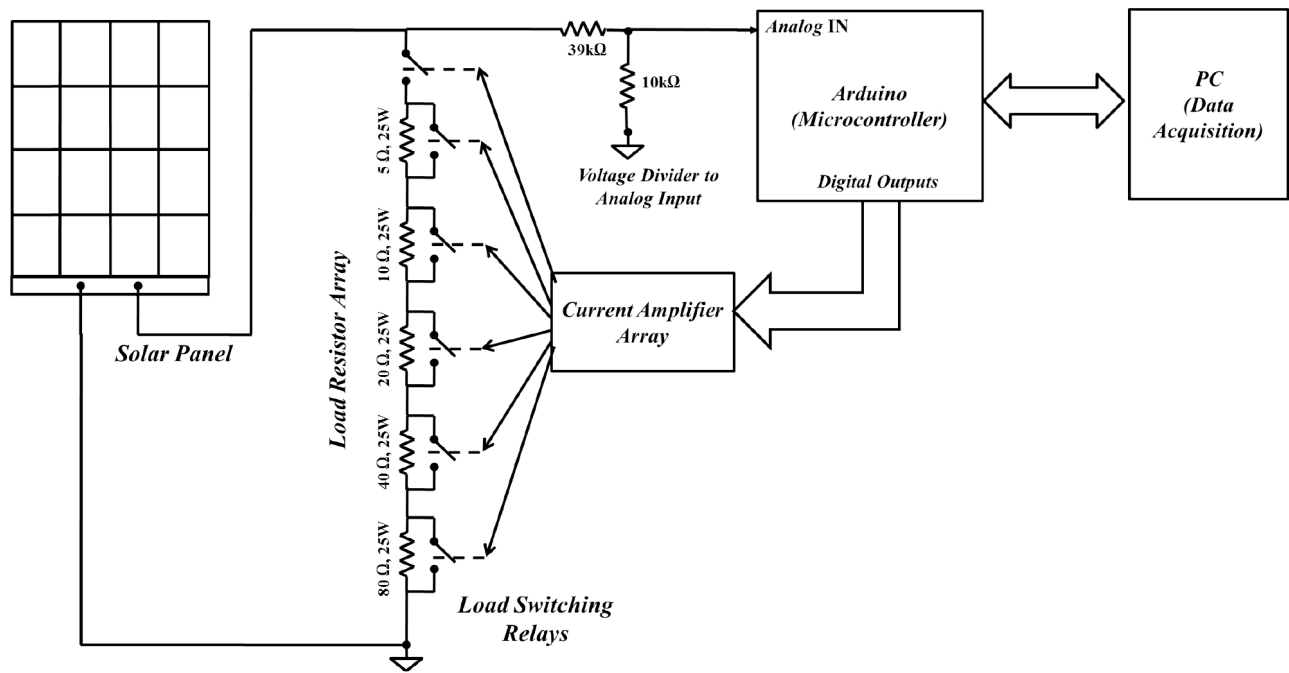
Figure 3.1: Hardware diagram. *From left to right, solar panel, serial load resistor array in series with disconnect-relay (to measure panel's open-circuit voltage), transistor-buffered resistor-shorting relays, voltage divider, Arduino (performing analog voltage input measurement and controlling load-re istor-shorting relays)*

.

A. Instrumentation

We studied the response of our solar panel - approximately 32 cm x 60 cm, so approximately 0.19 m^2 - using a simple single-board data acquisition system in communication with a dedicated laptop computer that is in turn in communication with the internet. Our panel is an off-the-shelf unit mounted at a tilt-angle of approximately 40° outside an approximately south-facing window with a reasonably clear view of the sun's path most of the day, most of the year. The panel's pointing and tilting are probably never perfectly optimal, but are a good compromise that receives better-than-average solar radiation throughout the year. Data acquisition and control are provided by an Arduino Duemilanouve (2009) [3], a low-cost easy-to-program open-design board that provides convenient access to the ATMega168 microcontroller's digital I/O, 10-bit analog input, PWM output, and serial communication pins. A program written in a C-like language using a simple API on a PC is more-or-less invisibly compiled and downloaded via a USB channel on which data are subsequently also returned. Digital output pins are transistor-buffered and diode-protected to safely switch the coils of relays that short-out a series-array of {5, 10, 20, 40, 80} ohm power resistors to provide 5 to 155 ohm load in 5 ohm steps - plus open-circuit - across the solar panel. A measurement sequence is initiated and recorded every 10 minutes. Independently but also every 10 minutes, a USB webcam captures a sky picture. The data files and sky pictures are stored "in the cloud" using Dropbox [2]. As a practical matter, the ATMega168's ADC's rudimentary analog input circuitry and 10-bit resolution do not provide precise or accurate measurements. But they do appear to be stable, which is all that is really required for the present experiments, wherein we are interested primarily in reaching qualitative

