



Abstract

We present a novel method for modeling time-dependent statistics in the power signal generated by a photovoltaic (PV) system. Our white-box machine learning method is interpretable and auditable, based on principles of multiperiodic basis functions and convex optimization [1]. Our proposed method of time dilating the daily signal to remove night time values results in a novel representation of PV power signals, evocative of a ‘Bundt cake’. The proposed model describes the marginal distribution of power output as a function of date and time. The resulting probabilistic model of a PV system can be used to perform a variety of tasks, and here, we demonstrate the application of clear sky detection.

Method

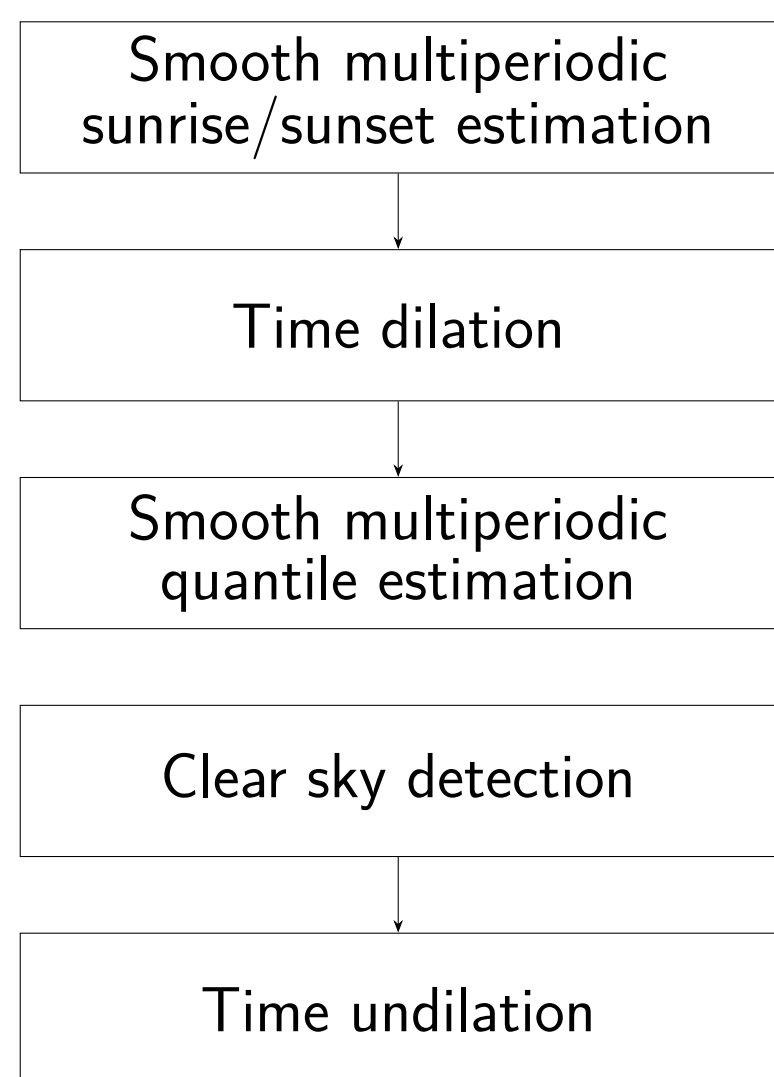


Figure 1. Main steps of the proposed method.

Given a PV power signal $y = (y_1, \dots, y_T) \in \mathbf{R}^T$ possibly with missing values, we use a set of N multiperiodic basis functions $\phi_i: \mathbf{R} \rightarrow \mathbf{R}$, $i = 1, \dots, N$ and estimate power output at each step as a linear combination of these basis functions as follows:

$$\hat{y}_t = \sum_{i=1}^N \theta_i \phi_i(t).$$

These basis functions are sinusoids with different periods and harmonics. For instance, they can be:

$$\begin{aligned} \phi_1(t) &= \cos(2\pi k_1 t / \Pi_1), \\ \phi_2(t) &= \sin(2\pi k_2 t / \Pi_2), \\ \phi_3(t) &= \cos(2\pi k_1 t / \Pi_1) \sin(2\pi k_2 t / \Pi_2), \end{aligned}$$

where Π_1, Π_2 are periods and k_1, k_2 are harmonics. Notice that ϕ_3 is a cross term of ϕ_1 and ϕ_2 . Figure 2 shows some basis functions used in the proposed method. Then, to estimate the coefficients θ_i , we solve the the generic convex optimization problem

$$\text{minimize } \sum_{t=1}^T \ell(y_t, \hat{y}_t) + \sum_{i=1}^N r(\theta_i) \quad (1)$$

where ℓ is a loss function which is determined based on which estimation task we want to perform and r is a smoothness penalty to ensure that the estimates change smoothly.

