SIPS: Solar Irradiance Prediction System

Stefan Achleitner*

Computer Science and Engineering

The Pennsylvania State University, University Park, PA 16802

Email: stefan.achleitner@cse.psu.edu

Ankur Kamthe, Tao Liu and Alberto E. Cerpa

Electrical Engineering and Computer Science

University of California, Merced, CA 95344

Email: {akamthe, tliu5, acerpa}@andes.ucmerced.edu

Abstract—There is high interest in up-scaling capacities of renewable energy sources such as wind and solar. However, variability and uncertainty in power output is a major concern and forecasting is, therefore, a top priority. Advancements in forecasting can potentially limit the impact of fluctuations in solar power generation, specifically in cloudy days when the variability and dynamics are the largest. We propose SIPS, Solar Irradiance Prediction System, a novel sensing infrastructure using wireless sensor networks (WSNs) to enable sensing of solar irradiance for solar power generation forecasting. In this paper, we report the findings of a deployment of a hierarchical WSN system consisting of 19 TelosB nodes equipped with solar irradiance sensors, and 5 MicaZ nodes equipped with GPS boards, deployed in the vicinity of a 1 MW solar array. We evaluate different irradiance sensor types and the performance of different novel prediction methods using SIPS' data and show that the spatial-temporal cross-correlations between sensor node readings and solar array output power exists and can be exploited to improve prediction accuracy. Using this data for short-term solar forecasting for cloudy days with very high dynamics in solar output power generation –the worst case scenario for prediction–, we get an average of 97.24% accuracy in our prediction for short time horizon forecasting and 240% reduction of predicted normalized root mean square error (NRMSE) compared to state-of-the-art methods that do not use SIPS data.

Keywords—*wireless sensor networks, solar energy forecasting, sensor data processing, forecasting algorithms*

I. INTRODUCTION

The increasing utilization of photovoltaics (PV) power systems has created much interest in the technical and economic impacts of the solar resource variability and uncertainty [1]. Much like wind power, the output from a PV plant is non-dispatchable and variable in nature, and has high costs of storage. Balancing the power grid in real time is a difficult task since only limited low-cost storage and spinning reserves are generally available. A sudden drop of the power level can adversely affect the local grid stability and power quality, with possible correlated effects on adjacent power nodes [2]. Accurate forecasting techniques of power generation, especially in the case of solar energy, are essential for a smart and dynamic power grid that can absorb certain decreases of the generated electrical power. To enable real-time dispatch, which is an important component of successful solar power plant operation, accurate short-term energy forecasting is critical.

To cope with dynamics, the PV power output is either curtailed or backed-up by ancillary generators during high variability conditions [1], [3]. To limit the variability impacts, the need for accurate short-term forecasting is high on the agenda of the solar energy industry. With proper forecasting, there are possibilities to reduce variability impacts by helping the management of curtailment and ancillary generators [1]. Other possibilities can also arise with accurate short-term (sub-minute) forecasting such as smoothing with ultra-capacitors during large ramping events [4], [5], or increasing inverter efficiency with Maximum Power Point Tracking (MPPT) [6] using solar irradiance forecasted information. In this work, we focus on two important time horizons for power plant operation, real-time dispatch and, to a lesser degree, trading (0-2 and 2-10 min ahead) [1].

Forecasting PV power output is a complex problem because solar irradiance is extremely sensitive to the presence of clouds. Satellite observations do not have the spatial or temporal frequency to provide cloud details involved in the high frequency fluctuations of the solar irradiance. Although sky imagers are great alternative tools to resolve the cloud situation at the local level, problems still exist due to the glaring caused by the circum-solar scattering of sunlight, severely reducing the quality of small time scale resolutions [7], [8]. A single point sensor, such as the Eppley Precision Spectral Pyranometer (PSP), does not adequately characterize large power plants at small time scales because it is able to measure solar irradiance at only one location and hence, it cannot describe the net power output of a power plant covering a large region. In addition, sky imagers and pyranometers are costly devices, making it prohibitive to install in large numbers. This necessitates a cheap, distributed infrastructure for sensing cloud movement over spatio-temporal scales. Wireless sensor networks (WSNs) have been deployed in a wide-variety of monitoring applications [9], [10], [11] and offer a viable alternative. However, even if using a WSN, it is first required to have a relatively inexpensive sensor that consumes low power and is capable of sensing solar irradiance with sufficient spatial-temporal granularity over long periods of times. This would enable deployment in large numbers for sensing over the necessary spatio-temporal scales. In addition, the software infrastructure within the sensor network should be capable of high data delivery rates, support time synchronization over large areas and ensure that the network lifetime is enough for a long term application of the system.

In this paper, we describe the development of SIPS, a wireless network system to enable high-resolution solar irradiance forecasts. To our knowledge, this is the first work of applying wireless sensor networks for solar irradiance estimation. Our goal is to develop a forecasting infrastructure using a deployment of sensor nodes distributed in the proximity of a 1 MW PV solar power plant. We propose several forecasting models that utilize multiple readings from spatially distributed sensors

---

*Work performed while being an EECS Ph.D. student at UC Merced

225

to compute future predicted values of the plant solar power output at short time granularity. We demonstrate the forecasting approach by comparing to the actual solar plant output power and evaluating the performance of our forecasting models by comparing to state-of-the-art prediction methods without using SIPS data. The contributions of this paper are as follows:

- We tested commercial-off-the-shelf (COTS) light sensors under real outdoor deployment conditions, and evaluated their effectiveness in matching the performance of a NIST calibrated instrument, an Eppley PSP radiometer, over four months.

- We designed, built, deployed and experimented in an outside deployment for several weeks, SIPS, a new hierarchical wireless sensor network system comprised of: (a) Solar Irradiance Motes (SIMs), which integrate the COTS light sensor assembled with a neutral density filter on a new board interfacing with a TelosB sensor node; and (b) Gateway Time-Synchronizer (GTS) motes, which consist of MicaZ sensor nodes that provide time synchronization using GPS data to a group of SIMs.

- We developed novel forecasting models able to predict the output of a 1 MW solar plant within a specific time horizon using our sensor data. We evaluated our forecasting models by comparing them to state-of-the-art prediction schemes, such as time series ANN prediction models, showing that our best prediction model can provide average accuracy of 97.24% (maximum of 99%) and reduce NRMSE prediction error by an average of 240% (maximum of 426%) with respect to prediction models that do not use SIM sensor data.

The rest of the paper is organized as follows: A review of related work is provided in Section II. In Section III, we explain our choice of irradiance sensor, hardware and software infrastructure and we describe the correlation between sensor values and solar plant energy output. In Section IV we introduce our prediction models to perform solar energy forecasting. In Section V, we evaluate the performance of our solar energy prediction models. In Section VI we discuss some training and time horizon issues. Finally, in Section VII, we summarize our findings and conclude.

II. BACKGROUND AND RELATED WORK

Several solar irradiance forecast methodologies have been presented for various time horizons. Numerical weather prediction (NWP) models have been tested [12], [13], [14], [15] for medium-term solar forecasting of 6-hours to a few days ahead. Satellite-based forecasts are competitive with NWP models for less than 1–6 hours ahead [16]. Satellite models have also been applied to 30-minute ahead forecasting [17]. Besides these physics-based models, time-series based solar forecasting methods have also been extensively applied, (e.g. [18], [19], [20], [21], [22], [23], [24], [25]). In some of these time-series applications, direct cloud information is not used [23], [21], [19], [20], [25] while some incorporate cloud information as exogenous inputs [22], [24]. The time-series forecasting methodologies have been applied successfully for various time-horizons including 1-hour [22], [23], [21], and 24-hours [20].

The conventional approach for short-term forecasting problem has been to use sky imagers. One particular type of sky imager is the TSI-880 (TSI), developed by Yankee Environmental
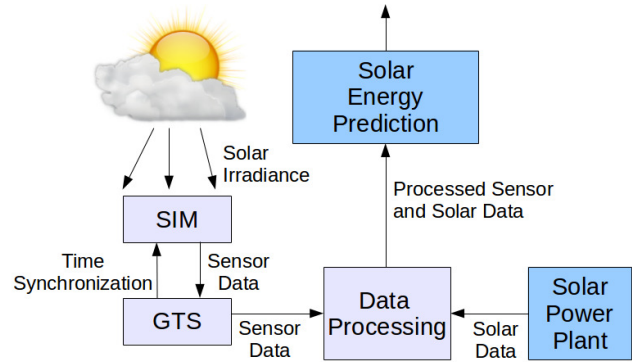

Fig. 2: SIPS Architecture Overview

Systems [26]. The TSI is composed of a hemispherical mirror with a CCD camera located above it. The mirror contains a sun tracking shadow-band that continuously covers the mirror from the direct sunlight in order to protect the camera from the sun's reflection. Recently, Chi et al. [7] used TSI images to produce instantaneous values of Global Horizontal Irradiance (GHI) single-point solar irradiance sensors located in the University of California San Diego meteorological network. They analyzed cloud field propagation by computing matching errors of the future image given the current one translated in the direction of the computed velocity field. Their results show a mean forecasting error of $\sim$8.65% for sub-minute time horizons (30 and 60 seconds) and $\sim$30% for a 5-minute forecast. In our work, we reduced the sub-minute mean forecast error by more than three times, and more than eight times for best case scenario. In [8], cloud indices obtained from the TSI built-in cloud classification algorithms and Infrared Radiometric measurements were reported to improve results for 1-hour forecasting of GHI. In terms of distributed infrastructure approaches relevant to solar irradiance monitoring at short-time scales, [27] deployed 24 LI-COR irradiance sensors around a 1MW PV plant in Hawaii. Data from the sensor was collected using NI WSN-3212 module. The goal of the paper was to study cloud movement, to be able to predict short-term characteristics of the PV plant. The authors report that they were unable to track cloud movement as the cloud patterns appeared chaotic. In contrast, we show that by using sensors at different locations we are able to determine a correlation between cloud movement and solar plant output power. In addition, our proposed choice of sensor is very inexpensive ($<$ \$1 per irradiance sensor compared to the LI-COR irradiance sensor ($\approx$ \$200)).

