Data Analysis and Green Computing: Profiling HPC Power and Tracking CO2 Emissions

Green Team Report

Supervisor: Mohsen Seyedkazemi Ardebili

April 2024

Contents

Abstract

In the following document we try to collect all possible observations and intuitions related to our work, whose goal is to profile High Power Computing power consumptions and track CO2 emissions related to it.

Contributors:
Devletiyarov Baibek
Liberatoscioli Martina
Squarzoni Lorenzo
Zavaroni Sofia

1 Tools and methods used for the analysis

1.1 STL (Seasonal-Trend Decomposition)

1.2 Meta's Prophet

2 Power and Energy Analysis

The data distribution spans from April 2020 to October 2022, showing a discrete amount of information that can be enough to make some initial observations and guessings. While there isn't a discernible pattern across the entire period, comparing plots for specific nodes (r205n01), the sum of all nodes, and individual racks reveals certain trends.
The lines in the plots that have a strange or unusual pattern are the result of the horizontal and vertical interpolation applied to fill some empty spaces in the distribution. No further manipulations have been applied to the data.

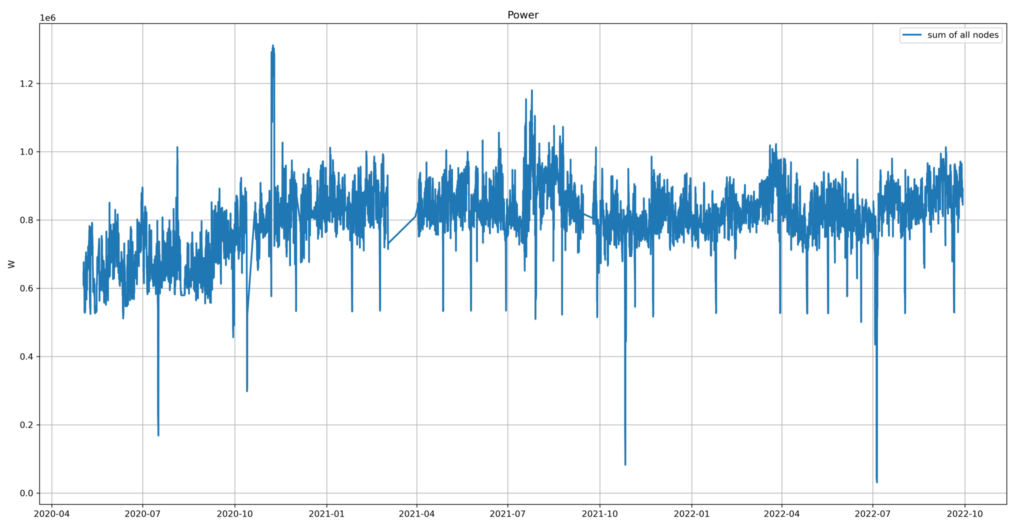

Figure 1: PWR total value (sum of all nodes in the server)

count 2.110200e+04
mean 7.997859e+05
std 1.008300e+05
min 3.101649e+04
25% 7.600741e+05
50% 8.096636e+05
75% 8.581640e+05
max 1.311606e+06

In terms of pure power consumption, we see a visible peak that reaches a value of 1.3 MW, while the general mean stays close to 0.8 MW. The data fluctuates a lot throughout the days, but it is difficult to find any particular pattern or repetition at this level of depth; what we can guess is that all o most of the lowest points in the plot are given by a moment or period of maintenance for the server, while the highest values might indicate an episode of testing for the capabilities of the server in terms of maximum computational power.