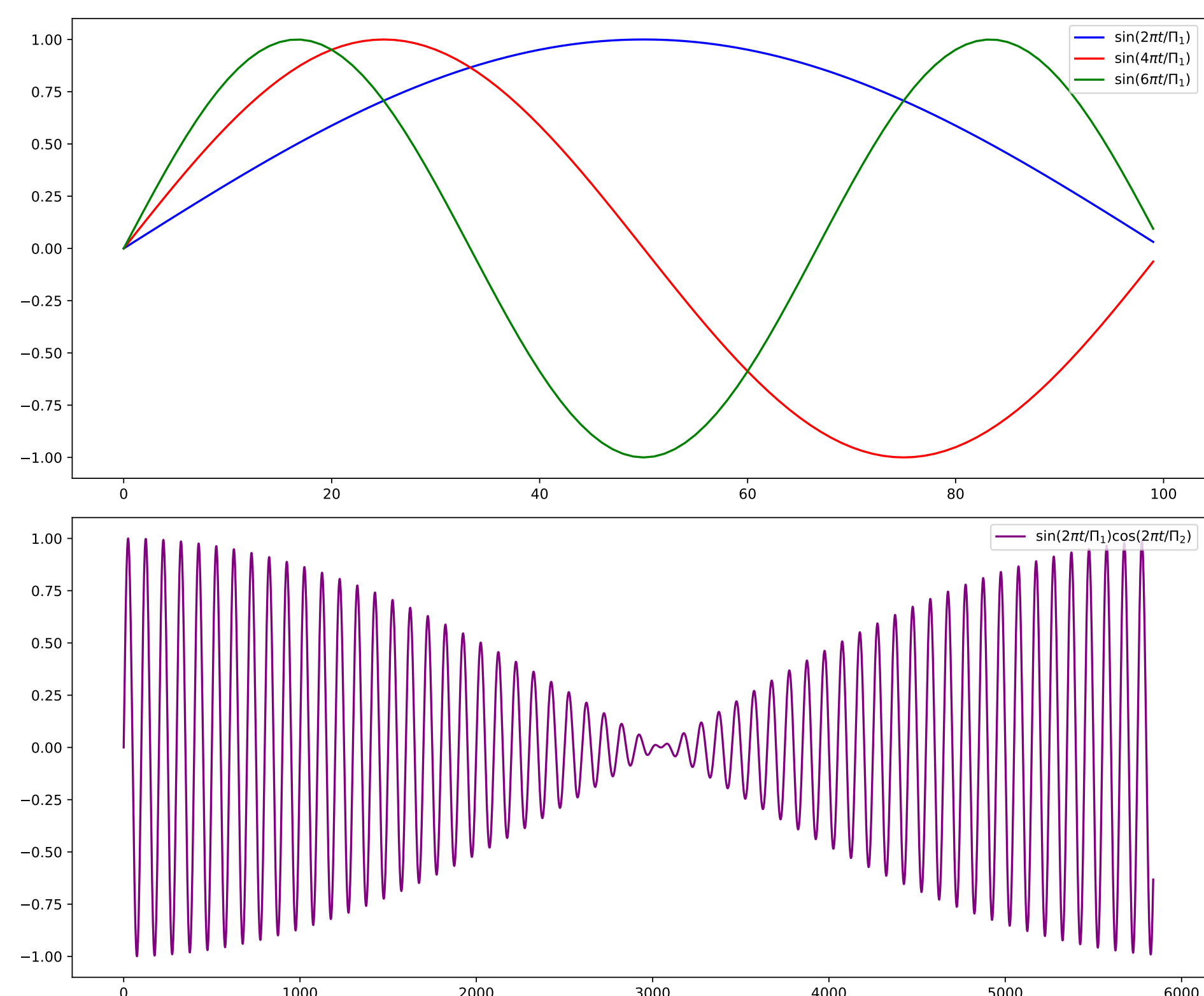


Figure 2. Some basis functions used in the proposed method. *Top* First 3 harmonics of sin with period $\Pi_1 = 100$. *Bottom* First cross terms of sin and cos with periods $\Pi_1 = 100$ and $\Pi_2 = 36522$.