III. SIPS DESCRIPTION

In this section, we describe the motivation, hardware and software components and deployment of SIPS, which is located in the proximity of a 1 MW solar plant.

A. Motivation

Solar irradiance at the ground level depends on the presence of specific gases in the atmosphere like CO_2, H_2O, etc., aerosols, solar position and cloud cover. The most important factor for solar irradiance at ground level, which directly maps to the output power of a solar plant, is the cloud coverage. Clouds are a random phenomenon and models for cloud movement have been proven elusive [27], [28].
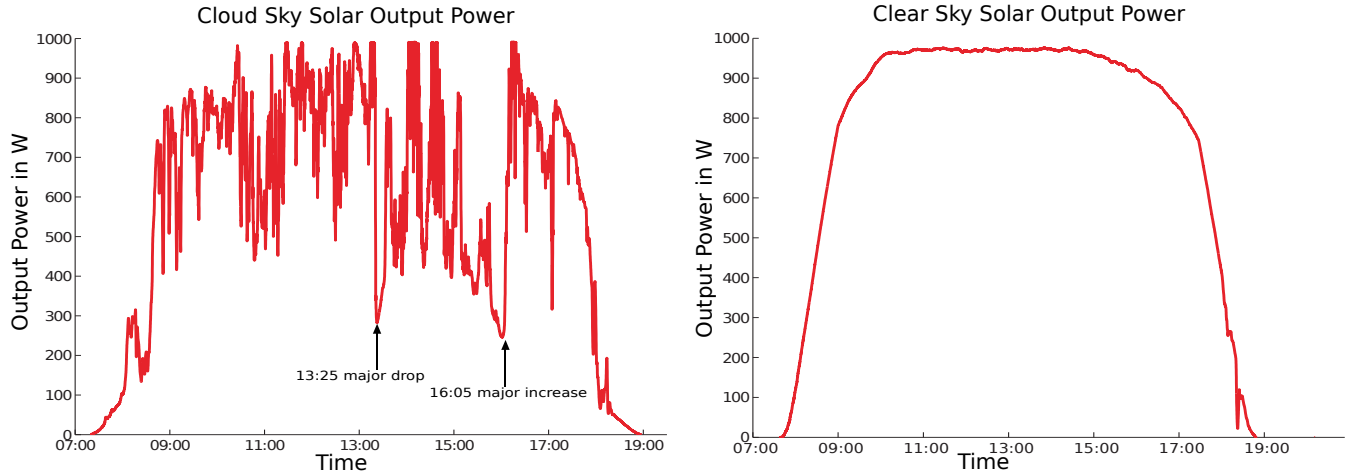
Fig. 1: Solar power output for very cloudy (left) and clear sky conditions (right). On cloudy days, power increases/drops up to 700 kW.
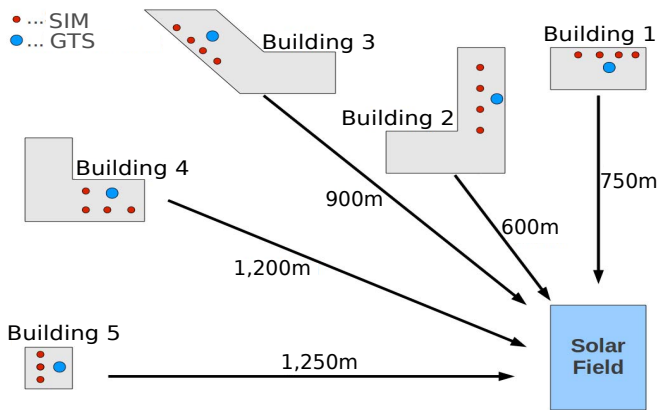

Fig. 3: SIPS deployment overview

Predicting solar irradiance at ground level on a cloudy day is a non-trivial and error prone process. Figure 1 shows the output power of the solar plant on a representative day with cloudy conditions and on a representative day with a clear sky. It can be observed that the output of the solar array on a cloudy day rapidly changes by substantial amounts within a short period of time. Cloud cover alternating with cloud opening cause a significant amount of variability in the power output of a solar plant. In the cloudy day, output fluctuations of up to 700 kW within a few minutes can be observed (e.g. 13:25 and 16:05), which is a *drop/increase of 70% of the overall output power of the plant*. To be able to handle fluctuations like this, a system for short-term prediction of solar irradiance is critical.

B. System Architecture

SIPS comprises of many different components. Figure 2 shows the architecture with our system components. First, a hier-archical wireless sensor network system consisting of: (a) Solar Irradiance Motes (SIMs) measuring the current solar irradiance at the lowest level of the hierarchy, and (b) Gateway Time-Synchronizer (GTS) nodes, that provide overall time synchronization and data delivery to a group of SIMs. Second, a data collection component running on the back-end, which

collects the information provided by all the GTS/SIM nodes as well as the real-time solar plant energy output measurements and stores it in a database for future use. Third, a data processing component that takes the data collected, sanitize it in case of missing sensing/power-plant data and performs on-line model training (if required). Finally, a prediction model to forecast the output of the solar plant for a specific time horizon.

C. System Deployment

Our deployment consists of groups of 1 GTS connected to many SIMs (in our deployment 3-4). All SIMs have single hop connectivity to the GTS. Each group acts independently from each other, since the distance between the different GTS/SIMs group of nodes makes communication using low-power radios infeasible. Figure 3 shows an overview of the SIPS deployment.

To perform short-term solar irradiance prediction it is required that a live sensor data stream is available. Our system of 19 SIMs and 5 GTS nodes is deployed on rooftops in the proximity of a 1 MW solar power plant. Sensed data from the SIMs gets forwarded to the GTS, where the data is forwarded to our server via the building's Wifi (when available) or using 4G Wifi hotspots. SIMs are deployed on the roof-tops of buildings, since it is imperative that the SIMs sensors do not get occluded by shadows produced by infrastructure, such as buildings and other structures. All SIMs deployed on the same rooftop, get a spatial separation of approximately 70-80 meters maximum in each direction from the GTS node (i.e. covering a maximum length of 150 meters for the longest rooftops). It is also possible to deploy our SIM motes on non roof-top locations, but it is important that the sensors are not shadowed by objects blocking the sunlight.

