Time Dilated Bundt Cake Analysis of PV Output

Giray Ogut[1], Bennet Meyers[2], Aramis Dufour[1], and Stephen Boyd[1]
[1] Stanford University, Stanford, CA, 94305, USA
[2] SLAC National Accelerator Laboratory, Menlo Park, CA, 94025, USA

***Abstract*—We present a method for modeling the power generated by a photovoltaic (PV) system that takes into account seasonal variation. The method is interpretable and auditable, and works directly from observed PV output data, which can include missing data. It relies on multiperiodic basis functions and convex optimization, and so is reliable and efficient. The first step is to model the variation in PV sunrise and sunset times across the year. We then time dilate the original PV signals, given in uniform time segments, into uniform 'PV days', which start at PV sunrise and end at PV sunset, which vary over the year. A 3D plot of this time dilated data shows the variations of cloud and obstruction effects across a year, and resembles a Bundt cake, which gives the method its name. We can then use a multiperiodic basis to fit marginal quantiles of PV output, taking into account variation over the year and within one PV day. These quantiles can be used for several applications, such as anomaly detection or automatic clear sky modeling.**

***Index Terms*—convex optimization, time dilation, quantile estimation, clear sky detection**

I. Introduction

Given the growing amount of data produced by photovoltaic (PV) power generation systems, there is a need for data driven methods for modeling, analyzing, predicting, and controlling PV systems. Classic approaches rely on physical models, as described in [1], and typically only provide point, not distributional, estimates. Other data driven approaches, such as those based on neural networks, are black-box and not interpretable or auditable [2]. In this paper we present a white-box machine learning method based on convex optimization [3] for modeling the time dependent statistics in measured PV power signals. This model describes the probability of observing a certain power output from the PV system on a particular date and time. We demonstrate an application of this model here for detecting 'clear sky' periods, when the system was unaffected by cloud occlusions. However, our future application of this method will be as a step in a pipeline for estimating the joint probability distributions of a fleet of PV systems. This work is a continuation of our previous work on modeling PV power signals [4], [5] and modeling smooth, periodic processes [6]. Novel to this work is the introduction of a 'time dilation' which removes nighttime values and creates daily signals with an equal number of measurements, from sunrise to sunset.

A. Related work

a) Time dilation: Time dilation, also known as dynamic time warping, is a technique used to align two time series by warping time to minimize distortions, allowing for the comparison of signals that may vary in speed [7]. Time dilation is employed in various fields such as biology, economics, signal processing, finance, and robotics for tasks including pattern recognition, signal prediction, and measuring similarities between time series [8]–[11]. Despite its versatility, time dilation can produce sharp irregularities, known as singularities, where multiple points in one signal map to a single point in another [12]. To address these issues, recent advancements like Generalized Dynamic Time Warping (GDTW) frame time dilation as a continuous optimization problem, penalizing misalignments and smoothing the warping function [13].

b) Quantile estimation: The quantile or pinball loss has been long used to estimate the conditional quantiles in statistical fitting problems [14], [15], and empirical quantile estimation continues to be an active area of research [16]. This research belongs to the broader domain of nonparametric probability models [17]. The fitting of time-varying quantiles, specifically related to seasonality (*i.e.*, with periodic structure), has been explored in hydrological sciences [18], but remains relatively unexplored in the PV space.

c) Clear sky detection: Clear sky detection is the task of labeling time series power or irradiance data when no clouds were occluding the system, and is typically carried out by comparing the measured signal to a physical model of clear sky irradiance [19]. There exist at least 70 models published in the literature for estimating the clear sky solar irradiance at a given location [20]. These models determine the local solar position and influence of atmospheric conditions at the site based on a variety of parameters that can be generally grouped into two categories: solar geometry inputs and atmospheric parameters inputs. These methods require significant inputs besides a power (or irradiance) time series and are not robust to issues like site shading. Additionally, these models can be difficult to tune for local site conditions [21]. Our work here is follows in the spirit of our previous work on directly estimating clear sky models from data, by assuming smooth, multiperiodic statistics and minimizing a quantile loss [22].

d) PV sunrise/sunset estimation: Traditional methods for PV sunrise/sunset estimation use geographic coordinates and date to compute solar position angles [23], and these methods are implemented in popular software tools like `pvlib` [24] and `PVsyst` [25].

Data driven methods have been increasingly applied to estimate PV sunrise and sunset times, and the time dilation work presented here is in this category. Packages such as `solar-data-tools` [26] use signal decomposition framework to find PV sunrise and sunset times [27]. The advantage of these approaches is that system location (*i.e.*, latitude and longitude) are not required; the analysis can be carried out

directly on the power time series of interest without the need for external data. These approaches also capture the effects of occlusions.