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). The data is collected at 15-minute intervals and there are no missing data.

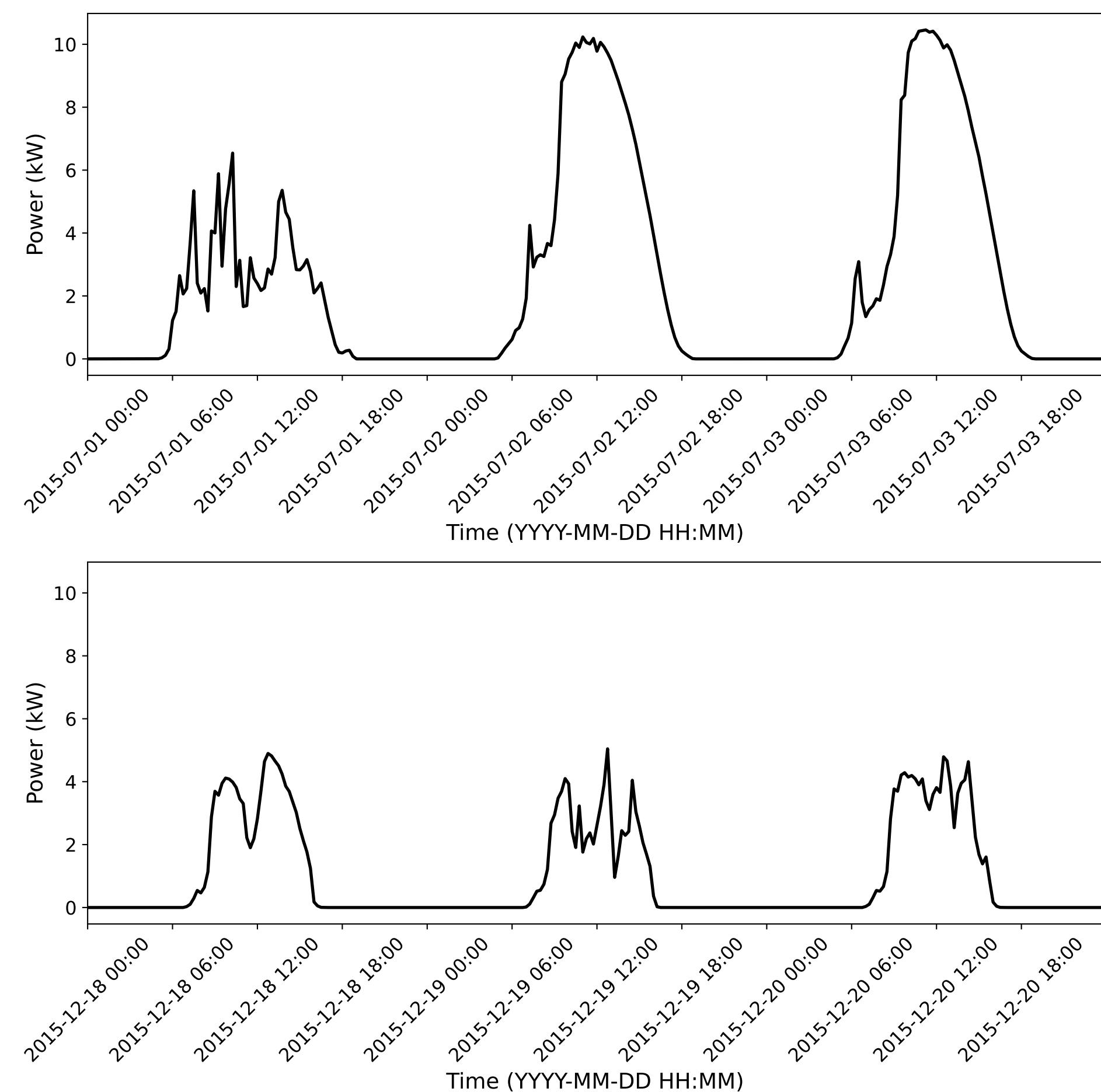


Figure 3. Subdaily PV output during 2 different seasons. *Top*. Summer. *Bottom*. Winter.

Figure 3 shows subdaily data for 3 days in summer and winter. As expected during nighttime power output is zero. During daytime, power output is highly variable due to the effect of the weather but is mostly similar between consecutive days. This behavior indicates a daily periodicity. We also see that there is a drop in power output in the middle of the day for winter days. Upon further inspection, we noticed that this drop was caused by the shading of the PV system by a nearby building.