conclusions. Of course, since the measurements do seem to be stable, after-the-fact calibration can be undertaken if subsequently it seems valuable.

B. Weather Forecast

The solar panel and its Arduino-plus-Windows-laptop based monitoring system are located at an off-campus location, which is secure and has a good south-looking view with a large open-sky solid-angle. On campus another Windows PC that has reliable access to the Internet periodically downloads present and predicted weather information from the National Oceanic and Atmospheric Administration (NOAA) through the Yahoo! Weather RSS Feed [1], in the form of XML files. Since, weather conditions tend to vary slowly, we recorded the weather conditions every hour, every day. In order to be able to match forecast with current condition, we used the 48 standard categories provided by the weather service. To minimize the prediction error, we choose here to use the forecast just before sunrise. Other more complicated methods could take into account the evolution or tendency of the forecast.

C. Data aggregation

The question of how to aggregate the data may seem simple at first sight, but it is in fact extremely complex. We choose to work on a one-full-day basis, because it provides a natural, regular cycle. Further works could deal with energy prediction with a shorter or a longer time horizon. Hence to compute the energy produced by the panel over one day we need to start from the power measurement obtained every 10 minutes. Our first step consists of choosing from each of these series the maximum power. In this way we simulate an ideal MPPT. Then under the assumption that everything remains equal for the following ten minutes we integrate over the whole day to obtain the total energy produced. The assumption introduces an

error for quickly changing conditions (as for instance a sunny day with some clouds). In fact, the measurement could have been done when the cloud is just over the panel. We believe that the introduced error averages out because of the frequency and the uniform nature of the sampling. In fact, if there are a lot of clouds, more often than not the measurement will be done under reduced illumination approximately proportional to the average coverage. Since weather conditions fluctuate during the day, to obtain a global "for the day" weather classification, we choose to aggregate by majority vote all the classifications of the National Weather Service reported during the daylight hours of that day. In other words, we choose to label the day based on the most frequent NWS classification; and we focus our attention only on the hours when there should be light (between sunrise and sunset). So, if it rains for only one hour during the day and it was, for the rest, a sunny day, it is labeled as a sunny day (notice that this is not the case for weather services). Although the solar panel data, the sky pictures, and the downloaded weather data are not perfectly synchronized, for the purpose and nature of the experiments described their imprecise - and occasionally inconsistent - alignment is inconsequential.

IV. ENERGY AND WEATHER PREDICTION

Based on the data described above the challenge is to predict, before the sun rises, the energy that we will by produced during that day. All methods can be grouped in two large families: The direct ones, where the energy value is computed by a "black-box" algorithm (usually regression like, as for instance Neuronal Networks [7] [8]) and the indirect ones, where first a weather class is predicted and based on it an average electricity is predicted.

In this paper, we choose to explore the performance of the latter. This approach allows using the power of the national weather forecast services, without any further modifications, to predict the energy. The general formulation has the advantage of opening the range of possible algorithms that can be used. In particular, we choose here to use Fuzzy Decision Trees, which are not only able to deal with symbolic and numerical classes simultaneously, but also provide an explanation to the prediction.

A. Fuzzy Decision Trees

Fuzzy decision trees (FDTs) are an extension of classical decision trees. They have been introduced in Machine learning to handle training sets that contains numerical and/or fuzzy values [13] [14] [15]. Moreover, such trees introduced a soft classification of examples that leads to a smoother decision. Thus, degrees of decision and degrees of membership to classes are provided as a result of a classification by means of the FDTs.