II. Bundt cake representation

We are given a time series (or signal) of power output of a single PV system

$$y = (y_1, \ldots, y_T) \in (\mathbf{R}_+ \cup \{?\})^T,$$

where $\mathbf{R}_+$ denotes the nonnegative real numbers, and $y_t = ?$ means that the value y_t is missing. The time index $t = 1, \ldots, T$ corresponds to a fixed time interval such as 15, 5, or 1 minutes, and y_t is the average power over that interval. We let P denote the number ot time periods in one 24 hour period, so the examples above correspond to $P = 96$, $P = 288$, and $P = 1440$. When indexing data the index t is an integer, but we will also use it to refer to times in between the time interval boundaries, where it is a real number. For example $t = 123.4$ refers to the time instant 40% of the way through the 123rd time interval.

When converting to absolute date/time, t represents the beginning of the time interval. For example if $t = 0$ corresponds to midnight, and $P = 96$, then $t = 5$ refers to the 15 minute time interval from 1:15–1:30, p_5 is the average PV power over that interval, and $t = 5.33$ corresponds to 1:20.

When t corresponds to night, we have $y_t = 0$. During day, the PV signal varies over a day, with a pattern that changes over the year due to varying geometry (including possible obstructions), and in addition is subject to changes (mostly reductions) due to cloud cover. The statistics of the weather induced PV output changes also varies over the year. We seek to model these patterns directly and more or less automatically from the observed data.

Our first step is to estimate the *PV sunrise* and *PV sunset* over the year, defined as the time when the PV output typically first rises above zero and the time when the PV output reduces to zero, in the absence of clouds. The PV sunrise and sunset vary smoothly over the year, and are given by values of t that need not be integers. We define the *PV day* as the period between the PV sunrise and sunset. The length of the PV day varies over the year.

We then transform the data from a time series indexed by time index t into a matrix where each row represents one PV day, divided into M uniform time segments. We denote these row vectors as $(x_d)_m^T$, where $x_d \in \mathbf{R}^M$. The index d represents a day (date), which ranges from $d = 1$ to $d = D = \lceil T/P \rceil$. The index $m = 1, \ldots, M$ denotes the PV day time index. We interpret $(x_d)_m$ as the PV output in day d in the interval m of the PV day, *i.e.*, at a time that is the fraction m/M through the PV day. Since the length of the PV day varies over the year, the time interval represented by the index m corresponds to different amounts of (real) time over the year. We denote the dilated data as a matrix

$$X \in (\mathbf{R}_+ \cup \{?\})^{D \times M}.$$

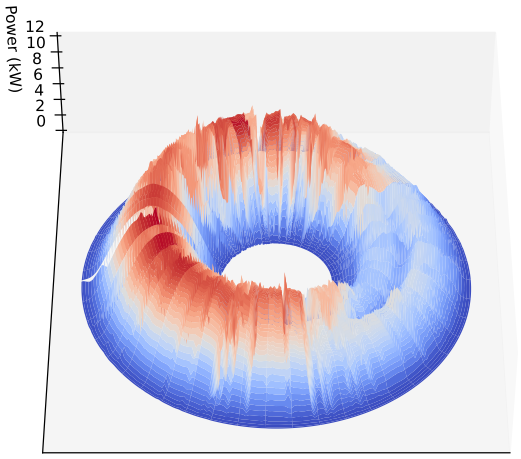

Fig. 1: Bundt cake plot of time dilated PV output from 07/01/2015 to 06/30/2016.

We expect the dilated data X to have approximate annual periodicity, *i.e.*, we expect $X_{d+365,m} \approx X_{d,m}$. We also expect the rows to begin and end near zero, *i.e.*, $X_{d,m}$ is small when m is small or near M. On a clear day we expect $X_{d,m}$ to rise as m increases to PV noon, $m \approx M/2$, and decrease after that as t increases to PV sunset, $m = M$.

The time dilated PV output can be visualized by plotting X in 3D with height denoting PV output on an annulus, with radius corresponding to the PV time, and angle corresponding to the day. An example (with no missing data) is shown in figure 1. (We describe the details of this data later.) The outer ring corresponds to PV sunrise and the inner ring corresponds to PV sunset. The angle corresponds to the day of the year. For obvious reasons we call this a Bundt cake plot. An alternative 2D visualization uses a heat map on an annulus, shown in figure 2. In the next subsections we describe the details of how to carry out this time dilation.