Smooth multiperiodic sunrise/sunset estimation

We start by estimating sunrise and sunset times for each day in the dataset. We do this by solving the problem (1) with logistic loss

$$\ell(y_t, \hat{y}_t) = \begin{cases} -\hat{y}_t + \log(1 + \exp(-\hat{y}_t)), & y_t \geq \epsilon, \\ \log(1 + \exp(-\hat{y}_t)), & y_t < \epsilon. \end{cases}$$

where ϵ is a small power threshold.

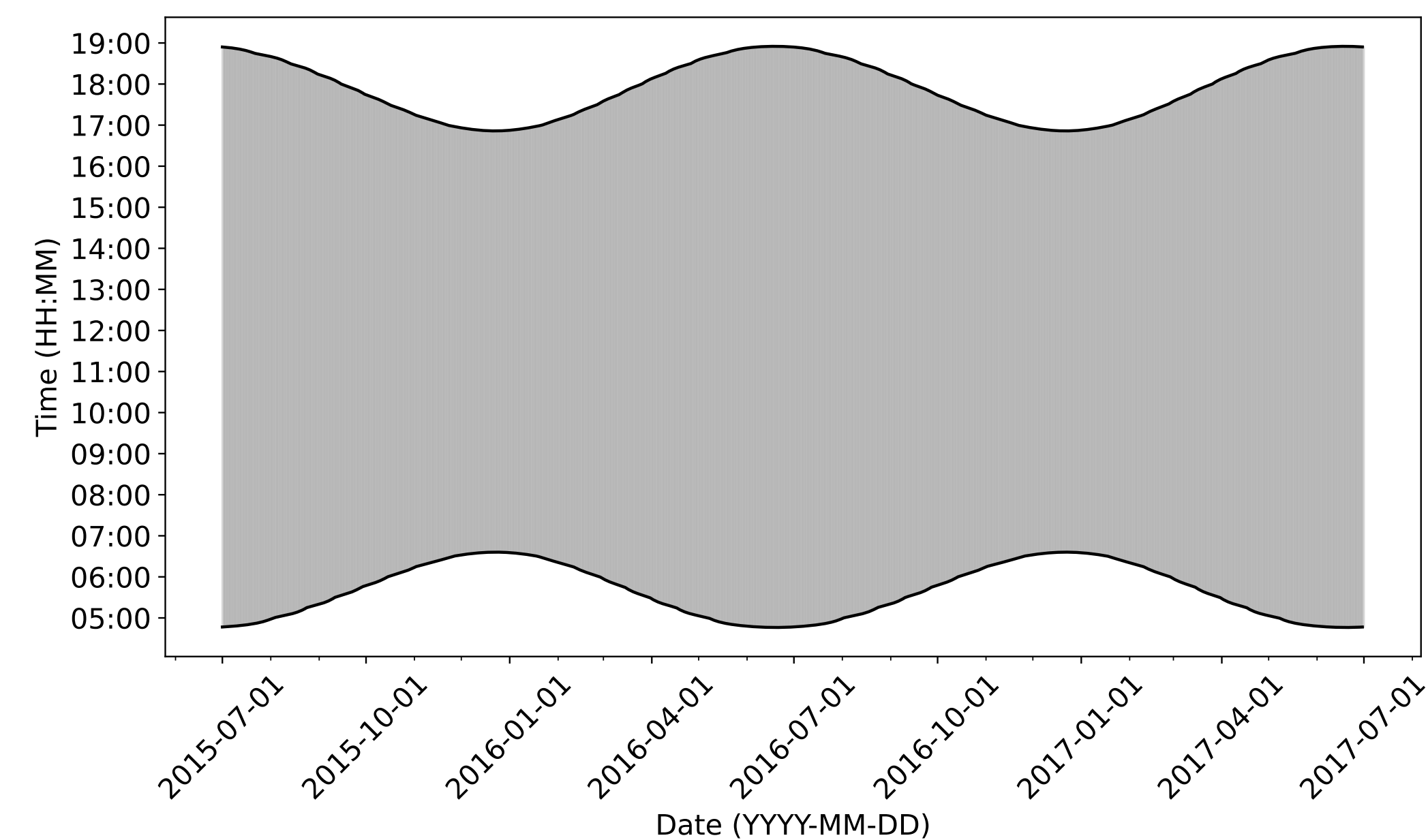


Figure 4. Estimated sunrise and sunset times over entire dataset. Shaded regions indicate daytime for each day.