The construction of a FDT from a training set $T = \{e_1,...,e_n\}$ is based on the well-known ID3 [16] or the CART algorithms [17]. A fuzzy decision tree is made up from its root to its leaves by sequentially partitioning T into subsets. Each partition is obtained from a comparison on the values of a selected attribute. This comparison made up a node of the tree.

Let each example e_i from T described by means of a set of values for attributes $A = \{A_1, ... , A_m\}$. Where each attribute A_j can take a fuzzy, numerical, or symbolic value v_{jl} in the set $\{v_{j1},..., v_{jm}\}$. An example's *description* is a m-tuple of attribute value pairs (A_j, v_{jl}). Each description is associated to a class c_k from $C = \{c_1,..., c_K\}$ to make up the *training* example e_i. A fuzzy value v_{jl} is associated with a membership function μ_{vjl} from T that associated to each e_i of T the degree of having the value v_{jl}. Similarly, each c_k is supposed to be associated with a membership function μ_{ck}.

At each step of the construction of the FDT, an attribute is selected by means of a measure of discrimination, for instance, the well-known Shannon entropy from Information theory [16], [17], that orders the attributes according to their increasing correlation to the C in the local training subset. The discrimination power of each attribute is valued with regard to the classes [18]. The attribute with the highest discriminating power is selected to construct a node. Well-known fuzzy measures of discrimination are the fuzzy entropy (that is an extension of the Shannon entropy to fuzzy events) [15], and the measure of ambiguity [13]. A new measure, the gradual discrimination measure, has been introduced in [19]. This measure is interesting in our case because it values the discrimination power of the values of an attribute with regards of the values of the class and takes into account a monotonic relation between these values if there exists (see [19] for a full explanation on that measure).

The aim of a FDT is to classify any forthcoming example, not necessarily present in T. To classify an example e, paths in the FDT are followed from the root to leaves of the tree, according to the values of the attributes of the description of e. At each node of a path, a membership degree for e is valued depending on the value of e for the attribute presents in the node and the fuzzy values that label vertices going out that node. On a path, all the membership degrees valued from the root to the leaf are aggregated thanks to a conjunctive operator (typically, a t-norm). The membership degrees for e obtained for the whole leaves of the FDT are aggregated thanks to a disjunctive operator (typically, a t-conorm). That leads to value a membership degree for e to belong to each class c according to the FDT. Various pairs of t-norms and t-conorms can be used to aggregate the membership degrees. The most classical ones are the Zadeh's operators (minimum, maximum), or the Lukasiewicz operators. More details can be found in [15]. A FDT can also be used as a crisp decision tree: the alpha-cuts of level 0.5 of each fuzzy membership functions are used to replace the fuzzy sets. Such crisp use of a FDT enables the tree to produce a single class, non fuzzy, as result of classification of an example.

B. Baseline prediction

In order to measure the improvement obtained by our method, we need to define a distance measure and a baseline. To assess the extent to which we can predict the energy production of a solar panel, we calculate the average of the absolute values of the differences between the predicted energy and the observed energy for the proposed models.

To enrich the analysis we propose three baselines:

- *Constant average prediction*: we assume that the average energy for a region and for a period of time can be perfectly predicted, but is constant for all period. To achieve this we compute, after the fact, the average energy observed during the whole period. Notice that this is an ideal point that cannot be achieved, in real predictions conditions. Any constant prediction will augment the proposed energy distance.

- *Energy tomorrow equals the one produced of today*: this is a standard method used for time series and in particular in weather forecast prediction.

- *Pure weather forecast based prediction*: we propose to use the weather forecast as the predicted energy class. This approach corresponds to the natural way we would address the problem: "If today is going to be sunny and on a sunny day we produce on average energy E then today we should observe energy E."