A. PV sunrise/sunset estimation

In this section we describe a simple method for automatically modeling PV sunrise and sunset as a smooth periodic function of the day of the year d. We denote these times as T_d^{rise} and T_d^{set}, for $d = 1, \ldots, 365$. These are real numbers, not integers, measured in the original time series interval units. We expect successive values to be offset by around one day, and to be periodic over the year, *i.e.*,

$$T_{d+1}^{\text{rise}} \approx T_d^{\text{rise}} + P, \qquad T_{d+365}^{\text{rise}} = T_d^{\text{rise}} + 365P,$$

and similarly for T_d^{set}.

We will model these as the alternating zero crossings of an annually periodic function f which has one increasing zero

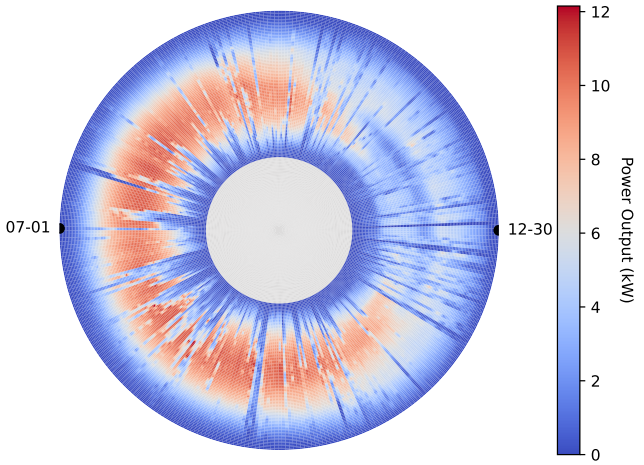

Fig. 2: Heat map annulus plot of time dilated PV output from 07/01/2015 to 06/30/2016.

crossing and one decreasing zero crossing each day. Roughly speaking $f(t) > 0$ corresponds to PV day and $f(t) < 0$ corresponds to PV night.

We now decribe how to construct the function f from the given data. We start by forming a Boolean time series z from our original one y, that indicates whether there is nonzero PV output:

$$z_t = \begin{cases} 1 & y_t \geq \epsilon \\ 0 & y_t < \epsilon \\ ? & y_t = ?, \end{cases}$$

for $t = 1, \ldots, T$. Here ϵ is a positive threshold such as $\epsilon = 0.005 \max_t y_t$, *i.e.*, we consider PV output as nonzero if it exceeds 0.5% of the maximum observed value.

To capture the desired properties of f, *i.e.*, that it have one increasing and one decreasing zero crossing per day, which vary smoothly over the year, we define a set of N basis functions, $\phi_i : \mathbf{R} \to \mathbf{R}$, for $i = 1, \ldots, N$, and take $f(t) = \sum_{i=1}^{N} \alpha_i \phi_i(t)$. We use $N = 9$, with basis functions the constant function 1, the daily Fourier basis functions

$$\cos(2\pi kt/P), \qquad \sin(2\pi kt/P), \quad k = 1, 2,$$

and the yearly Fourier basis functions

$$\cos(2\pi kt/(365P)), \qquad \sin(2\pi kt/(365P)), \quad k = 1, 2.$$

More sophisticated bases can also be used, but we have found this basis gives good results.

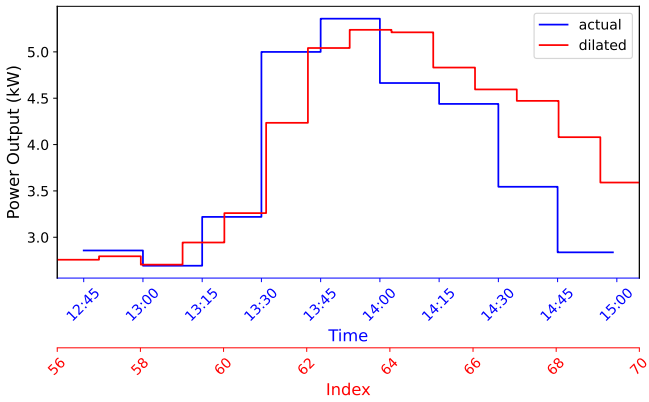

Fig. 3: Actual signal and time dilated signal corresponding to the same time interval.

We find the N coefficients α_i by minimizing the logistic loss over the known data,

$$\sum_{t \neq ?} \begin{cases} -f(t) + \log(1 + \exp(f(t))) & z_t = 1 \\ \log(1 + \exp f(t)) & z_t = 0. \end{cases}$$

This is a smooth convex function and readily minimized. We can include a regularization term when fitting, but given the small number of parameters ($N = 9$), this is not needed in practice.