Figure 4 shows estimated sunrise and sunset times across all days in the dataset. We observe that sunrise and sunset times vary throughout the year and are not symmetric around noon. As expected, sunrise is earlier and sunset is later in summer, resulting in longer daytime in summer than in winter.

Time dilation

After estimating sunrise and sunset times for each day in the dataset, we standardize the length of daytime data using linear interpolation. This process, which is reversible, ensures that each day, regardless of its initial length (with winter days typically shorter than summer days), is adjusted to a consistent, standard length.

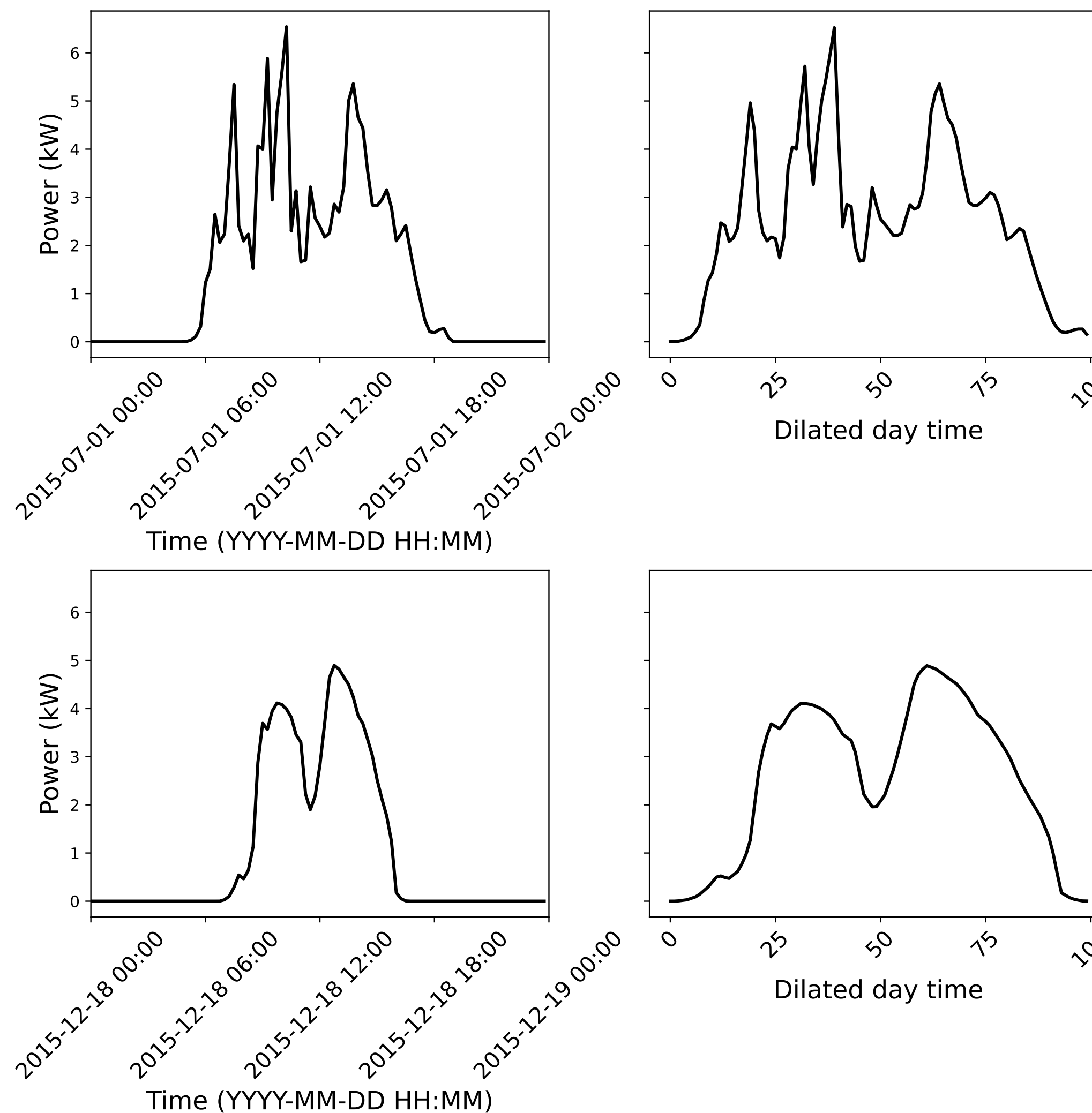


Figure 5. Time dilation of day time PV output. *Top*. Summer. *Bottom*. Winter.

