Solar Harvest Prediction Supported by Cloud Cover Forecasts

Christian Renner
University of Lübeck
Lübeck, Germany
renner@iti.uni-luebeck.de

ABSTRACT
Solar harvest prediction is used in energy-harvesting sensor networks to achieve perpetual node operation. Existing approaches only exploit local knowledge and thus fail in unforeseeable, changing weather conditions. We investigate the benefit of incorporating global knowledge in terms of fractional sky cloudiness, so-called cloud cover. We propose and evaluate two methods that combine local information of a node’s harvest pattern with global cloud cover forecasts. We evaluate their performance with solar traces collected by three solar-harvesting sensor nodes and compare the results with existing prediction algorithms. We find that (i) harvest predictions using cloud cover forecasts improve overall prediction precision, (ii) prediction errors in changing weather conditions are considerably reduced, and (iii) coarse-grained cloud cover forecasts require low extra network traffic while sacrificing little prediction precision.

Categories and Subject Descriptors
C.2.4 [Computer-Communication Networks]: Distributed Systems; G.3 [Probability and Statistics]: Time-Series Analysis

General Terms
Algorithms, Performance, Design, Measurement

1. INTRODUCTION
Energy-harvesting sensor networks enable perpetually operating sensor networks. In particular, solar harvesting has drawn considerable attention [12, 3, 8], since it is applicable in most outdoor deployments. The amount of harvested energy, also called harvest, is yet limited, so that sensor nodes must still employ power-saving mechanisms such as radio duty cycling. Moreover, the harvest is unstable due to massive inter-day and intra-day fluctuation, seasonal effects, dirt, and hardware aging. In addition, the environment has a notable impact, e.g., shades of buildings and trees reduce the harvest considerably. For a network of tens or hundreds of nodes, planning the dimensions of the solar cells and the duty cycle for each node is noneconomical or even impossible.

For this reason, algorithms have been devised that adjust the duty cycle or task schedule of sensor nodes online to achieve energy-neutral (i.e., depletion-safe) operation [5, 11, 4, 15]. Rather than saving as much energy as possible, these algorithms aim at maximizing the energy-neutral node consumption in order to achieve goals such as low network delay or high sensing rates. They rely on methods to predict the future harvest [5, 13, 1, 2], preferably long-term predictions of at least one day. Existing prediction algorithms only exploit local historical observations of the harvest. Particularly in changing weather conditions, when a cloudy day follows a sunny day and vice versa, these methods fail and produce erroneous predictions. In the worst case, they hazard energy-neutral, perpetual operation with too optimistic predictions, and they waste energy by causing full energy buffers with too pessimistic predictions.

Using global weather forecasts has the potential to elevate the precision of harvest predictions. Here, cloud cover is promising, because it exhibits a correlation with solar irradiance [14, 6, 10]. In addition, cloud cover forecasts have a small data size—they usually have a resolution of up to one hour and are provided as eights of the sky being covered with clouds—so that they can be distributed into the network within a single packet per hour. However, cloud cover information cannot reflect the local harvest pattern, that is, the environment has a notable impact, e.g., shades of buildings and trees reduce the harvest considerably. For a network of tens or hundreds of nodes, planning the dimensions of the solar cells and the duty cycle for each node is noneconomical or even impossible.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

ENSSys’13, November 13 2013, Roma, Italy
Copyright is held by the owner/author(s). Publication rights licensed to ACM. ACM 978-1-4503-2432-8/13/11 ...$15.00.
http://dx.doi.org/10.1145/2534208.2534210

2. HARVEST PREDICTION FOR SENSOR NETWORKS
Several researchers have tackled the problem of light-weight algorithms for generating solar harvest predictions for sensor nodes. Here, the main focus is on low computation power and local reck-
Figure 1: Excerpt of a solar current trace (red / filled curve) and mean harvest in slots (blue) for $S = 12$

2.1 Harvest Predictions with Time Slots
A day is divided into $S$ time slots, where each slot is assigned a slot value $\mu_{d,s}$ on day $d$. Most algorithms assume static time slots of equal length, e.g., 12 slots with a length of 2h each. In each day $d$ and time slot $s$, the mean value $h_{d,s}$ of the harvest—which can be, e.g., the solar current or power— is calculated and assigned to the corresponding slot value, i.e., $\mu_{d,s} = h_{d,s}$. Figure 1 shows an example of the actual harvest and the mean values $h_{d,s}$. The fundamend of all harvest prediction algorithms is to smooth the values $\mu_{d,s}$ of the same slot $s$ along subsequent days $d$. These smoothed values, denoted $\tilde{\mu}_{d,s}$, are then used to predict future slot values $\tilde{\mu}_{d,s}$. In the following, we explain different methods for averaging slot values and deriving predictions.