TABLE I.   ENERGY BASED ON OBSERVED CURRENT CONDITIONS

Majority Weather	April - July 2010		
	Nr of Days	*Watt-hr*	*Std Deviation*
Fair (day)	38	449.6	119.4
Partly cloudy (day)	12	396.9	136.3
Mostly cloudy (day)	8	261.6	114.8
Cloudy	14	149.1	96.2
Showers	5	70.2	122.6
Globally	**77**	**342.6**	

V. Data analysis

Between end of April and beginning of July 2010, we collected data for 77 successive days. The average energy produced per day was 342 watt-hour with standard deviation 178 W-h. Table I shows that roughly half the days are "fair" and half are "cloudy" or "rainy". As expected, "fair" days tend to produce more energy than "partly cloudy", which are better than "mostly cloudy", "cloudy", and "shower" days in that order. This conformity of semantic and energetic descriptions gives us confidence that our model, and in particular the majority aggregation process, are suitable. The variability of the daily energy production is rather large, but more or less constant for each category.

The accuracy of the weather prediction for the studied period, using the standard set of categories, was of 60% (of correct prediction at sunrise for the day). This surprisingly small proportion can be explained by two phenomena: aversion to risk in the prediction and mismatch of categories. In Table II, which shows the number of forecast weather conditions, we can observe a shift towards an increased number of rainy days (predictions). We observed 5 "shower" days, but 37 "showers" or "thunderstorms" predictions. This discrepancy may come from the aversion to risk of the weather forecaster. In fact, if it should rain for only an hour in day the weather forecast will be "rainy day". But our majority observation, suitable for the energy prediction, would be sunny day, with consequent category mismatch. Moreover, by comparing labels on Table I

and 2, we notice that the number and labeling of categories differs in the two sets, thus more-or-less guaranteeing mismatches. Labels appearing in the forecast do not appear in the current weather observations. For instance there are no "cloudy" predictions and no "sunny" forecasts. This reveals an even more profound and structural problem: class boundaries are fuzzy. In fact, if we predict "mostly cloudy" and we observe "partly cloudy" it will be considered a mismatch. New weather classes could be created by grouping labels, as for instance "cloudy" with "partly cloudy" in an "overcast" class; but preliminary work showed that the prediction accuracy does not improve, because the descriptions then become too vague or arbitrary.

TABLE II.   ENERGY BASED ON FORECASTED CONDITIONS

Forecast at Sunrise	April - July 2010		
	Nr of Days	*Watt-hr*	*Std Deviation*
Sunny	12	533.3	54.5
Fair (day)	10	489.2	82.2
Partly cloudy (day)	16	381.5	151.5
Mostly cloudy (day)	2	239	103
Showers	8	98.8	109.6
Isolated thunderstorms	7	380.3	79.7
Scattered thunderstorms	19	242.8	135.4
Thunderstorms	3	146.7	134
Globally	**77**	**342.6**	**94.4**

Improved weather forecast based prediction: To increase the prediction quality due to what is described above, the total mismatches (no sunny day observation) were manually matched to the closest class: sunny to fair, any thunderstorms type to showers, etc.

VI. Results

Table III shows the energy prediction difference. By assuming that a solar panel produces more-or-less the same (constant prediction, baseline) we observe an average discrepancy of 152 W-hr compared with what is really observed. If we use the naïve model that assumes that tomorrow energy is equal to what was observed today, we observe that difference predicted-observed is increased. This proves that the energy tends to change rather quickly and that a constant assumption is a good baseline not easy to beat.

If we focus our attention to the improved (with manual match of fuzzy classes) method based only on the weather forecast, we observe a reduction of 12% with respect to the constant average estimation.

We used the fuzzy decision trees to predict the energy. In this approach, we use the Salammbô software [15] to build a FDT from the whole dataset. From a training set, the Salammbô software provides us with a FDT with fuzzy set

values that label vertices going from a node associated with a numerical attribute.

Numerical attributes are automatically discretized (as a fuzzy partition) by means of the software, at each step of selection of an attribute to build a node of the tree. Attributes to build nodes of the FDT are selected by means of a discrimination measure [18]. In this experiment, we use the gradual discrimination measure introduced in [19]. The predicted energy class has been discretized in 4 intervals, from 0 (0 to 180 W-hr) to 3 (greater than 500). The classification of an example by means of the FDT provided a set of membership degrees to each intervals that define the class. In order to obtain the predicted energy of the example, median values of each interval weighted by the corresponding membership degrees are aggregated to provide a predicted energy.