Bundt cake representation

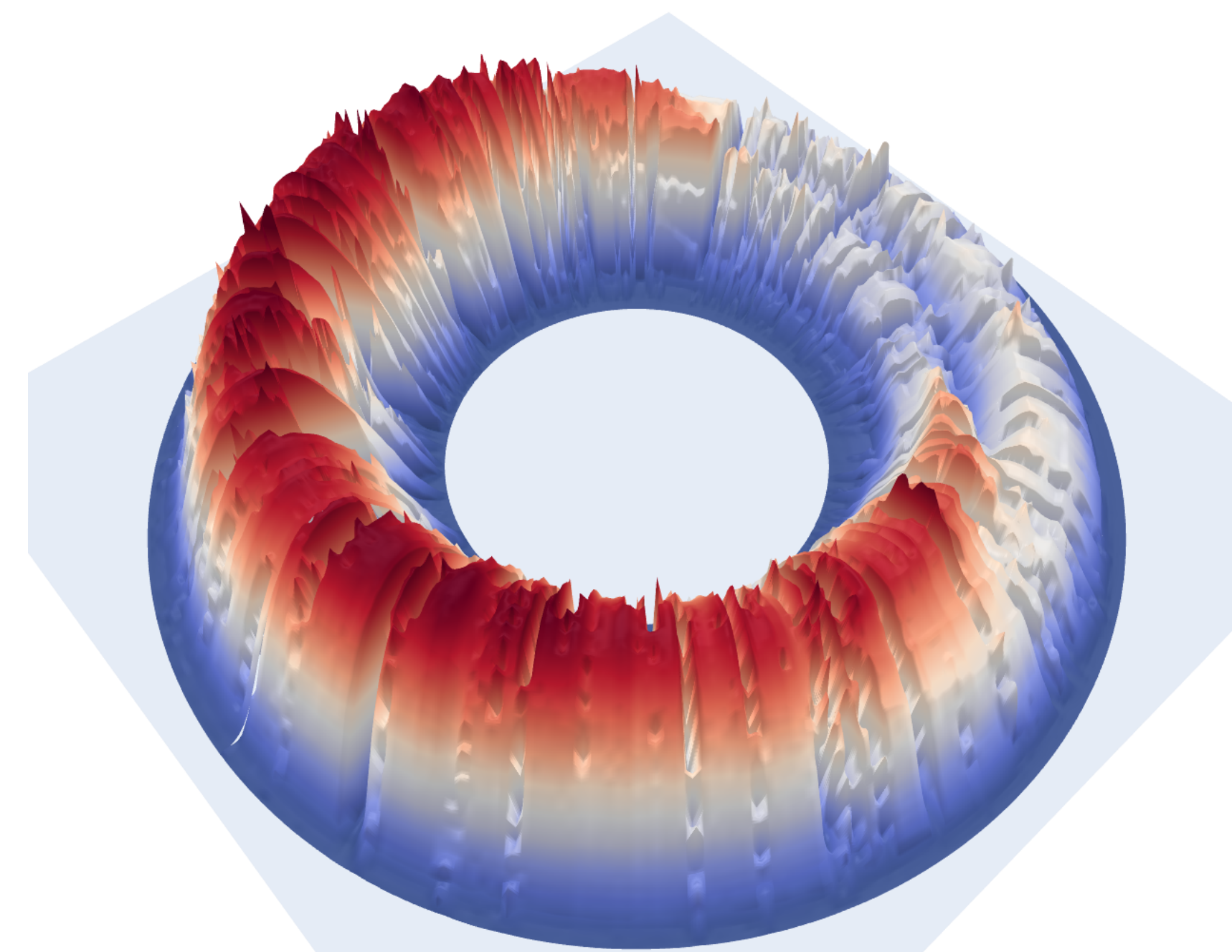


Figure 6. Bundt cake representation of time dilated PV output between 07/01/2015 and 06/30/2016.

Figure 6 shows the Bundt cake representation of dilated daytime data between 07/01/2015 and 06/30/2016. Here, each day is represented as a slice of the Bundt cake and the height of the Bundt cake represents the power output for each timestep. A heatmap is used such that as power output increases, the color changes from blue to red.