The area of the solar plant is 170 meters in width (West to East) and 230 meters in length (North to South), covering an area of approximately 10 acres. The array is comprised of 488 single-axis, solar trackers each with 10 PV panels. The SIMs are deployed in clusters out of radio range from each other, covering an arch of about 45° north-west around the solar field.
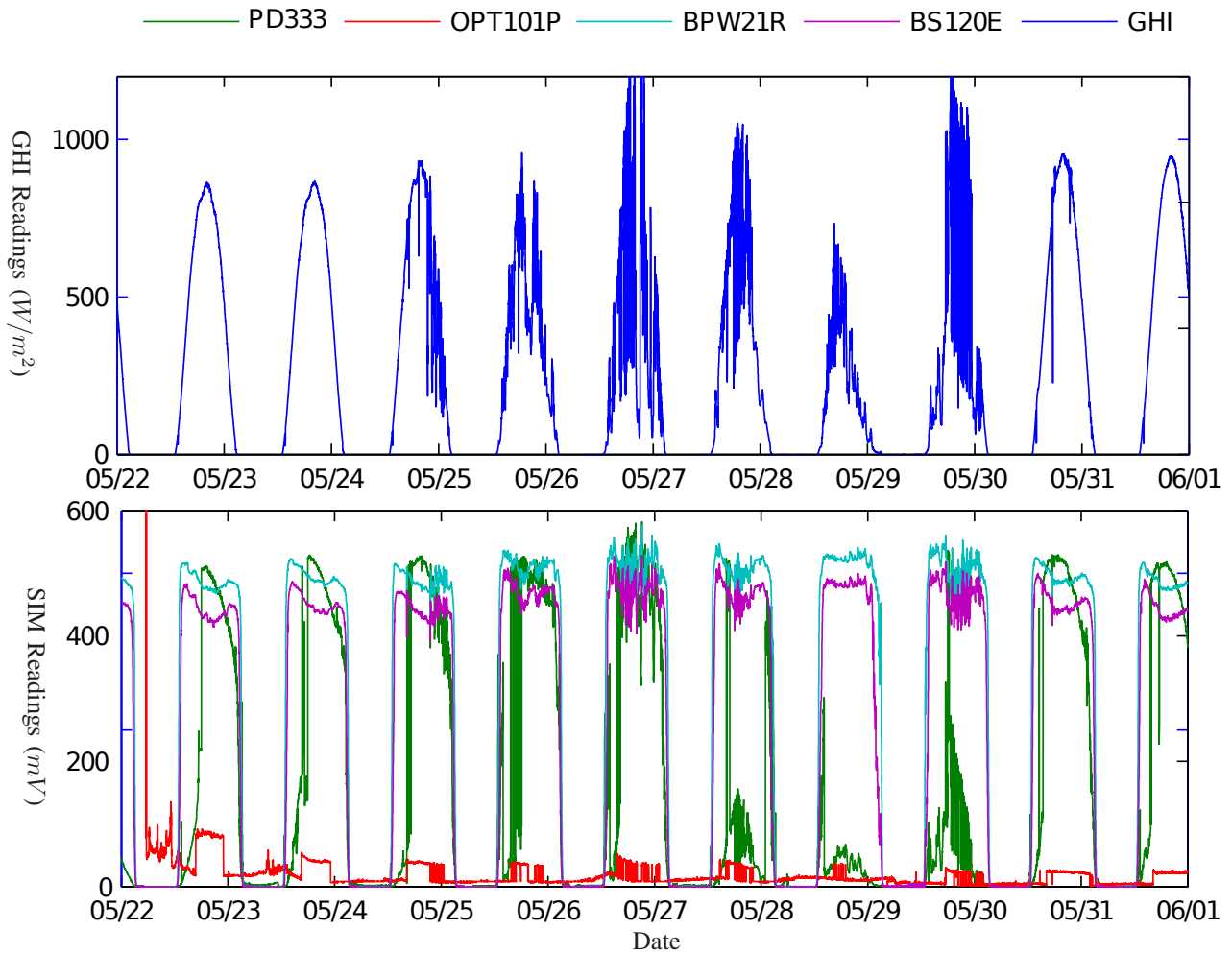
Fig. 4: Global Horizontal Irradiance (GHI) from the Eppley PSP and voltage readings from solar irradiance sensors (PD333, OPT101P, BPW21R and BS120E0F) measured using a datalogger. In this plot, the sky was cloudy from 05/25 - 05/30. Note the y-axes for GHI is in W/m^2, whereas the one for the SIM readings is in millivolt.

An ideal SIMs deployment would cover all geographical directions around a solar field to monitor all approaching clouds. Covering the proximity of a solar field that is in the correct wind direction the majority of time is the desired approach. To determine the ideal geographical direction of deployment, we use wind speed and wind direction data available in the National Oceanic and Atmospheric Administration (NOAA) database. By covering the area where the cloud density is most likely to approach, our sensor deployment acts as an early warning system for solar energy prediction.

D. Hardware Infrastructure

The GTS node consists of a Crossbow MICAz connected to a Nemsic MTS420/400CC sensor board. The MICAz has a low-power ATmega128L micro controller, 128KB memory, 512KB flash storage and a Texas Instruments CC2420 radio for wireless communications. Over a 51-pin expansion connector, the MICAz is connected to a MTS420/400CC module containing a GPS receiver as well as various other sensors. The MICAz is connected via serial to a plug computer (or Raspberry Pi) to get Internet access.

Our SIM is comprised of a Moteiv Tmote Sky module interfaced with a custom board populated with three solar irradiance sensors further explained in Section III-E. The Moteiv Tmote Sky (or TelosB) module is comprised of an ultra low power Texas Instruments MSP430 F1611 micro controller featuring 10KB of RAM, 48KB of flash, 1024KB of external flash and a Texas Instruments CC2420 radio for wireless communications. For sampling external sensors, the mote has 8 channels of 12 bit ADC inputs. These inputs are exported through the expansion connectors on the mote. The solar irradiance sensors are interfaced to the mote's ADC inputs.

The SIMs are deployed on top of buildings as described in Section III-C. For reasons of redundancy and potential differences in sensor hardware calibration, we are using three irradiance sensors of the same type and take the average value of the three sensors. Values with a difference larger than a std. dev. from the average are discarded from the calculation (reading may indicate a faulty sensor). A data packet of 18 bytes, containing a time stamp and values of three sensors, is sent to a GTS node, where the collected data is transferred to our servers.

We estimate the cost of a SIM node as the sum of a TMote ($80), sensor board ($10), water proof box ($15), utilities ($3) and power sources (batteries or cords $10-20). This makes an
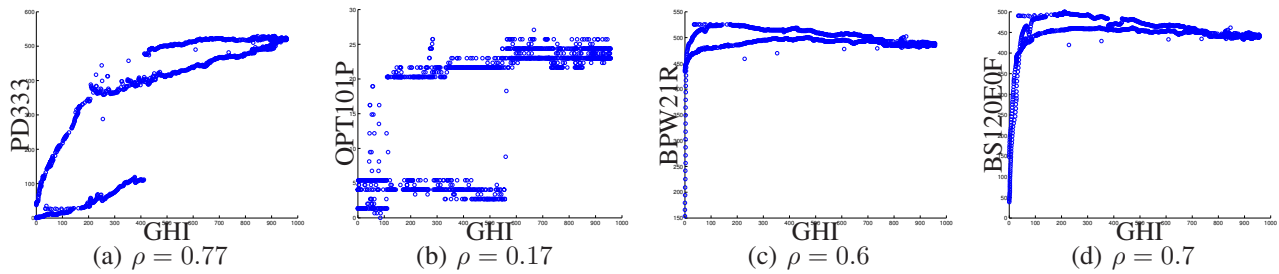
Fig. 5: Scatter plots of Global Horizontal Irradiance (GHI) from the Eppley PSP and voltage readings from solar irradiance sensors (PD333, OPT101P, BPW21R and BS120E0F). ρ indicates the Pearson's correlation coefficient between GHI and sensor readings.
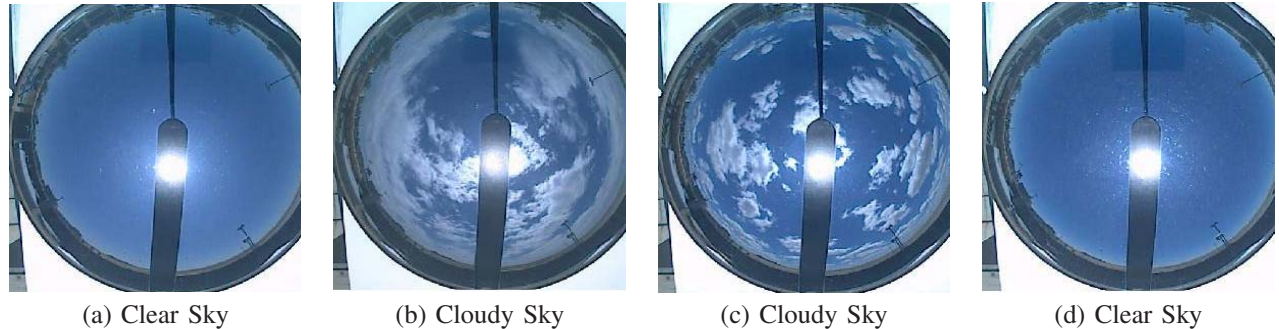

Fig. 6: Images of the sky at noon from the total sky imager co-located with our sensors on particular days shown in Figure 4.

TABLE I: List of solar irradiance sensors.

Sensor Type	Sensor Part#	Spectral Range
Photodiode	BPW21R	350-750nm
Photodiode	BS120E0F	400-800nm
PIN Photodiode	PD333	400-1100nm
Light to Voltage Converter	OPT101	320-1100nm

overall cost of approximately $120 per SIM node. Our SIM node is a self made prototype, a commercially manufactured SIM node can be produced with lower costs and cheaper components.