The FDT constructed from the whole training set (77 examples) is composed of 38 paths, with a maximum of 7 nodes on a path, and an average number of 5.1 nodes on a path. Some instances of paths are:

- If the majority weather at sunset is mostly cloudy, and if the temperature max is *lower[1] than 20* then the predicted energy ranges from 370 to 500 (class 2).

- If the majority weather at sunset is cloudy or showers, and if the temperature min is *greater[2] than 9* and the weather at sunrise is fair then the predicted energy ranges from 180 to 370 (class 1).

We recall that a path in a FDT is equivalent to a fuzzy rule: premise of the rule is composed of the attribute values that pertains to the path, and the conclusion of the rule is the value of the class presents in the leaf of the path.

We investigate the validity of this approach by means of a *leave one out* experiment with the whole collected data set. Results are presented in Table III.

With a crisp use and a crisp output of the FDT, the FDT products a single weather class as output. In that case, we can observed (column "Crisp") that the prediction is worse than the baseline one.

The accuracy of energy prediction can be further improved by taking into account the fuzzy classification provided by the FDT. The use of FDT with either min-max t-norms or Lukasiewicz tnorms to aggregate the membership to the vertices on paths from the root to the leaves (see [15]) provides an important improvement of the prediction. The min-max weighting scheme provides excellent results reaching a 33% improvement compared to the baseline, with an average energy difference of 106 W-hr. Good results are also obtained by means of the Lukasiewicz weighting scheme that provides a 26% improvement compared to the baseline, with an average energy difference of 112 W-hr.

---

[1] *Lower than 20* is a fuzzy set deduced automatically during the construction of the FDT. It is a piecewise linear membership function with a support equals to $(-\infty, 21]$ and a kernel equals to $(-\infty, 19]$.

[2] *Greater than 9* is a fuzzy set deduced automatically during the construction of the FDT. It is a piecewise linear membership function with a support equals to $[7, +\infty)$ and a kernel equals to $[11, +\infty)$.

TABLE III.   AVERAGE ENERGY DIFFERENCE BETWEEN THE DIFFERENT PREDICTION MODELS, COMPARED TO BASELINE (BEST CONSTANT PREDICTION)

	Prediction Models Comparison					
				Fuzzy Decision Trees		
	Best Constant (baseline)	Today equals Tomorrow	Improved Weather Forecast	Crisp	Min-max	Lukasi-ewicz norms
Average Energy Difference (watt-hr)	152	170	134	223	106	112
Compared to baseline	--	worse	-12%	worse	**-30%**	-26%

VII.   CONCLUSIONS AND FUTURE WORK

The use of the weather forecast service allows improving the energy production prediction. It not only improves compared to *any* fixed prediction (based on average of other studies), but also compared to a naïve sequential approach.

Since the weather forecast is wrong forty percent of the time - based on the predictor's own categories – it is necessary either to manually add coherence by realigning the fuzzy categories or use a machine learning algorithm (as here the fuzzy decision trees) to automatically discover the underlying rules. These rules can be used in a second step to setup efficient controllers, as for instance fuzzy Takagi Sugeno ones. But it is important to point out that without such a study, any controller would perform poorly, due to complex relationship existing between weather class, weather forecast and energy production.

One of future works should focus on comparing the performance of this approach with other regression algorithms, such as neuronal networks – although the problem of the symbolic weather classes remain a challenge. The interest will be, not only to compare the performance with a dedicated blackbox, but also, on addressing the challenge of incorporating knowledge in these types of systems, improving the overall performance. Another potential possibility is to test prediction techniques that include temporal evolution, as for example Markov models. Improved prediction models could take advantage of available data sources not incorporated into this first attempt at analysis, e.g., the recorded images of the sky and the locally measured reported temperature: allowing to correct national versus local measurement bias.