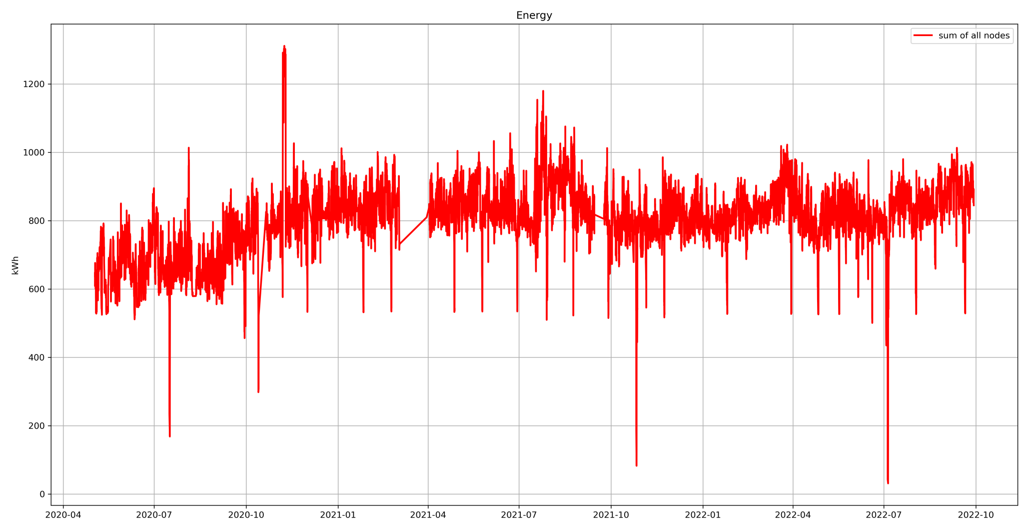

Figure 2: E total value (sum of all nodes in the server)

The energy computation was conducted based on power consumption data to derive energy values in kWh.
Distinct regions in the plot exhibit consistent energy levels around the mean, notably in June and August, alongside regions with negative peaks (e.g., July and October) and others with positive peaks. Understanding these peaks in the context of high-power computing offers valuable insights;
for example, cooling needs vary with external conditions like temperature. Peaks may coincide with hotter periods, necessitating more energy for cooling.
Also, changes in online service or data processing demand can influence energy consumption, e.g., heightened activity leading to increased power usage.

2.1 PWR r205

At rack or even node level we see a similar behaviour: a lot of peaks rising from an horizontal line that indicates the mean power consumption.