E. Solar Irradiance Sensors

We experimented with different kinds of light sensors before narrowing down our choice of sensors for the Solar Irradiance Motes. Table I lists four different types of sensors, which were part of our extended evaluation, along with their spectral bandwidth ranges. The photo-diode works as follows: when light falls on the active area of the device, a photo-current is generated proportional to the amount of incident light. By connecting a load resistor (R_L) across the photodiode, we can measure the voltage across R_L, which is proportional to the amount of incident light. In our case, we use $R_L = 1k\Omega$. The light to voltage converter is in IC form with integrated photo-diodes and opamp circuitry. It operates on a supply voltage of >2.7V same as the TelosB mote (3V, 2AA batteries), making it suitable for use in the SIM. We created a PCB to mount these sensors and measured the sensor voltage readings using a Campbell Scientific CR1000 datalogger. Our sensors were co-located with an Eppley Precision Spectral Pyranometer (PSP) [29] and a Total Sky Imager (TSI). Eppley PSP is a radiometer designed for the measurement of total solar radiation (or GHI). Images of the sky are taken at 20 seconds

intervals at the solar observatory station using Yankee Environmental Systems, Inc. TSI-880 instrument [26]. We recorded readings from our sensors and the PSP for four consecutive months. The goals of this evaluation were two-fold: (i) to see how well our sensors measure the total solar radiation in comparison to the Eppley PSP from the solar observatory, and (ii) to see if these off-the-self sensors are able to withstand the harsh conditions (temperature) of an outdoor deployment. Figure 4 shows the readings of GHI from the Eppley PSP and corresponding voltage readings from our solar irradiance sensors for a representative subset of the days. Figure 5 shows the scatter plots of the voltage readings from each sensor and GHI. From Figures 4 and 5, we can observe that out of the 4 sensors, the PD333 output shows the best correlation with the GHI. The BPW21R and BS120E sensors displayed a bimodal output indicating that they are prone to saturation, and are unable to capture the solar irradiance dynamics. The OPT101 sensor was also saturated, did not display any correlation. Figure 6 shows a picture of the sky captured using the TSI. From the sky images, we can see that our PD333 sensor was able to function (correlate with GHI) correctly on days with clear as well as cloudy skies. Based on this data, we decided to proceed with the PD333 PIN photodiode sensor. To improve the correlation between GHI and our sensor, we decided to reduce the saturation effect using a neutral density filter. A neutral density filter reduces and/or modifies intensity of all wavelengths or colors of light equally, thus making it ideal for purpose of reducing the saturation effects. From Figure 7, we observe that the PD333 sensor shows even better correlation with respect to GHI when we introduce a neutral density filter. The Pearson's coefficient between the two is 0.98, indicating very good correlation. Thus, using the PD333 with a neutral density filter is a good choice for inexpensive sensing of solar irradiance.
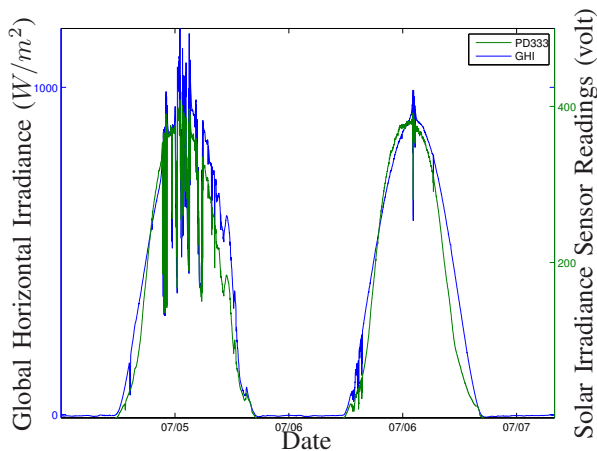
Fig. 7: Global Horizontal Irradiance (GHI) from the Eppley PSP and voltage readings from PD333 sensor when using a neutral density filter (Pearson Correlation $\rho = 0.98$).

F. Software Architecture

In this section, we describe the main components of our software stack that is comprised of time synchronization, data collection, data processing, training and data prediction. Figure 8 shows an overview of our software stack.

1) Time Synchronization: While time synchronizing is critical for our task, our accuracy needs are not very demanding. For our application, time synchronization accuracy needs only to be in the order of high sub-seconds to correlate cloud movement patterns. In addition, we need all sensor nodes to be synchronized in order to forecast at sub-minute intervals. A hard constraint in our design is that GTS/SIM nodes from one area deployment (see Section III-C) cannot communicate with nodes from other *different* areas due to being too far from each other when using the CC2420 RF transceiver. However, all nodes in the system must be synchronized at high sub-second granularity of a common global clock. We explore a number of options to achieve this (e.g. using the 4G hotspot clock, running NTP, etc). We decided to use GPS boards on the GTS nodes to achieve time synchronization among all GTS nodes, since this provides a high accuracy common clock with no bandwidth usage cost. Since we only need to deal with a clock drift of 3.5 seconds every 24 hours on a TMote Sky [30], and in our application all the SIMs are within radio communication range of the GTS, we decided to use a very simple time synchronization protocol, instead of using more accurate time synchronization available [31]. GTS nodes get the GPS time every 30 minutes (to save energy) using the National Marine Electronics Association (NMEA) ZDA sentence which returns the current UTC date and time and synchronize their clocks. Then, the GTS nodes send broadcast timesync packets to all SIM nodes within range with a timestamp of its newly synchronized clock and all SIM nodes synchronize to it. This very simple yet effective method allows us to keep all nodes in the system time synchronized while minimizing complexity and nodes' energy usage.

2) Data Collection: In our SIPS application, one important system parameter we must consider is the sampling rate. There is a clear tradeoff between the sampling rate frequency and the energy consumption. In our case, we considered the dynamics of the phenomena interacting with solar irradiance (i.e. cloud movement and speed), together with the physical separation
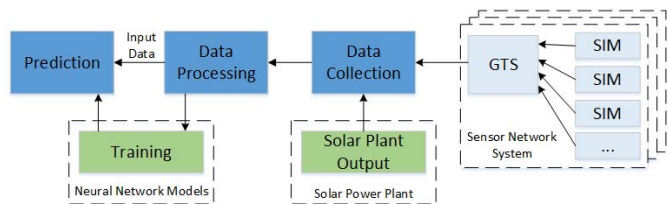

Fig. 8: Software Architecture Overview

distance between the different SIMs in the same deployment area (see Section III-C). In our deployment, SIMs collect a solar irradiance sample every 5 seconds. A high sampling rate is necessary since the irradiance values change quite fast due to cloud movement. High sampling rates enable the sensors to detect a cloud front earlier and more accurately. This is important to achieve high prediction accuracy. However, sampling too fast also increases energy consumption and reduces system lifetime. We tried different sampling rate values but we found that 5 seconds was a reasonable tradeoff between detection accuracy and system lifetime and it worked the best for our application and local conditions.

To perform short term prediction a live data stream of sensor values is necessary. After recording an irradiance sample a SIM sends a data packet containing the sensor value and the current timestamp to the GTS where the data is forwarded to our server via the building's WiFi (when available) or using a 4G WiFi hotspot, and processed to perform prediction. Data collection is only done during times where the solar irradiance is above a specified threshold. This leads to a collection of approximately 8000 irradiance samples per day and sensor.

To train and evaluate our prediction models, it is also necessary to have access to the ground truth, which is the actual energy output of the solar power plant. We put forth substantial engineering efforts to guarantee reliable continuous sampling to obtain an uninterrupted stream of solar plant ground truth data, which proved to be a non-trivial task. All of our equipment was designed to be resistant to various weather conditions. Although the solar array is not in a remote location, the only access to the communication module of the solar array is completely exposed to weather and animals, most commonly cows and sheep (grazing contract) and mice, rabbits, hares and raccoons. We used a rain resistant steel case for the sampling computer, and securely fixed the case to an existing structure above the ground to avoid animals and crawling insects. Another issue is the electromagnetic interference generated by the solar array itself. At certain times of the day, the 500 kW solar power inverters generate interference strong enough to completely block the serial communication used in the sampling computer, which lead to significant data loss. We tried many ways to shield the computer, including sealing the computer in multiple metal boxes of different sizes one inside another one, minimize hole size for running power and serial cables, drilling the smallest holes possible on opposite sides of alternating boxes, filling the space with steel wool, and grounding heavily each box. After many failed shielding attempts, we replaced the serial communication with wireless to solve the problem (we discovered some COTS USB chips are not designed to work in environments with heavy electromagnetic interference). Our system gathers the real time energy output by directly accessing the web interface of the
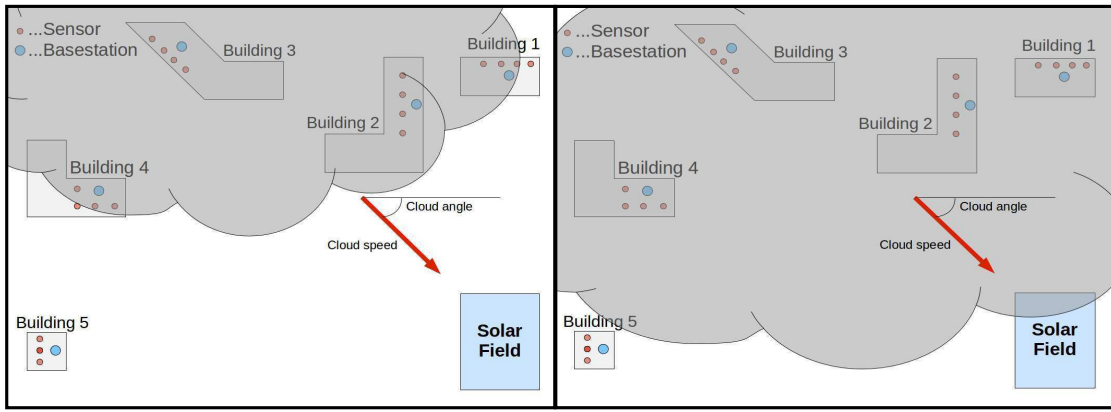
Fig. 9: Illustration showing a cloud front hitting the sensor deployment before covering the solar power plant.
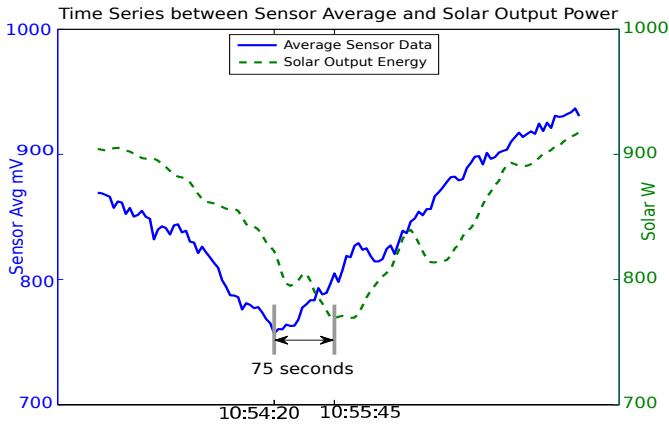