To approximate the zero crossings of f we use simple linear interpolation. To find the PV sunrise times we identify integers t with $f(t) < 0$ and $f(t+1) > 0$ and take

$$T_d^{\text{rise}} = t - \frac{f(t)}{f(t+1) - f(t)}$$

(which lies between t and $t+1$) as the sunrise time, where d is the day corresponding to t. To identify sunset times we indenify integers t with $f(t) > 0$ and $f(t-1) < 0$, and use a similar interpolation.

B. Time dilation

For each day $d = 1, \ldots, D$, we define a set of $M+1$ uniformly spaced time points between T_d^{rise} and T_d^{set},

$$\tau_m = T_d^{\text{rise}} + \frac{m}{M-1}(T_d^{\text{set}} - T_d^{\text{rise}}), \quad m = 1, \ldots, M.$$

These define M intervals over the PV day, $[\tau_m, \tau_{m+1}]$, $m = 0, \ldots, M-1$. We take $X_{d,m}$ as the integral over the $(m-1)$th interval of the PV output, assumed constant with value y_t over the interval $[t, t+1]$. (This ensures that the total energy over a day is preserved in the time dilation.) If at any time in the interval the PV output data is missing, we assign $X_{d,m} = ?$.

Figure 3 illustrates how a 15-minute average power signal (aligned on quarter hours) is time dilated into a signal of length 100 over the PV day. Here we show dilation between 12:45 and 15:00 on 07/01/2015. Estimated PV daytime on that day was ≈ 813 minutes, so each interval of the dilated PV signal corresponds to around 8.13 minutes.

III. Smooth multiperiodic marginal quantile estimation

In this section show how to estimate marginal quantiles of PV output over the PV days. Let

$$0 < \eta_1 < \cdots < \eta_L < 1$$

denote the L quantiles we wish to estimate. We will find the quantile estimates as $Q^i_{d,m}$, $i = 1, \ldots, L$. We interpret $Q^i_{d,m}$ as our estimate of the η_i-quantile of PV output in time period given by d, m.

We add the constraint

$$Q^{i+1}_{d,m} \geq Q^i_{d,m}, \qquad d = 1, \ldots, D, \quad m = 1, \ldots, M,$$

to prevent the quantile crossing problem, *i.e.*, $Q^{i+1}_{d,m} < Q^i_{d,m}$. We also enforce

$$Q^i_{d,m} \geq 0, \qquad d = 1, \ldots, D, \quad m = 1, \ldots, M,$$

so that all quantiles have nonnegative values.

We model the quantiles using smooth multiperiodic functions as

$$Q^i_{d,m} = c_0 + \sum_{j=1}^{N} c_j B^j_{d,m},$$

with $B^j_{d,m}$ denoting the jth basis function evaluated for entry (d, m). Here $c_0, \ldots, c_N$ are coefficients that we will fit.

Daily basis functions are

$$\sin\left(\frac{\pi k m}{M}\right), \quad k = 1, \ldots, 10.$$

Note that we just use sines as our basis functions for daily periodicity, and start with a period that is twice the overall interval width. This is justified by expanding a Fourier series of a function over an interval that is zero at each endpoint. Yearly basis functions are

$$\cos\left(\frac{2\pi k d}{365}\right), \qquad \sin\left(\frac{2\pi k d}{365}\right), \quad k = 1, 2, 3.$$

We also consider products of daily and yearly basis functions, *i.e.*, cross terms, which allows us to model a daily pattern that varies over the year. An example of a cross term is

$$\cos\left(\frac{2\pi 2 d}{365}\right) \sin\left(\frac{\pi 3 m}{M}\right).$$

Cross terms help us capture joint daily and yearly variations in the signal. For instance, a clear sky noon in winter is expected to be different than a clear sky noon in summer especially if there are seasonal shading effects. An additive model without cross terms would not be able to capture this difference. All together we have 10 daily basis functions, 6 yearly basis functions and 60 cross terms, so $N = 76$. Including the constant coefficient c_0 we have 77 parameters to fit for each quantile.

To fit all quantiles using known entries $X_{d,m} \neq ?$, we solve the problem

$$\begin{array}{ll} \text{minimize} & \sum_{l=1}^{L} \sum_{d=1}^{D} \sum_{m=1}^{M} \ell^{\text{pin}}_l (X_{d,m} - Q^l_{d,m}), \\ \text{subject to} & Q^{l+1}_{d,m} \geq Q^l_{d,m}, \quad l = 1, \ldots, L-1, \\ & Q^1_{d,m} \geq 0, \quad d = 1, \ldots, D, \quad m = 1, \ldots, M. \end{array}$$

Here,

$$\ell^{\text{pin}}_{\eta_j}(u) = \begin{cases} (\eta_j - 1)\,(u) & u \geq 0 \\ \eta_j\,(u) & u < 0 \end{cases}$$

is the pinball or quantile loss for quantile η_j.