Smooth multiperiodic quantile estimation

Next, we estimate smooth multiperiodic quantiles of the data. We do this by solving the problem (1) with ℓ being pinball loss corresponding to the quantile we want to estimate. Namely, for $\eta \in [0, 1]$,

$$\ell(y_t, \hat{y}_t) = \begin{cases} (1 - \eta)(\hat{y}_t - y_t), & y_t \geq \hat{y}_t, \\ \eta(y_t - \hat{y}_t), & y_t < \hat{y}_t. \end{cases}$$

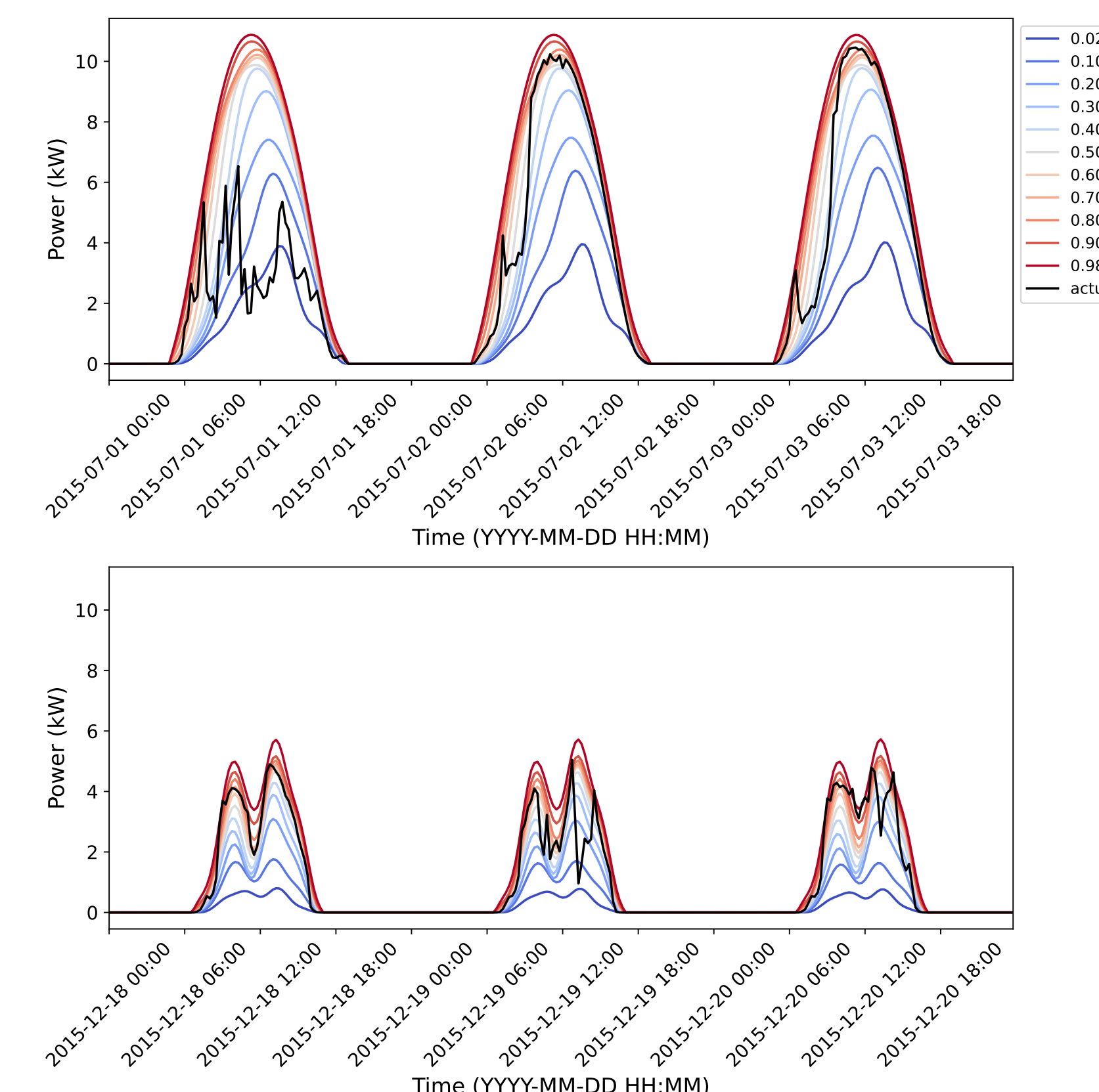


Figure 7. Estimated quantiles shown on undilated PV output. *Top*. Summer. *Bottom*. Winter.