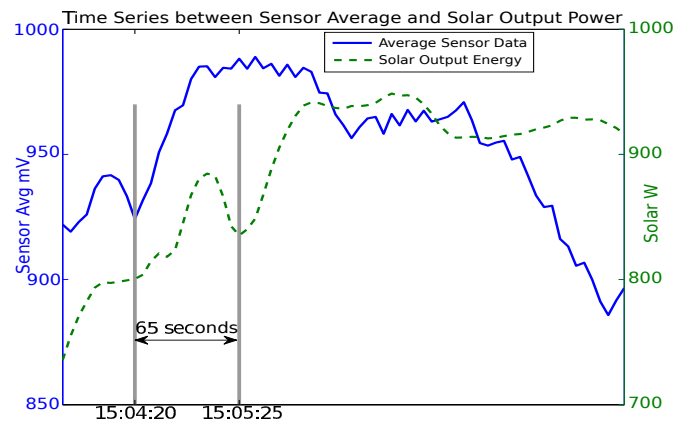Fig. 10: Sensor detecting a cloud 75 seconds earlier than the plant.

Fig. 11: Sensor detecting a cloud 65 seconds earlier than the plant.

internal communication system of the solar array. All the data streams from SIPS are stored in a central database.

3) Data Processing: After the data is collected on our server the values from all SIMs and the solar plant output values are synchronized into time slots of 5 seconds and stored in a data matrix. In order to sanitize the data, if a sensor value of a specific time slot is missing or has a difference larger than a std. dev. from the average, the average of the closest sensors is used to fill this gap. Similarly, if a solar plant data is missing, we use the average of adjacent temporal sample values. The redundancy in our system, explained in Section III-D, helps us addressing the problem of failing sensors and missing data. Since we are using GPS for time synchronization to establish a common timeline, all nodes' readings are sufficiently synchronized. The generated data matrix can now be used by our prediction models to forecast the upcoming solar array output.

We want to predict the energy output of the solar field introduced in Section III-C for different time horizons based on past values of our sensor deployment and past values of the solar plant output power. The data collection and data processing procedure creates multiple parallel time series of the sensor values $x_i(t), x_i(t-1), x_i(t-2), x_i(t-3), ...$ and a time series of the solar field energy output values $y(t), y(t-1), y(t-2), y(t-3), ...$

IV. PREDICTION MODELS

A. Exploratory Analysis

As explained in Section III-D our sensor deployment covers 45 degree slice of geographical area north-west of the solar field and functions as an early warning system for approaching clouds. The location of the sensor deployment makes it possible to sense an approaching cloud front before it hits the solar panels of the power plant and causes a change in the output power as shown in Figure 9. Our sensors are able to measure a decrease or increase of the solar irradiance earlier then it is reflected in the output power of the solar field. Our sensors record a cloud phenomenon between 30 to 120 seconds earlier than the solar field. We are able to determine if clouds are approaching our sensors from the north-west based on the wind speed and direction provided in the NOAA database for the location of our solar field.

Figure 10 and 11 show the behavior between the average sensor output and the actual output power of the solar field. In Figure 10 the solar plant output power is dropping from about 900kW to 770kW within a few minutes. The average sensor values accurately capture these dynamics ahead of time. The solar field output reaches the lowest point of about 770kW at 10:55:45, the sensor in comparison observes the lowest value at 10:54:20 and therefore 75 seconds earlier then the solar plant. Figure 11 shows a similar situation, where the sensor motes observes a specific pattern in the cloud movement 65 seconds
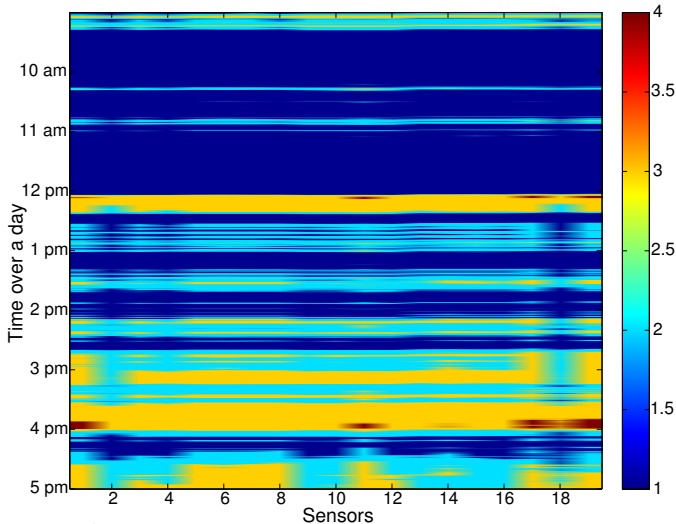
Fig. 12: Cloud map of all 19 sensors over a full day.

earlier compared to the energy output of the solar plant. The time span between recording an approaching cloud at the SIM location and observing a decrease or increase in the output power of the solar field can be increased by moving the SIM sensors to different locations with a greater distance from the solar field. There is a tradeoff between the time span to increase the potential prediction horizon and the correlation accuracy between sensor values and output power. Clouds change in size and composition while they are moving. This means the further away a sensor is placed from the actual solar array, the lower the correlation between sensor values and solar output power. Figure 12 shows a map of all 19 sensors displaying the different irradiance levels over a full day. In the cloud map, a value of 1 (blue) represents a clear sky and a value of 4 (red) represents a very dense cloud causing a huge drop in the solar power output. The map shows the clouds' dynamics and how a cloud front affects the sensors at different times. As shown in Figure 12, our sensors experience huge differences in cloudiness during one day and we are still able to calculate an accurate prediction as presented in Section V.

Our first attempt to create a prediction model was a mechanistic approach to calculate the speed and angle of an approaching cloud front. Dynamic cloud models like this have been proven elusive as stated in [27], [28]. It cannot be assumed that an approaching cloud front has the shape of a straight line, this makes it almost impossible to calculate the angle of an approaching cloud front with a sensor network deployment as ours. Since clouds are not uniform and their density is constantly changing, it is also very difficult to calculate the speed of clouds based on the readings of two sensors at different locations. Due to these conditions, our approach to calculate cloud speed and cloud angle from our sensor data did not work well. An approach to determine the cloud speed and angle by deploying a wind speed and directions sensor did not show very good results since the wind speed and direction at the ground level do not necessarily correlate to the speed and direction of clouds at a higher elevations. These conditions require a prediction model that is able to calculate an energy output prediction only based on SIM sensor readings and does not require additional information such as cloud speed or angle. To address these conditions, we implemented three different prediction models. A Neural Network without

external sensor input, a Neural Network with external sensor input and a novel algorithm called Peak Matching Algorithm (PMA). We introduce these prediction models in the following sections.

B. Nonlinear Autoregressive Neural Network without External Input (NAR)

A Nonlinear Autoregressive Neural Network with no external input (NAR) predicts the solar field output only based on a time series of the solar plant energy trace. This model does not use any sensor data. It represents the current state of the art to perform solar plant output prediction. The NAR model predicts a series of future values for a time horizon of Δt based on n past values of the solar plant output power $y(t)$ as follows:

$$\hat{y}(t + \Delta t) = f(y(t), y(t-1), ..., y(t-n)) \qquad (1)$$

where $y(t)$ stands for the output of the solar field in kW (kilo Watt) at time t, $\hat{y}(t + \Delta t)$ stands for the predicted value at time $t + \Delta t$. The architecture of a 3-layer (input-hidden-output) neural network with 15 hidden neurons and an input and feedback delay of 2 showed the best results in our evaluations. To create this model we used the Matlab Neural Network toolbox for time series prediction.