We denote the tensor of estimated quantiles as $Q \in \mathbf{R}_+^{D \times M \times L}$. The goodness of fit of our marginal quantiles can be measured using the continuous ranked probability score (CRPS).

IV. Clear sky detection

Clear sky detection is a binary classification task with

$$C_{d,m} = \begin{cases} 1 & \text{clear sky interval} \\ 0 & \text{non clear sky interval} \\ ? & \text{missing data.} \end{cases}$$

The naïve method is to take

$$C^{\text{naïve}}_{d,m} = \begin{cases} 1 & X_{d,m} \geq 0.8 Q_{d,m,L} \\ 0 & \text{otherwise,} \end{cases}$$

i.e., if the average power is at least 80% of the highest quantile estimated for that time interval, and 0 otherwise. This naïve method leads to many transitions between clear and non clear sky intervals, making interpretation harder. To alleviate this problem we introduce a post-processing step that smooths our clear sky estimate.

We form a graph with two nodes per time interval, corresponding to clear and non clear intervals. We have edges between each pair of nodes of adjacent time steps. All together then we have $2M$ nodes and $4(M-1)$ edges. We attach zero weight to each horizontal edge, *i.e.*, an edge that connects two adjacent clear sky or non clear sky edges. We attach a positive weight σ to each non horizontal edge, *i.e.*, an edge that represents a transition between clear sky and non clear sky edges. We attach losses to each node that correspond to the naïve clear sky estimates. If $C^{\text{naïve}}_{d,m} = 1$ we attach a loss 0 for the node corresponding to clear sky, and 1 for the node corresponding to non clear sky. If $C^{\text{naïve}}_{d,m} = 0$ we attach a loss 1 for the node corresponding to clear sky, and 0 for the node corresponding to non clear sky. A path through the graph, from one of the nodes corresponding to $m = 1$ to one of the nodes corresponding to $m = M$, corresponds to a clear sky estimate $C_{d,m}$. The weighted cost of the path is exactly equal to the number of intervals for which $C_{dm} \neq C^{\text{naïve}}_{d,m}$, plus σ times the number of transitions between a clear sky and a non clear sky. A shortest path corresponds to an estimate that minimizes this cost function. Larger values of σ yield estimates that are smoother, *i.e.*, have fewer transitions between clear sky and non clear sky. Such a path is readily found via dynamic programming in $O(M)$ operations. This is illustrated in figure 4.

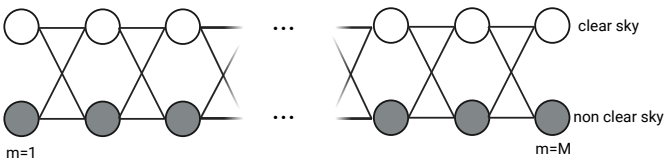

Fig. 4: A path from left to right corresponds to a clear sky estimate. Horizontal edges have no weight, and diagonal edges have weight σ. The nodes have weight zero or one, depending on the naïve clear sky value.

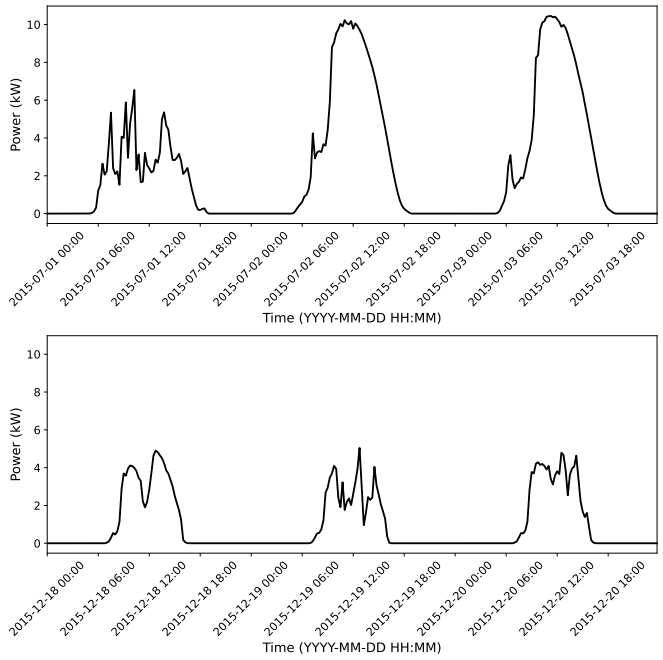

Fig. 5: Subdaily PV output during three consecutive days. *Top.* Summer. *Bottom.* Winter.