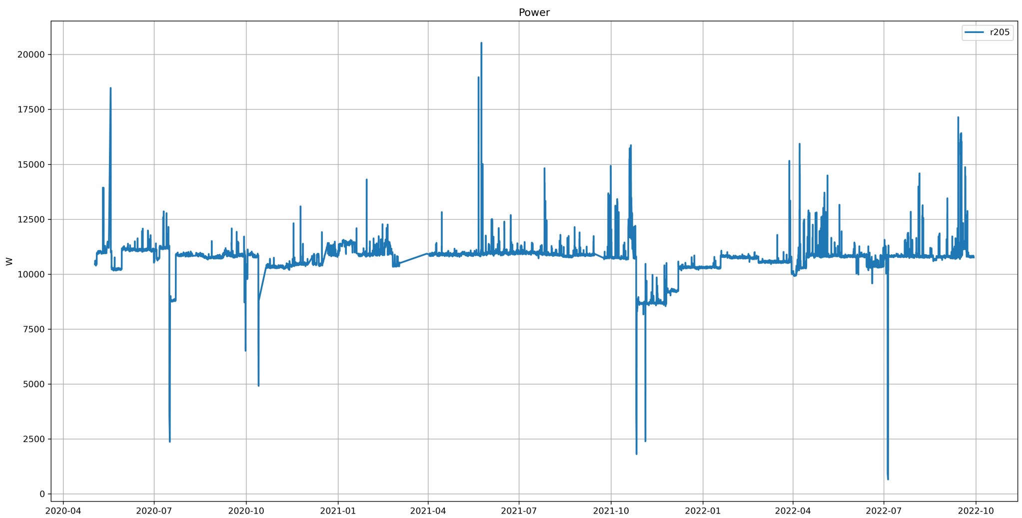

Figure 3: PWR r205

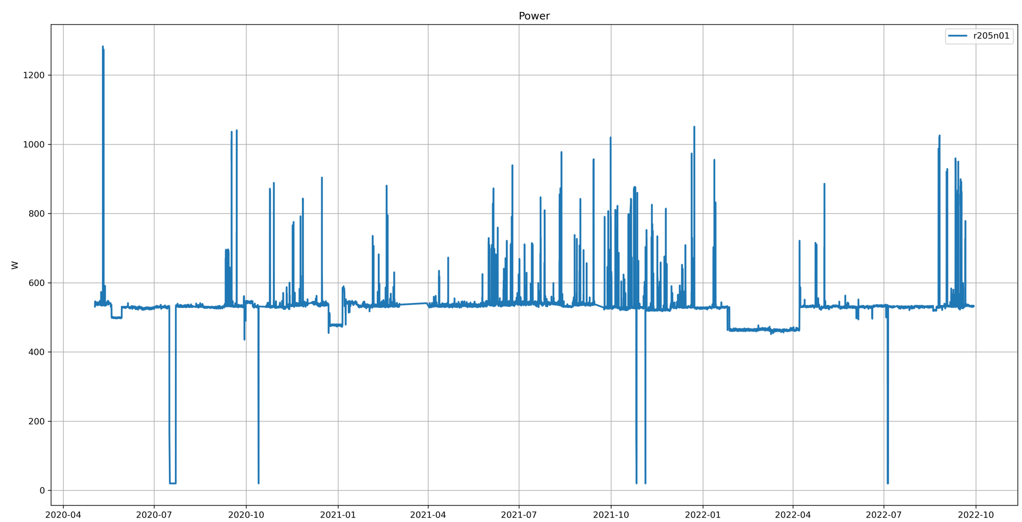

Figure 4: PWR r205n01

2.2 PWR r206

The first rack (r205) seems to be the only one with such a regular and stable power consumption; indeed just looking at a different rack like this one (but every other rack is more similar to this) we see a much different and oscillating plot. The only hypothesis we can make is that the first rack is much less used that all the others, or maybe it is used for a different purpose.

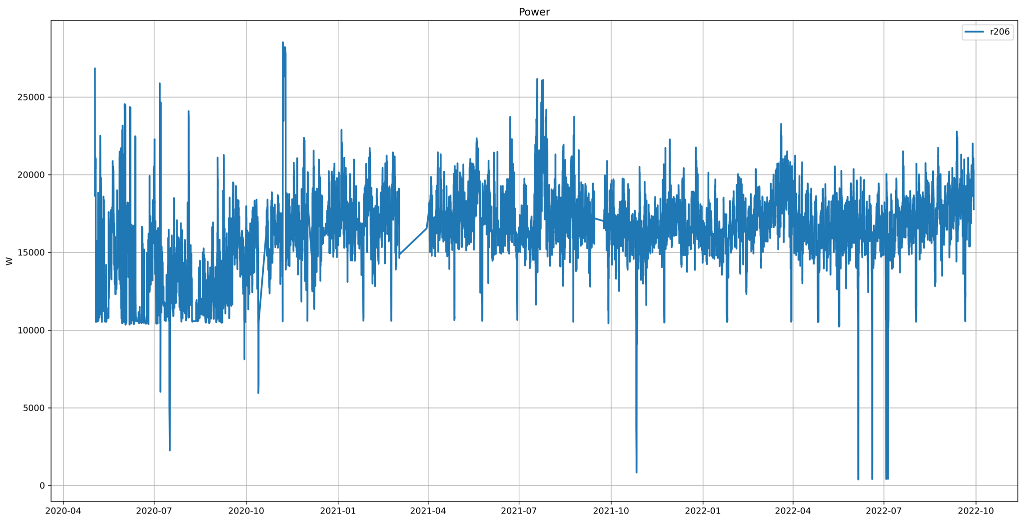

Figure 5: PWR r206

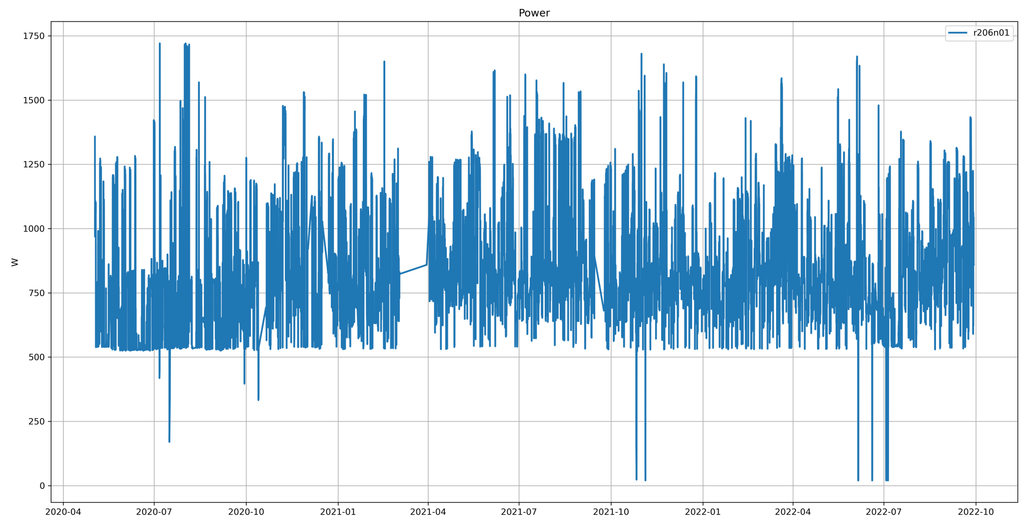

Figure 6: PWR r206n01

2.3 PWR r206n01 STL

Even making a seasonal-trend decomposition it's hard to highlight any specific trend, since the data is full of variations and outliers. Taking advantage of a different tool we'll try to make any seasonality clearer.