C. Nonlinear Autoregressive Neural Network with External Input (NARX)

The NARX prediction model extends the NAR model by including external sensor values. We try two different external inputs, one based on the GHI data collected with a single point Epply PSP radiometer (NARX GHI), and another based on data from our SIM sensors (NARX SIPS). Using the latter as an example, the NARX model predicts a series of future values for a time horizon of Δt based on the previous n values of the solar plant output power $y(t)$ and the previous n values of the SIPS network $x(t)$ as follows:

$$\hat{y}(t + \Delta t) = f(y(t), y(t-1), ..., y(t-n), \\ x(t), x(t-1), ..., x(t-n)) \qquad (2)$$

where $y(t)$ stands for the output of the solar field in kW (kilo Watt) at time t, $x(t)$ stands for the value of our SIM sensors at time t and $\hat{y}(t + \Delta t)$ represents the predicted energy value at time $t + \Delta t$. As for the NAR model, we are also using a 3-layer (input-hidden-output) architecture with 15 hidden neurons and an input and feedback delay of 2. To perform a prediction with the NAR and the NARX model, training on multiple days of data is required. The selection of the training data is critical, especially since the solar irradiance varies depending on the time of year. For example, a training set collected in February is not applicable to perform prediction of the power output in August.

D. PMA: Peak Matching Algorithm

The Peak Matching Algorithm (PMA) calculates a prediction by dynamically comparing the stream of sensor data to the stream of solar plant output data. For the PMA prediction method no training of any kind is required. This gives the PMA a big advantage over Neural Network based prediction models, since the selection of a good training set is a very time

consuming and difficult task in the case of solar irradiance data. The PMA is based on calculating the current time difference between a sensor recording a cloud phenomenon and the solar plant output reflecting this cloud phenomenon. This is done by matching local maxima between the sensor data stream and the solar plant output power stream. Algorithm 1 shows the PMA. To calculate a prediction with the PMA, a continuous stream of sensor data and solar plant power output data is required. In these data streams, which are represented as vectors $\vec{S}$ and $\vec{G}$, the algorithm determines local peaks within a moving window W_l as shown in the first while loop in algorithm 1. After all largest local peaks in sensor vector $\vec{S}$ and solar plant power vector $\vec{G}$ are found, the algorithm compares each peak in $\vec{SensorPeaks}$ to each peak in $\vec{PowerPeaks}$, which is done in the nested for loops shown in algorithm 1. If the timestamp of $SensorPeak$, is smaller than the time stamp of $PowerPeak$ and the difference in time between $PowerPeak$ and $SensorPeak$ is less than P_d, the algorithm matches these peaks. If two matching peaks are found, the current prediction horizon is calculated as well as a factor representing the ratio between the solar plant power output value and the sensor value at the peak. This procedure is repeated within the nested for loops until the best matching peaks within the local peak window W_l are found. Now the predicted solar plant output power value at time $t + horizon$ can be calculated by multiplying the current sensor value at time t with the $factor$ value. After a prediction value is calculated for all 19 sensors, the RSME within the window t-4 to t (window size 5) is calculated for every sensor. The predicted value of the sensor with the lowest error rate is selected as the predicted value for the current time step. This procedure is done for every prediction time step. The window size should be small so that the algorithm is able to react fast and switch to sensors that provide improved prediction quality. The PMA is able to calculate a prediction for a time step based on number of peaks n detected in each local peak window size W_l with a complexity of $O(n^2)$. For the evaluation results presented in Section V we are using a local peak window size W_l of 15 data points and a peak matching distance P_d of 15 data points which correlates to 75 seconds in real time. This prediction method dynamically updates the prediction horizon, depending on the current speed and direction of the cloud movement. To make this method comparable to a neural network with a fixed prediction horizon, we calculate the average prediction horizon of the PMA over a full day and compare the results to a neural network prediction of the same day with the same horizon.

V. EVALUATION

We collected readings from our 19-node deployment over a time period of 6 weeks in March/April 2013. Within this time, we only focus on prediction of days with very cloudy conditions. Solar irradiance prediction on clear days has been extensively evaluated [22], [23], [21].

A. Prediction model evaluation metrics

To evaluate the forecasting quality of our prediction models, we calculate the Root Mean Squared Error of our prediction results compared to the actual solar plant power output. We calculate the RMSE as follows:

Algorithm 1 PeakMatching($\vec{S}$, $\vec{G}$, n, W_l, t, P_d)

$\vec{S}$: vector of the n most recent sensor values
$\vec{G}$: vector of the n most recent solar plant power values
W_l: local peak window size
t: current time stamp
P_d: peak matching distance
n: number of data points in $\vec{S}$ and $\vec{G}$

{Comment: Determine largest peaks in $\vec{S}$ and $\vec{G}$ in the window $i - W_l$ to i and store them}
$i = 0$
while $i \leq n$ **do**
    Add largest peak in $\vec{S}(i - W_l : i)$ to $\vec{SensorPeaks}$
    Add largest peak in $\vec{G}(i - W_l : i)$ to $\vec{PowerPeaks}$
    $i = i + 1$
end while

for all $SensorPeak$ in $\vec{SensorPeaks}$ **do**
    for all $PowerPeak$ in $\vec{PowerPeaks}$ **do**
        {Comment: If SensorPeak is earlier than the PowerPeak and the difference is less than P_d, peaks are matched}
        if SensorPeakTime <PowerPeakTime AND (SensorPeakTime - PowerPeakTime) <P_d **then**
            $horizon$ = PowerPeakTime - SensorPeakTime
            $factor$ = PowerPeakValue / SensorPeakValue
        end if
    end for
end for
{Comment: The predicted value at time t+horizon is the sensor value at time t times the factor}
$prediction(t + horizon) = \vec{S}(t) * factor$
return $prediction$

$$RMSE = \sqrt{\sum_{t=1}^{n} \frac{(s(t) - p(t))^2}{n}} \qquad (3)$$

where $s(t)$ is the solar output power and $p(t)$ is the predicted solar output power at time t. In our evaluation, we use the normalized RMSE, i.e. $NRMSE = RMSE/(s_{max} - s_{min})$, with $s_{max} = 1000kW$ and $s_{min} = 0kW$ and the value expressed as a percentage of the output power.

We also evaluate our prediction results with a metric introduced in [28]. By applying the proposed metric we compare our model to a clear sky persistence model, which predicts the next time step $y(t + 1)$ by comparing the measured irradiance to the clear sky irradiance. Our clear sky persistence model is based on the data of a clear day without any clouds. The second plot in Figure 1 shows the output power trace of the solar field on a clear day. The prediction of the output power for the next time step $P(t + 1)$ in the persistence model is calculated by multiplying the power output of the clear sky model $P_{clr}(t + 1)$ with the ratio of the power output $P(t)$ and the clear sky power output $P_{clr}(t)$ as follows:

$$P(t + 1) = P_{clr}(t + 1)\frac{P(t)}{P_{clr}(t)} \qquad (4)$$

This simple persistence model relies on $P_{clr}(t)$, the power output on a clear day at time t, which is time and location dependent. Since we are using data collected at the location of the introduced solar field to calibrate the clear sky persistence
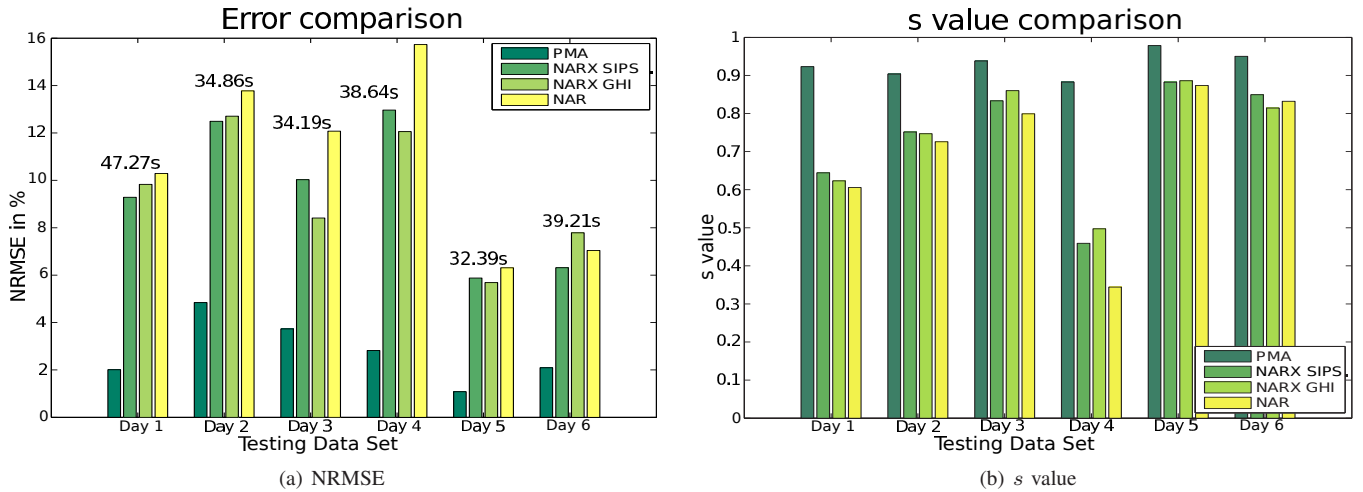
(a) NRMSE (b) s value

Fig. 13: Normalize Root Mean Square Error (NRMSE) and s values of the different prediction models. The numbers on top of the bars in Figure 13(a) indicate the time horizon used.

model, it will only work at our current location and may be error prone for different geographical locations. As stated in [28] we calculate the model quality metric s as follows:

$$s = 1 - \frac{RMSE_M}{RMSE_P} \qquad (5)$$

where $RMSE_M$ is the Root Mean Squared Error of our prediction model and $RMSE_P$ is the Root Mean Squared Error of the clear sky persistence model. The calculated value s tells us the advantage of a forcasting model compared to the clear sky persistence model. A value of $s = 1$ means the prediction is perfect, a value of $s = 0$ means the data variability dominates the forecast. A persistence model has a forecast quality of $s = 0$ by definition. A value of $s < 0$ means that the forecasting model performs worse then a simple persistence model.