EWMA: Exponentially Weighted Moving Average
In [5] Kansal et al. obtain smoothed slot values by applying an EWMA filter:

$$\tilde{\mu}_{d,s} = \alpha \cdot \tilde{\mu}_{d-1,s} + (1 - \alpha) \cdot \mu_{d,s} \quad (0 < \alpha < 1).$$

Their intention is to compensate daily fluctuation, seasonal effects, and harvester aging. The averaged value $\tilde{\mu}_{d,s}$ of a time slot $s$ in day $d$ is used as harvest prediction for the same time slot $s$ in the next day $d + 1$, i.e.,

$$\tilde{\mu}_{d+1,s} = \tilde{\mu}_{d,s}.$$ (2)

The method is light-weight, because only $S$ values have to be stored and the arithmetic complexity is low. Theoretically, predictions with a horizon of multiple days are possible. However, this comes at the cost of achieving predictions with low precision as discussed in [16, 1]. In particular, there is no mechanism to react to changing weather conditions.

WCMA: Weather-Conditioned Moving Average
To solve the shortcomings of EWMA predictions w.r.t. changing weather conditions, the authors of [13] proposed the WCMA prediction algorithm. It differs in three aspects:

1. It averages slot values over the latest $D$ past days with

$$\tilde{\mu}_{d,s} = \frac{1}{D} \sum_{k=0}^{D-1} \mu_{d-s,k} \quad (D \geq 1).$$ (3)

2. A scale factor $\theta$ describes the relation of averaged slot values with those of the current day through

$$\theta = \left( \sum_{k=0}^{K-1} \frac{1}{k+1} \right)^{-1} \sum_{k=0}^{K-1} \frac{1}{k+1} \cdot \frac{\tilde{\mu}_{d,s-k}}{\bar{\mu}_{d,s-k}}$$

(4)
to compare the trend of the current day with the average.

3. The actual prediction of the time slot directly following slot $s$ is a weighted sum of the mean harvest of slot $s$ and the scaled average value of the following slot:

$$\tilde{\mu}_{d,s+1} = (1 - \beta) \cdot \tilde{\mu}_{d,s} + \beta \cdot \theta \cdot \tilde{\mu}_{d-1,s+1}.$$ (5)

Storage space is $S \cdot D$ slot values and computation cost is higher than that of EWMA. WCMA is designed to predict the directly following slot but can also be used to predict slot values other than the directly following ones. For this case our previous work in [16] suggests $\beta = 1$ due to a lack of correlation between the mean values of non-adjacent slots.

WCMA-PDR: WCMA w. Phase Displacement Regulator
Bergonzini et al. have improved WCMA in [2]. They noted a pattern of prediction errors that they call phase displacement. Their algorithm tracks the prediction error in each slot with an EWMA filter. A phase displacement value is calculated by comparing the prediction errors of adjacent time slots and is added to the WCMA prediction value. The evaluation in [2] shows an improvement of prediction precision for predictions of the following slot only. However, WCMA-PDR increases the memory and computation demand compared to WCMA and introduces additional parameters. Since the algorithm is particularly tailored for the case $\beta < 1$, we do not consider WCMA-PDR for long-term predictions.

Pro-Energy
The algorithm in [19] identifies characteristic harvest patterns (slot values) of typical days online—the authors name sunny, cloudy, and rainy days as examples. Harvest predictions are generated by picking the pattern that has the closest match within a given comparison window. The algorithm supports short-term and mid-term predictions (few minutes up to several hours) and increases the prediction precision compared to EWMA and WCMA. Long-term predictions (of a whole day) have not been evaluated by the authors. It is also not clear how well the prediction algorithms works during the night in absence of harvest, when patterns are equal.

2.2 Model-Based Harvest Prediction
Sharma et al. present a method for using global solar irradiance and cloud cover forecasts to predict the future harvest in [17]. They use a polynomial model for the relationship between time of a day and solar power. To cope with seasonal effects, one parameter set is generated for every month of the year. The authors suggest a linear down-scaling of the anticipated maximum solar power by means of the per cent cloud cover. The method does not handle influences of, e.g., hardware aging and shades caused by buildings or trees.