V. Results

a) Data: We consider a single residential PV system in Southern California and use data from 07/01/2015 to 06/30/2017 (*i.e.* 731 days), at 15 minute intervals, with no missing data. Figure 5 shows subdaily data for 3 days in summer and 3 days in winter. During daytime the power output is highly variable due to the effect of weather but is mostly similar between consecutive days. We clearly see that the winter days do not look like the summer days. We also see that there is a drop in power output in the middle of the day for winter days, caused by the shading of the PV system by a nearby building.

b) Estimated PV sunrise and sunset times: Figure 6 shows estimated sunrise and sunset times across all days in the dataset. We observe that PV sunrise and sunset times vary throughout the year and are not symmetric around noon. (Recall that PV sunrise and sunset mean the times when the PV output first becomes nonzero, and last is nonzero, and so are affected bu siting and obstructions.) As expected, PV

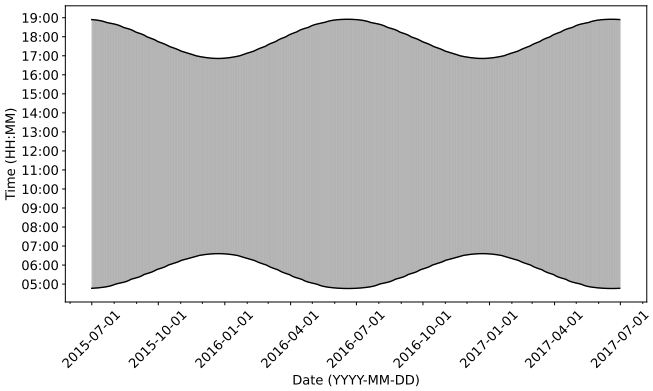

Fig. 6: Estimated PV sunrise and sunset times over entire dataset. The shaded region indicates PV daytime.

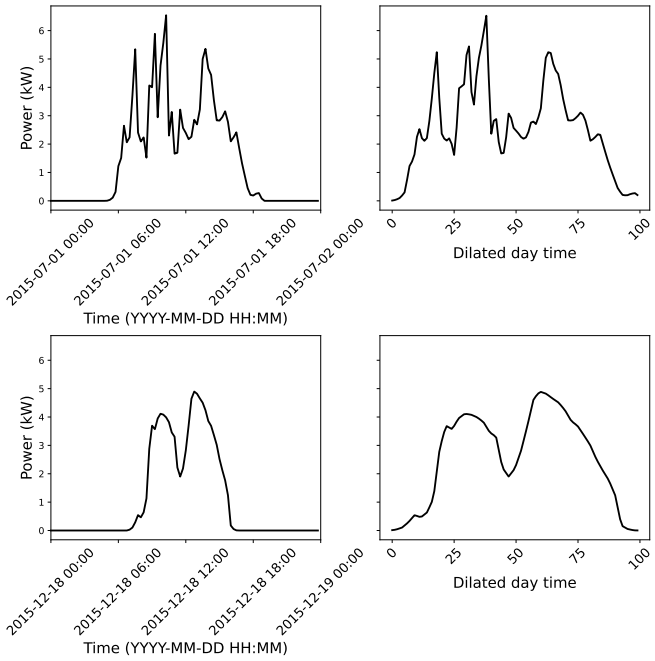

Fig. 7: Time dilation of PV output. *Top.* Summer. *Bottom.* Winter.

sunrise is earlier and sunset is later in summer, resulting in longer PV daytime in summer than in winter.

c) Time dilation: Figure 7 shows time dilation of daytime data for one day in summer and one day in winter, resolved to $M = 100$ samples during the PV day. The lefthand plots show the original data, with time stamps. The righthand plots show the time dilated data versus $m = 1, \ldots, M = 100$. Time dilation 'stretches' time more in winter than in summer.

d) Smooth multiperiodic quantiles: Figure 8 shows the estimated smooth multiperiodic quantiles on undilated PV output. We see that the quantiles are higher in summer than in winter. But not only the scale but also the shape of the quantiles change between different seasons. Also, the distance between quantiles is larger in summer than in winter,