B. Evaluation Results

In this section, we compare the prediction results of our NAR, NARX GHI, NARX SIPS and PMA prediction models on days with variable cloud coverage, since those are the most critical days for solar output power dynamics. To compare the forecasting results we calculate the Normalized RMSE (NRMSE) as well as the s value. We test our prediction models on 6 different days with cloudy conditions. To train the Neural Network based models we use data sets of 3-5 days with similar conditions as the training days. The prediction horizon of our testing days is determined by the average prediction horizon of the PMA. For example, if the PMA has an average prediction horizon of 40 seconds over one full day, we also perform a prediction of 40 seconds with the NAR and NARX models on the same day.

Figure 13(a) shows the NRMSE of the different prediction models over 6 different days with cloudy conditions. The average prediction horizon for each day is printed on top of the error bars. Our PMA model has an average accuracy of 97.24% on all testing days and shows significantly better performance than the NAR and both NARX models. The average PMA NRMSE reduction with respect to the NARX GHI, the state-of-the-art model that uses a single point Eppley PSP radiometer, is of 240% (i.e. 3.4 times less), with a maximum reduction of

426% (i.e. 5.26 times less). PMA performance improvement with respect to NAR is even greater, with an NRMSE reduction of almost 300% (maximum of 484%). The NARX SIPS and NARX GHI model only show small improvements compared to the NAR model, which predicts the output power only based on past output power values. Both NARX models show an average reduction of NRMSE of ~15% with respect to NAR, indicating that ANNs model had trouble finding the prediction patterns. We also evaluate the prediction ability of our models based on the s value introduced in section V-A. The s value represents the quality of the prediction model compared to a clear sky persistence model on a scale between 0 and 1, where 1 means perfect prediction and 0 means no improvement compared to a persistence model. Figure 13(b) shows a comparison of the s values for our prediction models calculated for all test days. The PMA model again performs the best compared to the ANN models by showing a very high s value for all testing data sets. The PMA average s value on all test cases is 0.93, with an average improvement of 26% over NARX SISP, 26% over NARX GHI and 33% over NRA. Similarly to the results from the NRMSE, the NARX models with either sensor input perform similarly on average, with an average s value of 0.74 and an average improvement over NAR of near 6%. Figure 14 shows a comparison of the instantaneous error rates between the different prediction models over a sample evaluation day (Day 1). The Y axis shows the under- and over-estimation of the prediction models over time. The PMA shows the lowest error values compared to the other prediction models. On rare occasions, the PMA prediction accuracy is lower than the other models (e.g. at around 13:30 in Figure 14). Sudden changes in wind direction can cause errors like this since the PMA accuracy depends mainly on the current stream of sensor data.

VI. DISCUSSION

The results of the PMA model are significantly better compared to the ANN based models. We believe that ANN models may require significantly more fine tuning of some of non-trainable parameters, such as the number of layers, number of hidden units per layer, etc, to perhaps obtain better results. Another advantage of the PMA model is that no training of any kind is required. This makes the potential application of
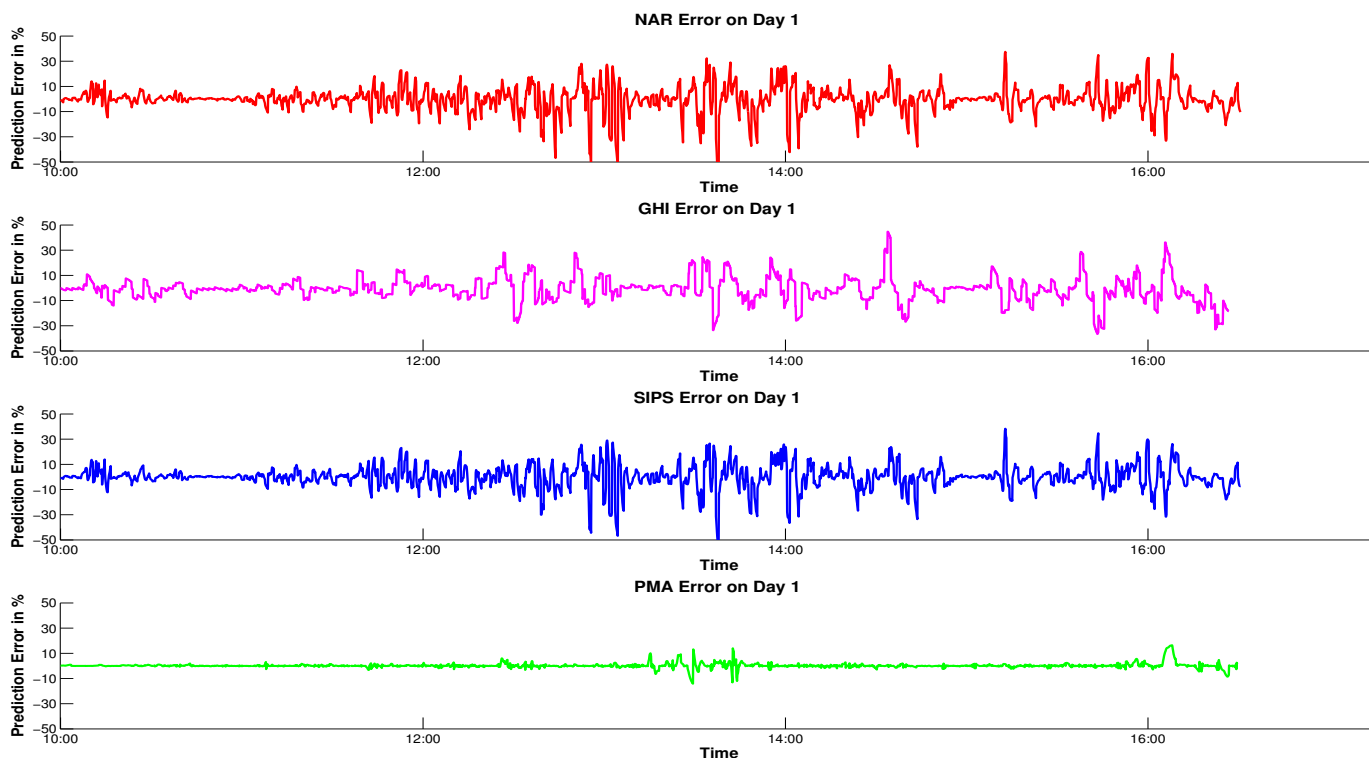
Fig. 14: Instantaneous prediction error over a sample day (10am to 5pm) for the different prediction models. Values around zero are best.
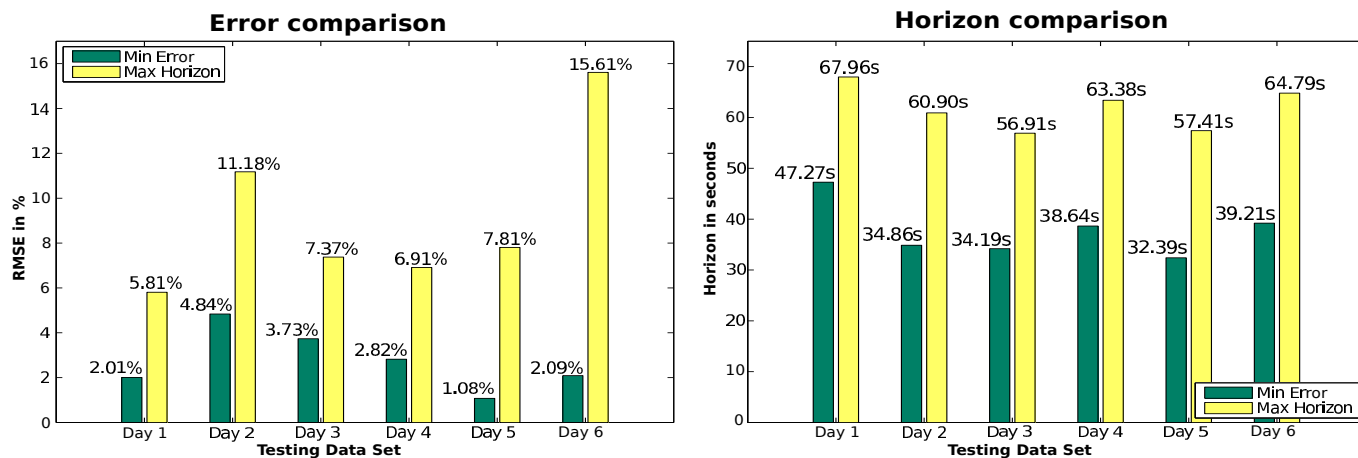

Fig. 15: Comparison between largest prediction horizon and lowest RMSE for the PMA algorithm

the PMA model easier to use in other environments, since ANN models require the creation of a good training data set, which is a very time intensive process. The only weakness of the PMA model is that the prediction horizon depends on the data set and can not be specified by the user. However, users can change the time horizons based on their needs by deploying closer or farther away from the target solar field. Every SIM sensor cluster is deployed in a different location with different distances to the solar power plant, as shown in Figure 3. Therefore, every sensor shows a different prediction time horizon. Our PMA is able to optimize a prediction value for either lowest RMSE or largest prediction horizon. This is done by running the prediction algorithm in parallel on all 19 sensors and dynamically selecting the sensor that either has the largest prediction horizon or the lowest RMSE, which is calculated based on the comparison of the most recent

solar plant output and prediction values within a specified window. Figure 15 show the results for our testing days when optimizing for lowest RMSE or largest prediction horizon.