Machine learning for predicting solar power was discussed in [18]. SunCast predicts sunlight conditions by finding the most similar
historical sunlight trace [9]. Unfortunately, these algorithms overshoot the resource demands offered by sensor nodes.

3. Harvest Prediction with Cloud Cover

Existing harvest prediction algorithms for sensor nodes succeed in identifying the average harvest pattern of a node. This is an important feature due to the individual harvesting patterns as explained in Sect. 1. In changing weather conditions, however, they fail due to their restricted, local knowledge. Using global knowledge, such as cloud cover forecasts, mitigates this weakness, but it does not incorporate a node’s individual harvest pattern. We hence argue that combining both methods satisfies the need for incorporating both local harvest patterns and global weather conditions into predictions.

In the following, we present methods that combine the advantages of EWMA predictions with cloud cover forecasts. First, we explain the relation between solar harvest (current or power), solar irradiance, and cloud cover. Second, we present our prediction methods. Third, we briefly discuss how to distribute cloud cover forecasts in the network.

3.1 Linking Cloud Cover and Solar Harvest

The photovoltaic effect causes a solar cell to produce a current that exhibits an almost linear dependency on solar irradiance \( R \), if the voltage of the solar cell stays reasonably below its maximum voltage. This relationship also holds for solar power and solar irradiance, if the terminal voltage of the solar cell is constant.

Solar irradiance and cloud cover are also related [14, 6, 10]. In specific, analytical models have been proposed that describe the relation \( \varrho \) between irradiance \( R_0 \) at a clear sky and irradiance \( R \) at a relative cloud cover \( C \). Here, \( C \) is the fraction (usually in eighth) of the sky that is covered with clouds, and \( \varrho \) is a function of \( C \) with

\[
R = R_0 \cdot \varrho(C)
\]  
(6)

Two well-studied models are those of Kimball and Laevastu [14, 6, 10]. Kimball proposes the linear model

\[
\varrho(C) = 1 - 0.71 \cdot C,
\]  
(7)

and Laevastu suggests using

\[
\varrho(C) = 1 - 0.6 \cdot C^3.
\]  
(8)

Other models are either variations of the presented ones or use additional information, such as the solar altitude at noon or the latitude of the deployment. Comparative studies as in, e.g., [6] have shown that these bring limited or no improvement over the methods of Kimball and Laevastu.

3.2 Harvest Prediction

We aim at providing a light-weight algorithm—in terms of computation power, memory demand, and the number of parameters—for harvest prediction using cloud cover forecasts. The general idea is to calculate for each slot the possible harvest under a clear sky with Eq. (6) and use it as slot value. To predict the future harvest, the slot value is scaled using the cloud cover forecast.

We suggest two methods that we elaborate in the following. Both methods are based on the EWMA algorithm and use the mean harvest \( h_{d,s} \), the mean relation \( \varrho_{d,s} \) in slot \( s \) of day \( d \), and the cloud-cover forecast \( \hat{\varrho} \). While \( h_{d,s} \) is locally computable by a sensor node, the values of \( \varrho_{d,s} \) and \( \hat{\varrho} \) have to be distributed in the network (cf. Sect. 3.3). Note that \( \hat{\varrho} \) embraces multiple values, one for each slot in the prediction horizon.

3.2.1 Combined Method

This method applies Eq. (6) for each slot by calculating the quotient of mean harvest and cloud cover and by using it as mean slot value

\[
\mu_{d,s} = \frac{h_{d,s}}{\varrho_{d,s}}.
\]  
(9)

The smoothed value \( \tilde{\mu}_{d,s} \) is calculated with an EWMA filter (cf. Sect. 2.1). The future harvest is predicted using the smoothed value and the cloud cover forecast. For slot \( s \) of day \( d + k (k > 0) \), the predicted harvest is

\[
\hat{\mu}_{d+k,s} = \tilde{\mu}_{d,s} \cdot \hat{\varrho}_{d+k,s}.
\]  
(10)

The memory footprint of this approach scales linearly with the number of slots \( S \). The effort of distributing cloud cover forecasts in the network depends on its forecast horizon and time resolution, i.e., the size of the forecast in bytes. Computation complexity is that of EWMA filtering plus one additional division and multiplication each.