Other future work might focus on more complex but more practical setups, for instance, sun-tracking panels, integration with storage batteries, etc. We believe that sun tracking will not dramatically change the conclusions of this work; though of course it will improve absolute collection efficiency. Storage batteries are obviously advantageous in that they give the system designer control over several time scales that are otherwise only in nature's hands, but with these additional handles comes additional complexity and uncertainty.

REFERENCES

[1] Yahoo! Weather RSS Feed [Online]. Available: http://developer.yahoo.com/weather/ (accessed: 2010, Jan)
[2] Dropbox Documentation [Online]. Available: https://www.dropbox.com/about (accessed: 2010, Jan)
[3] Arduino Duemilanuove Datasheet [Online]. Available: http://www.arduino.cc/en/Main/ArduinoBoardDuemilanove (accessed: 2010, Jan)

[4]   Peder Bacher, Henrik Madsen, Henrik Aalborg Nielson, "Online short-term solar power forecasting", Informatics and Mathematical Modelling, Richard Pedersens Plads, Technical University of Denmark, Denmark, 22 May 2009.

[5]   Lin Phyo Naing Srinivasan, D., "Estimation of solar power generating capacity", IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), 14-17 June 2010, Singapore

[6]   Hong-Tzer Yang, Jian-Tang Liao, Xiang-He Su, "A fuzzy-rule based power restoration approach for a distribution system with renewable energies", FUZZ-IEEE 2011: 2448-2453

[7]   Davide Caputo, Francesco Grimaccia, Marco Mussetta, Riccardo Enrico Zich, "Photovoltaic plants predictive model by means of ANN trained by a hybrid evolutionary algorithm", IJCNN 2010: 1-6

[8]   Francesco Grimaccia, Marco Mussetta, Riccardo Enrico Zich, "Neuro-fuzzy predictive model for PV energy production based on weather forecast", FUZZ-IEEE 2011: 2454-2457

[9]   Irwan Purnama, Yu-Kang Lo, Huang-Jen Chiu, "A fuzzy control maximum power point tracking photovoltaic system", FUZZ-IEEE 2011: 2432-2439

[10] Esram, T., Chapman, P.L., "Comparison of Photovoltaic Array Maximum Power Point Tracking Techniques", IEEE Transactions on Energy Conversion, Vol. 22 (2) pp. 439-449, 2007

[11] Chung-Yuen Won, Duk-Heon Kim, Sei-Chan Kim, Won-Sam Kim and Hack-Sung Kim, "A new maximum power point tracker of photovoltaic arrays using fuzzy controller", 25th Annual IEEE Power Electronics Specialists Conf. (PESC'94), pp. 396-403, Taipei, Taiwan, Jun 1994.

[12] Kyohei Kurohane, Tomonobu Senjyu, Atsushi Yona, Naomitsu Urasaki, Tomonori Goya, Toshihisa Funabashi: A Hybrid Smart AC/DC Power System. IEEE Trans. Smart Grid 1(2): 199-204 (2010)

[13] Yuan, Y. & Shaw, M. Induction of Fuzzy Decision Trees Fuzzy Sets and systems, 1995, 69, 125-139.

[14] Janikow, C. Z. Fuzzy Decision Trees: Issues and Methods IEEE Transactions on Systems, Man and Cybernetics, 1998, 28, 1-14.

[15] Marsala, C. & Bouchon-Meunier, B. An Adaptable System to Construct Fuzzy Decision Trees Proc. of the NAFIPS'99, 1999, 223-227.

[16] Quinlan, J. R. Induction of Decision Trees Machine Learning, 1986, 1, 86-106.

[17] Breiman, L.; Friedman, J.; Olshen, R. & Stone, C. Classification And Regression Trees Chapman and Hall, 1984.

[18] Marsala, C. & Bouchon-Meunier, B. Ranking Attributes to Build Fuzzy Decision Trees: a Comparative Study of Measures IEEE World Congress on Computational Intelligence, 2006, 1777-1783.

[19] Marsala, C. Gradual Fuzzy Decision Trees to Help Medical Diagnosis. IEEE World Congress on Computational Intelligence, 2012, Brisbane, Australia, June 2012 *(to appear)*