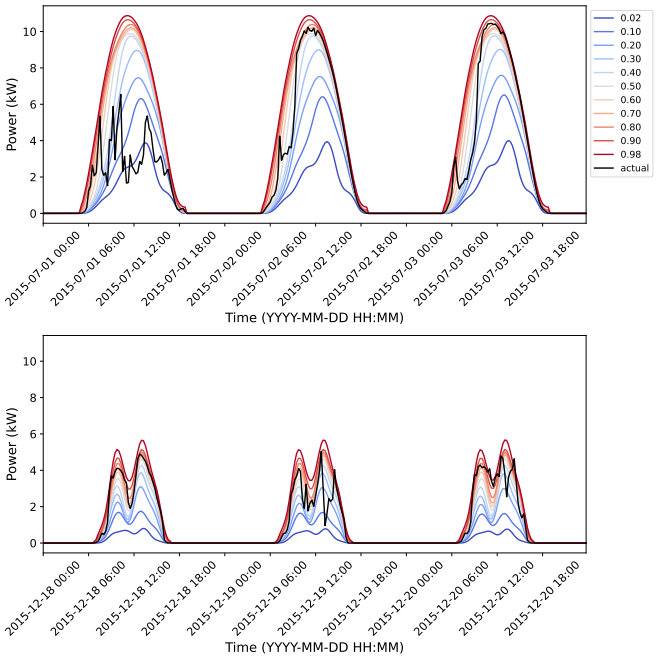

Fig. 8: Estimated quantiles shown on undilated PV output. *Top.* Summer. *Bottom.* Winter.

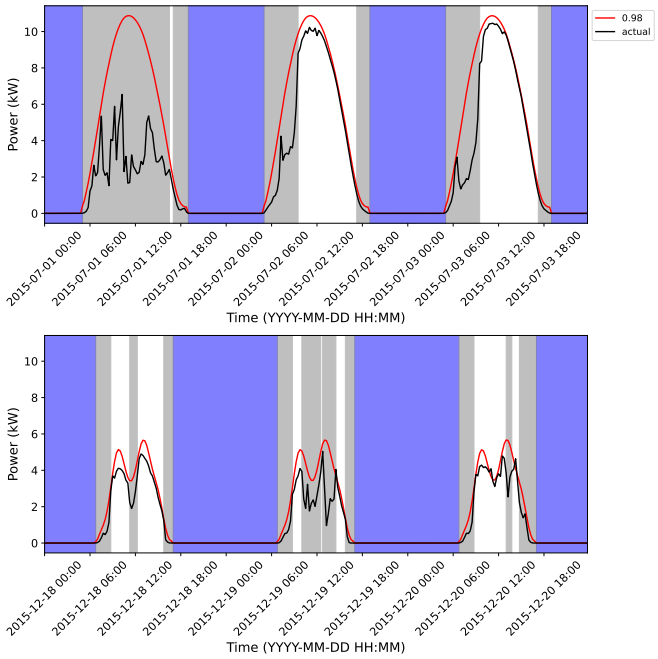

Fig. 9: Detected clear sky time intervals. *Top.* Summer. *Bottom.* Winter.

indicating more variability in PV output.

e) Clear sky detection: Figure 9 shows the detected clear sky time intervals, obtained using $\sigma = 2$, which means that a transition between clear and non clear sky costs a factor of 2 compared to disagreeing with the naïve clear sky estimate. In the plot day time clear sky is shaded white, non clear sky is shaded gray, and night is shaded blue. In these examples we observe that clear sky time intervals are typically longer in summer than in winter.

VI. Conclusions

We have developed an analysis method for PV output data that takes into account the variation in PV sunrise and sunset over the year, time dilating each day's PV output to a fixed length signal, so the PV output fits into a data matrix with columns corresponding to a fixed fraction through the PV day. From this we can form smooth marginal quantile estimates that take into account the variation in the shape of the daily PV over the year, due to trajectory of the sun, obstructions, and weather. We show how to detect clear sky time intervals, trading off deviating from the immediate (naïve) clear sky estimate and excessive transitions between clear and non clear sky. Our methods are based on convex optimization and are therefore reliable, fast, and entirely interpetable and auditable. The methods gracefully handle missing data.

All data and code described in this paper are available at https://github.com/cvxgrp/pv_bundt_cake.

Acknowledgment

This material is based on work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Solar Energy Technologies Office Award Number 38529. Stephen Boyd's work was funded in part by the AI Chip Center for Emerging Smart Systems (ACCESS).

References

[1] M. J. Mayer and G. Gróf, "Extensive comparison of physical models for photovoltaic power forecasting," *Applied Energy*, vol. 283, p. 116239, 2021.

[2] R. Ahmed, V. Sreeram, Y. Mishra, and M. Arif, "A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization," *Renewable and Sustainable Energy Reviews*, vol. 124, p. 109792, 2020.

[3] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2009.

[4] G. Ogut, B. Meyers, and S. Boyd, "PV fleet modeling via smooth periodic gaussian copula," 2023, accessed Jan 21, 2024. [Online]. Available: https://arxiv.org/abs/2307.00004

[5] B. Meyers and S. Boyd, "Signal decomposition using masked proximal operators," *Foundations and Trends in Signal Processing*, vol. 17, no. 1, pp. 1–78, 2023.