3.2.2 Separate Method

The second method tracks the smoothed mean harvest and cloud cover values separately. For each slot \( s \), two smoothed slot values \( \tilde{\mu}_{d,s}^h \) (for the harvest) and \( \tilde{\mu}_{d,s}^\varrho \) (for cloud cover) are stored. They are fed with the mean harvest \( \mu_{d,s}^h = h_{d,s} \) and the mean cloud cover \( \mu_{d,s}^\varrho = \varrho_{d,s} \) in slot \( s \) using EWMA filtering (with same filter coefficient \( \alpha \)).

The future harvest is predicted using these two smoothed values and the cloud cover forecast. In particular, the prediction for slot \( s \) of day \( d + k (k > 0) \) is calculated through

\[
\hat{\mu}_{d+k,s}^h = \frac{\tilde{\mu}_{d,s}^h}{\tilde{\mu}_{d,s}^\varrho} \cdot \hat{\varrho}_{d+k,s}.
\]  
(11)

The memory footprint and computation complexity are twice that of the combined method. The effort of distributing cloud cover forecasts in the network is the same.

3.3 Distribution of Cloud Cover Data

Until now, we have assumed that each node has access to cloud cover information—either in form of \( C \) or \( \varrho \)—for the current time slot and a cloud cover forecast. Apparently, this data is global information, that has to be distributed in the network. Naturally, this distribution should lead to as low communication overhead as possible. One way to solve this problem is to use a distribution method such as Trickle [7]. Another option is to piggy-back the corresponding data in (software) acknowledgments. This is particularly appealing, if there is regular data traffic in the network and since cloud cover information and forecasts are usually only available in an hourly resolution. We plan to investigate this point in detail in future work.

4. Evaluation

Next, we evaluate and compare harvest predictions with and without cloud cover forecasts. Before presenting the results, we introduce the data sets and evaluation methodology.
4.1 Data Sets

We deployed three energy-harvesting sensor nodes (A–C) on different window sills of our University building. Node A was placed on the inside, whereas nodes B and C were on the outside. All nodes were equipped with a miniature solar cell and a supercapacitor as energy buffer. The solar cell produces a maximum current of 35 mA, where the current roughly depends on the radiation (weather conditions) only. A shunt in the ground path of the charging circuit serves as solar current sensor. Further details on the hardware are provided in [15].

All nodes sampled the solar current every 3 s and sent 5 min-averages to the base station. We recorded solar current data from April 2013 until July (88 to 95 days of data per node). Figure 2 shows the statistical distributions of solar current for two nodes throughout the experiment time (nodes B and C performed similarly).

For the same time period, we recorded cloud cover forecasts from an online weather forecast service1. These forecasts have a resolution of one hour and a forecast horizon of ten days. Updated forecasts were available at every full hour.

4.2 Methodology and Metrics

Load adaptation algorithms (e.g., [11]) usually require long-term harvest forecasts to achieve stable node operation. We hence compare the prediction precision of our cloud-cover enabled harvest prediction algorithms from Sect. 3, EWMA and long-term WCMA predictions as described in Sect. 2.1. The prediction horizon is one day, where predictions are updated after every time slot. Initially, all smoothed slot values $\mu_{d,s}$ are zero. We performed an analysis for $S = 6, \ldots, 48$ slots and $\alpha$ from 0 to 1 in steps of 0.1. For WCMA we use $D = 3, 5, K = 3$, and $\beta = 1$. This setup has shown good performance in [16, 2, 1]. Predictions using cloud cover were generated with the relationship functions $g$ from Kimball and Laevastu and averaged to meet the slot length. We added a suffix to Kimball and Laevastu to differentiate between the combined (“1”, one slot value) and separate method (“-2”, two slot values).

To analyze the deviation of the prediction from the real harvest course (ground truth), we calculated the root-mean-square error (RMSE) for every single prediction. This metric shows the deviation within a day and is hence particularly relevant for small energy buffers (e.g., supercapacitors), where intra-day deviation may lead to depletion or a full buffer. To analyze the error at the end of a prediction horizon—i.e., not considering intra-day deviation—we calculated the mean error (ME) for every prediction. This metric is more relevant for large energy buffers that are rarely operated at extreme fill levels (e.g., rechargeable batteries).

Since there is one error value for each individual prediction, there are $S$ values per day of the solar traces. To allow for a setup phase, we only consider predictions after the fifth day (this is the rise time of an EWMA filter with $\alpha \leq 0.8$ to 67% of an asymptotic final value). Both RMSE and ME values are normed by dividing them through the overall mean harvest. Error magnitudes are thus directly compared to the harvest and the mean, energy-neutral consumption of a node. We did not calculate relative errors on a daily basis, because they are extremely large on days with low harvest and a too large prediction and thus distort the results massively.