VII. CONCLUSION

In this paper, we showed an alternative way to approach the problem of short term solar forecasting using a distributed wireless sensor network. Typically, single sky imagers and pyranometers are used for forecasting. However, these devices are limited because of problems such as glaring (for sky imagers) and single measurement point (for pyranometers). In addition, sky imagers and pyranometers are costly devices ($>$ \$2000), making it prohibitive to install in large numbers. We demonstrated that cheap ($<$ \$1) commercial-of-the-shelf light sensors can be adapted to measure the solar irradiance in real environments, and used these data to improve prediction accu-

racy. We evaluated the spatial-temporal correlations between of the power output of a 1 MW solar plant and the recorded sensor values measured by our SIPS infrastructure. For solar irradiance prediction, we introduced NAR and NARX Neural Network models, as well as a novel Peak Matching Algorithm (PMA) prediction model. PMA achieves average prediction accuracy 97.24% while reducing the average NRMSE by 240% when compared with state-of-the-art methods for different short-term time horizons. Our paper shows that wireless sensor networks offer enormous potential as a low cost but high accuracy approach for short-term solar forecasting.

VIII. Acknowledgment

References

[1] "Flexible Ramping Constraint, California ISO," Feb. 2010. 1

[2] H. T. C. P. Y. Chu and C. F. M. Coimbra, "Hybrid intra-hour DNI forecasts with sky image processing enhanced by stochastic learning," *The Official Journal of the International Solar Energy Society*, 2013. 1

[3] M. H. Albadi and E. F. El-Saadany, "Overview of wind power intermittency impacts on power systems," *Electric Power Systems Research*, vol. 80, no. 6, pp. 627–632, 2010. 1

[4] S. Teleke, M. E. Baran, S. Bhattacharya, and A. Q. Huang, "Rule-based control of battery energy storage for dispatching intermittent renewable sources," *Sustainable Energy, IEEE Transactions on*, vol. 1, no. 3, pp. 117–124, Oct. 2010. 1

[5] N. Cristian, A. T. Mahamadou, B. C. Mamadou, and D. Brayima, "Ultracapacitors and batteries integration for power fluctuations mitigation in wind-PV-diesel hybrid system," *International Journal of Renewable Energy Research*, vol. 1, no. 2, 2011. 1

[6] D. P. Hohm and M. E. Ropp, "Comparative study of maximum power point tracking algorithms using an experimental, programmable, maximum power point tracking test bed," in *Photovoltaic Specialists Conference, 2000. Conference Record of the Twenty-Eighth IEEE*, 2000, pp. 1699–1702. 1

[7] W. C. Chi, B. Urguhart, M. Lave, A. Dominquez, J. Kleissl, J. Shields, and B. Washom, "Intra-hour forecasting with a total sky imager at the UC san diego solar energy testbed," *Solar Energy*, vol. 85, no. 2011, pp. 2881–2893, Sep. 2011. 1, 2

[8] R. Marquez and C. F. M. Coimbra, "Intra-hour DNI forecasting based on cloud tracking image analysis," *Solar Energy*, vol. 91, no. 0, pp. 327–336, 2013. 1, 2

[9] G. Werner-Allen, J. Johnson, M. Ruiz, J. Lees, and M. Welsh, "Monitoring volcanic eruptions with a wireless sensor network," in *EWSN*, 2005. 1

[10] T. W. Hnat, V. Srinivasan, J. Lu, T. I. Sookoor, R. Dawson, J. Stankovica, and K. Whitehouse, "The hitchhiker's guide to successful residential sensing deployments," in *SenSys*, 2011. 1

[11] V. Erickson, S. Achleitner, and A. Cerpa, "POEM: Power-efficient occupancy-based energy management system," in *IPSN*, 2013. 1

[12] E. Lorenz, J. Hurka, D. Heinemann, and H. G. Beyer, "Irradiance forecasting for the power prediction of grid-connected photovoltaic systems," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 2, no. 1, pp. 2–10, 2009. 2

[13] R. Marquez and C. F. M. Coimbra, "Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database," *Solar Energy*, vol. 85, no. 5, pp. 746–756, May 2011. 2

[14] P. Mathiesen and J. Kleissl, "Evaluation of numerical weather prediction for intra-day solar forecasting in the continental united states," *Solar Energy*, vol. 85, no. 5, pp. 967–977, May 2011. 2

[15] C. Changsong, D. Shanxu, C. Tao, and L. Bangyin, "Online solar power forecasting based on weather type classification using artificial neural network," *Solar Energy*, vol. 85, no. 11, pp. 2856–2870, Aug. 2011. 2

[16] R. Perez, S. Kivalov, S. J., K. J. Hemker, D. Renne, and T. E. Hoff, "Validation of short and medium term operational solar radiation forecasts," *Solar Energy*, vol. 84, pp. 2161–2172, 2010. 2

[17] A. Hammer, D. Heinemann, E. Lorenz, and B. L. Ckehe, "Short-term forecasting of solar radiation: a statistical approach using satellite data," *Solar Energy*, vol. 67, pp. 139–150, Oct. 1999. 2

[18] A. Mellit, "Artificial intelligence technique for modelling and forecasting of solar radiation data: a review," *International Journal of Artificial Intelligence and Soft Computing*, vol. 1, pp. 52–76, 2008. 2

[19] A. Mellit and A. M. Pavan, "A 24-h forecast of solar irradiance using artificial neural network: application for performance prediction of a grid-connected pv plant at trieste, italy," *Solar Energy*, vol. 84, no. 5, pp. 807–821, May 2010. 2

[20] A. Mellit, H. Eleuch, M. Benghanem, C. Elaoun, and A. M. Pavan, "An adaptive model for predicting of global, direct and diffuse hourly solar irradiance," *Energy Conversion and Management*, vol. 51, no. 4, pp. 771–782, Apr. 2010. 2

[21] L. Martin, L. Zarzalejo, J. Polo, A. Navarro, R. Marchante, and M. Cony, "Prediction of global solar irradiance based on time series analysis: application to solar thermal power plants energy production planning," *Solar Energy*, vol. 84, pp. 1772–1781, Oct. 2010. 2, 9

[22] J. Cao and X. Lin, "Study of hourly and daily solar irradiation forecast using diagonal recurrent wavelet neural networks," *Energy Conversion and Management*, vol. 49, no. 6, pp. 1396–1406, Jun. 2008. 2, 9

[23] A. Sfetsos and A. H. Coonick, "Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques," *Solar Energy*, vol. 68, no. 2, pp. 169–178, Feb. 2000. 2, 9

[24] R. Marquez, V. G. Gueorguiev, and C. F. M. Coimbra, "Forecasting solar irradiance using sky cover indices," *ASME Journal of Solar Energy Engineering,*, vol. in press, 2012. 2

[25] R. Marquez and C. F. M. Coimbra, "A novel metric for evaluation of solar forecasting models," *ASME Journal of Solar Energy Engineering,*, vol. accepted, 2012. 2

[26] YES, *TSI-880 Automatic Total Sky Imager Manual*, Yankee Environmental Systems, Inc., Airport Industrial Park 101 Industrial Blvd. Turners Falls, MA 01376 USA, 2011. 2, 5

[27] S. Kuszamaul, A. Ellis, J. Stein, and L. Johnson, "Lanai high-density irradiance sensor network for characterizing solar resource variability of MW-scale PV system," in *Photovoltaic Specialists Conference (PVSC), 2010 35th IEEE*, Jun. 2010, pp. 000 283–000 288. 2, 8

[28] R. Marquez and C. F. M. Coimbra, "Proposed metric for evaluation of solar forecasting models," *Solar Energy Engineering*, vol. 135, no. 1, pp. 188–197, 2013. 2, 8, 9, 10

[29] "Eppley Precision Spectral Pyranometer." [Online]. Available: http://www.eppleylab.com 5

[30] M. Brzozowski and P. Langendorfer, "On prolonging sensornode gateway lifetime by adapting its duty cycle," in *Wired/Wireless Internet Communications - WWIC*, 2009. 6

[31] M. Maróti, B. Kusy, G. Simon, and Ákos Lédeczi, "The flooding time synchronization protocol," in *SenSys*, 2004. 6