[6] G. Ogut, B. Meyers, and S. Boyd, "Interpretable net load forecasting using smooth multiperiodic features," 2024, accessed May 21, 2024. [Online]. Available: https://web.stanford.edu/~boyd/papers/multiperiodic_forecasting.html

[7] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE transactions on acoustics, speech, and signal processing*, vol. 26, no. 1, pp. 43–49, 1978.

[8] S. B. Needleman and C. D. Wunsch, "A general method applicable to the search for similarities in the amino acid sequence of two proteins," *Journal of molecular biology*, vol. 48, no. 3, pp. 443–453, 1970.

[9] F. Itakura, "Minimum prediction residual principle applied to speech recognition," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 23, no. 1, pp. 67–72, 1975.

[10] C. Myers, L. Rabiner, and A. Rosenberg, "Performance tradeoffs in dynamic time warping algorithms for isolated word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 6, pp. 623–635, 1980.

[11] T. Rakthanmanon, B. Campana, A. Mueen, G. Batista, B. Westover, Q. Zhu, J. Zakaria, and E. Keogh, "Searching and mining trillions of time series subsequences under dynamic time warping," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2012, pp. 262–270.

[12] J. S. Marron, J. O. Ramsay, L. M. Sangalli, and A. Srivastava, "Functional data analysis of amplitude and phase variation," *Statistical Science*, pp. 468–484, 2015.

[13] D. Deriso and S. Boyd, "A general optimization framework for dynamic time warping," *Optimization and Engineering*, vol. 24, no. 2, pp. 1411–1432, 2023.

[14] R. Koenker and G. Bassett, "Regression quantiles," *Econometrica*, vol. 46, p. 33, 1978.

[15] R. Koenker and K. F. Hallock, "Quantile regression," *Journal of Economic Perspectives*, vol. 15, pp. 143–156, 2001.

[16] V. Chernozhukov, I. Fernández-Val, and A. Galichon, "Quantile and probability curves without crossing," *Econometrica*, vol. 78, pp. 1093–1125, 2010.

[17] A. J. Izenman, "Review papers: Recent developments in nonparametric density estimation," *Journal of the American Statistical Association*, vol. 86, pp. 205–224, 3 1991.

[18] A. Montanari, "Deseasonalisation of hydrological time series through the normal quantile transform," *Journal of Hydrology*, vol. 313, pp. 274–282, 11 2005.

[19] M. J. Reno and C. W. Hansen, "Identification of periods of clear sky irradiance in time series of ghi measurements," *Renewable Energy*, vol. 90, pp. 520–531, 2016.

[20] F. Antonanzas-Torres, R. Urraca, J. Polo, O. Perpiñán-Lamigueiro, and R. Escobar, "Clear sky solar irradiance models: A review of seventy models," *Renewable and Sustainable Energy Reviews*, vol. 107, pp. 374–387, 6 2019.

[21] M. Grigiante, F. Mottes, D. Zardi, and M. de Franceschi, "Experimental solar radiation measurements and their effectiveness in setting up a real-sky irradiance model," *Renewable Energy*, vol. 36, no. 1, pp. 1–8, 2011.

[22] B. Meyers, M. Tabone, and E. C. Kara, "Statistical clear sky fitting algorithm," *Conference Record of the IEEE Photovoltaic Specialists Conference*, 2018.

[23] I. Reda and A. Andreas, "Solar position algorithm for solar radiation applications," *NREL Technical Report TP-560-34302*, 2008. [Online]. Available: https://www.nrel.gov/docs/fy08osti/34302.pdf

[24] K. S. Anderson, C. W. Hansen, W. F. Holmgren, A. R. Jensen, M. A. Mikofski, and A. Driesse, "pvlib python: 2023 project update," *Journal of Open Source Software*, vol. 8, p. 5994, 12 2023.

[25] A. Mermoud, M. Villoz, B. Wittmer, M. Oliosi, and A. Bridel-Bertomeu, "PVsyst photovoltaic software." [Online]. Available: https://www.pvsyst.com/

[26] B. Meyers, S. A. Miskovich, D. Ragsdale, M. Victoriano, J. Goncalves, A. Dufour, C. Berschauer, A. Londono-Hurtado, T. Takahashi, E. Apostolaki, and S. Sheng, "slacgismo/solar-data-tools: v1.3.2," *Zenodo*, May 2024. [Online]. Available: https://doi.org/10.5281/zenodo.11111485

[27] B. Meyers, "PVinsight (final technical report)," *OSTI*, 2021. [Online]. Available: https://www.osti.gov/biblio/1897181