The estimated multiperiodic quantiles, displayed in Figure 7, reveal notable seasonal variations in undilated PV output. Summer quantiles are not only higher but also exhibit greater separation and distinct shape differences compared to winter.

Clear sky detection

Each timestep t is modeled as two states—clear sky ($z_t = 1$) and not clear sky ($z_t = 0$)—depicted as nodes in a graph. State transitions, shown as graph edges, can either persist or change. Identifying clear sky intervals is framed as a shortest path problem, focusing on minimizing total penalties for traversing nodes and edges.

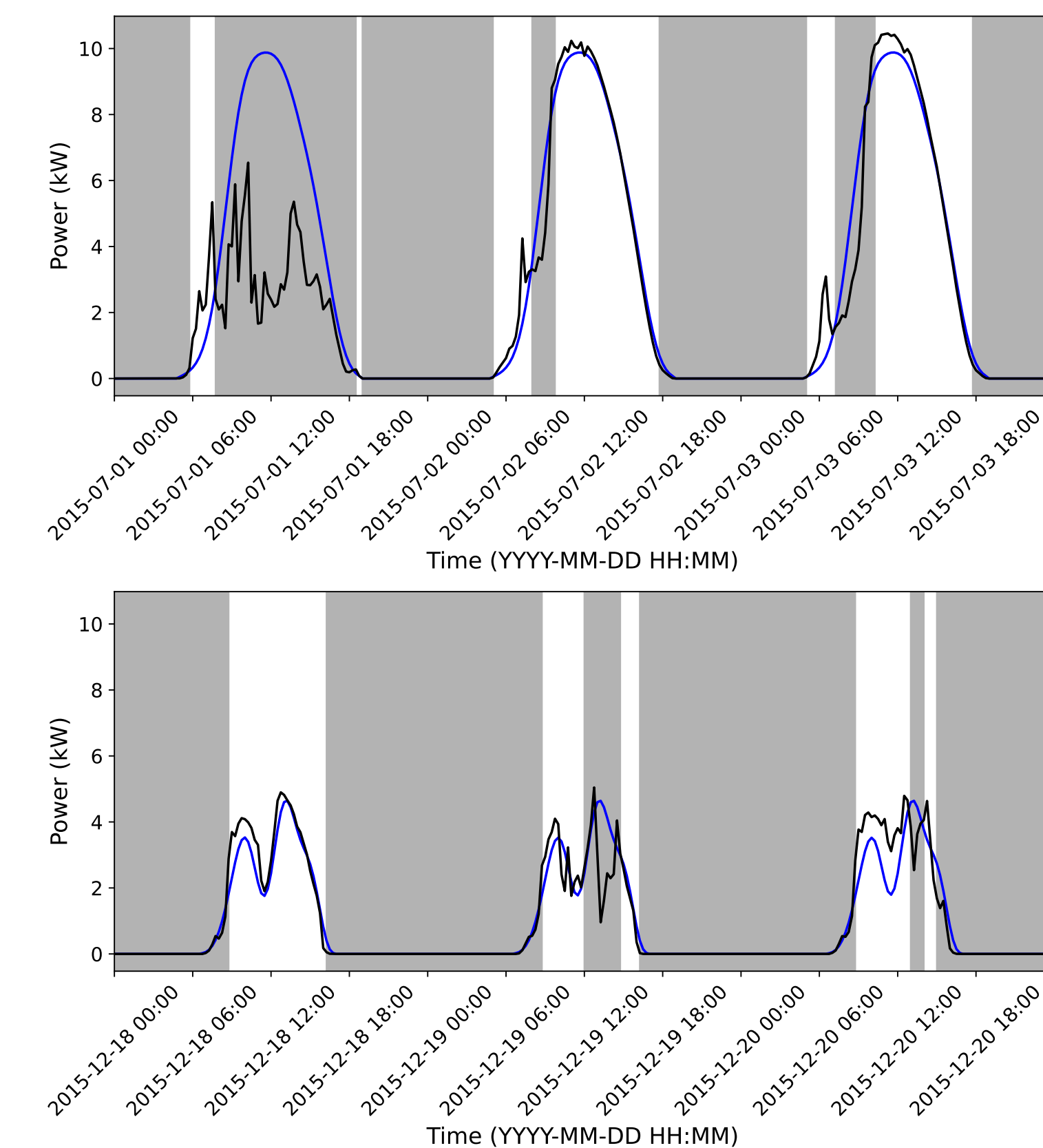


Figure 8. Detected clear sky time intervals. *Top*. Summer. *Bottom*. Winter.

References

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