4.3 Results

4.3.1 Comparison of Prediction Precision

The analysis of the RMSE shows a benefit of using cloud cover forecasts for harvest prediction. Using $S = 12$ or 24 gives a good trade-off between precision and memory consumption in all cases. Figure 3 portrays the mean normed RMSE for nodes A and C for different numbers $S$ of slots (results for node B are similar to node C). Results for different values of $\alpha$ (and $D$) are similar, where best values are achieved for $0.6 \leq \alpha \leq 0.9$ and $D = 3$.

The advantage of cloud-cover-enabled predictions ranges from 10–20%. Kimball conversion performs slightly better than Laevastu for our data set; the method of prediction (separate vs. combined) has a minor impact on the RMSE only. WCMA performs worst because its scaling method fails for slots further in the future—e.g., a

1www.wetterspiegel.de
better harvest than usual at noon does not (necessarily) imply a better harvest in the morning of the following day. The figure reveals that node C profits earlier from increasing \( S \), because its harvest pattern is more wide-spread over the day (cf. Fig. 2). The narrow harvest window of node A requires additional slots to improve the time resolution of prediction.

The RMSE in Fig. 3 may appear unexpectedly high (it ranges from 1.1 to 2.3 times the average harvest). The concept of time slots already produces a large normed RMSE between the actual traces and their slot representation (cf. Fig. 1); e.g., the normed RMSE for node A ranges from 0.81 (\( S = 48 \)) to 1.43 (\( S = 6 \)). For node C, it ranges from 0.50 to 0.96. We omit a comparison of prediction precision without the bias of time slots due to space constraints.

Next, we studied the distribution of the ME and found that all methods produce a relatively unbiased (close to zero) ME. However, EWMA and WCMA produce larger ME values, since they cannot adequately compensate for changing weather conditions. Figure 4 shows the cumulative distribution function (CDF) of absolute ME values. Both plots exhibit a notable improvement of Kimball-1 over EWMA and WCMA; e.g., 60\% of all ME values of node A are at most 0.37 for Kimball-1, 0.45 for EWMA, and 0.55 for WCMA. The results for other values of \( S \) and \( \alpha \) are comparable.

In addition to statistical metric evaluation, we investigated the temporal course of prediction errors. Figure 5 shows harvest and prediction ME for an excerpt of 21 days for node A. It shows that particularly upon changing weather conditions, Kimball-1 has an improved precision (ME closer to zero) over EWMA. This is, e.g., visible on day 61 and the series of days starting with day 71, where the weather conditions and thus harvest increase from day to day. In some cases, e.g., day 69/70, erroneous cloud cover forecasts (not shown in the figure) prevent an improvement. The figure also reveals that WCMA produces small errors particularly on days with low harvest (e.g., days 61 and 62). However, it produces large errors on days 76 and 79.

### 4.3.2 Influence of Cloud Cover Resolution

In the following, we explain the influence of the resolution of cloud cover forecasts. For this purpose, we analyzed the mean normed RMSE produced by Kimball-1 with resolutions ranging from 1 h to 24 h by averaging the Kimball-converted cloud cover values over the corresponding time window (resolution). Note that only the resolution of the forecast is altered. The values used in Eq. (9) still have the resolution of the slot length.

![Figure 4: Distribution of normed mean absolute errors for \( S = 24 \) and \( \alpha = 0.7 \) (\( D = 3 \) for WCMA)](image)

![Figure 5: Influence of the temporal resolution of cloud cover forecasts on the mean normed RMSE for Kimball-1 with \( \alpha = 0.7 \) for node C](image)

![Figure 6: Influence of the temporal resolution of cloud cover forecasts on the mean normed RMSE for Kimball-1 with \( \alpha = 0.7 \) for node C](image)
Acknowledgments

The author would like to thank Prof. Volker Turau from Hamburg University of Technology for providing energy-harvesting sensor nodes.

6. REFERENCES


Figure 5: Excerpt of the harvest trace (ground truth, upper plot) of node A and mean prediction errors at the beginning of each slot. Prediction parameters are $\alpha = 0.7$ ($D = 3$ for WCMA) and $S = 12$ with a prediction horizon of one day. Values closer to zero are better.