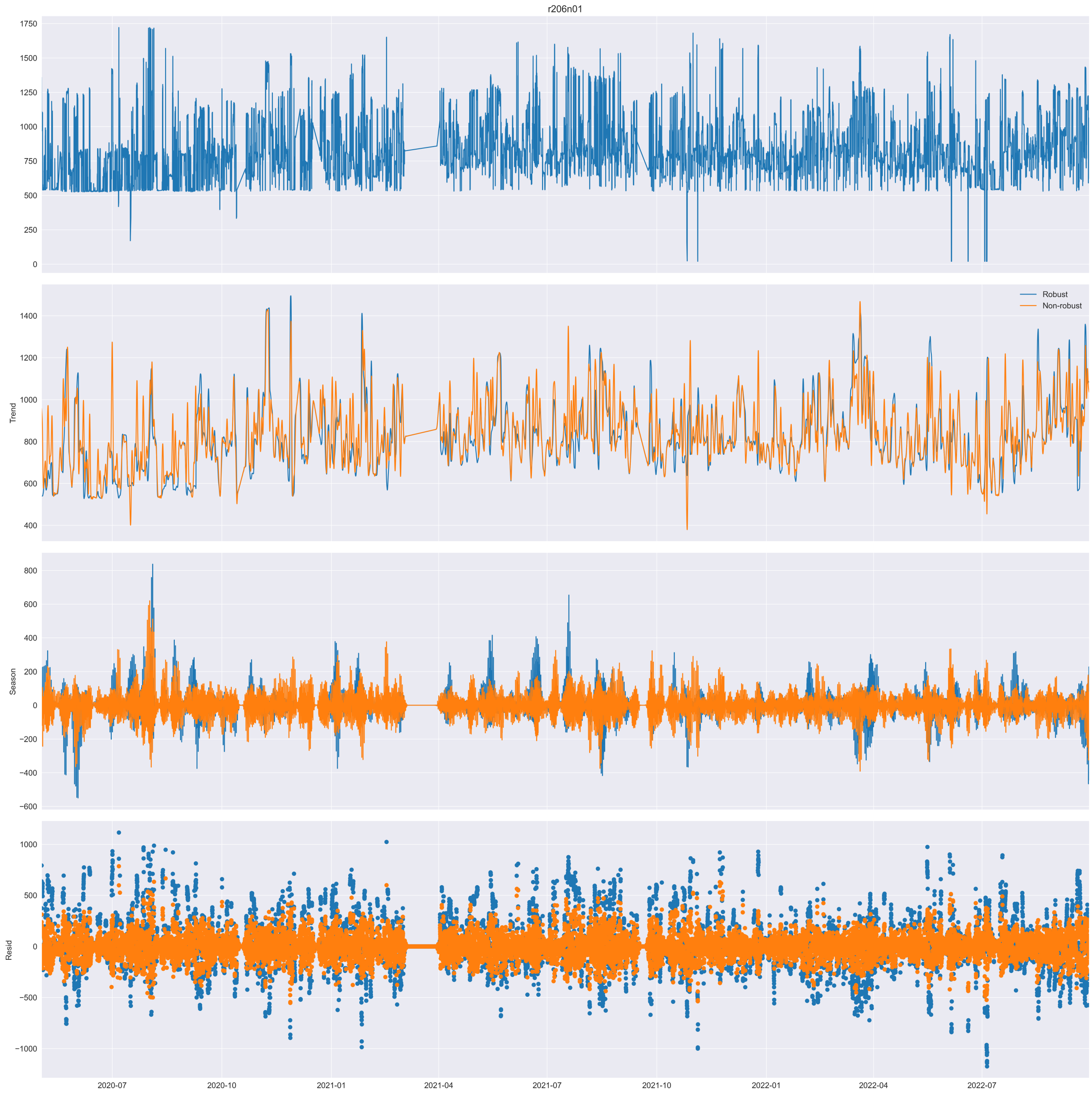

Figure 7: PWR r206n01 STL

2.4 PWR analysis using Meta's Prophet

Thanks to the usage of this method we are able to extract a clearer representation of the possible trends of power consumption in our server: good choices, purely looking at the following data, might be to increase the nodes' usage during the night, during week days and also during hot seasons, all periods in which power consumption results to be lower.

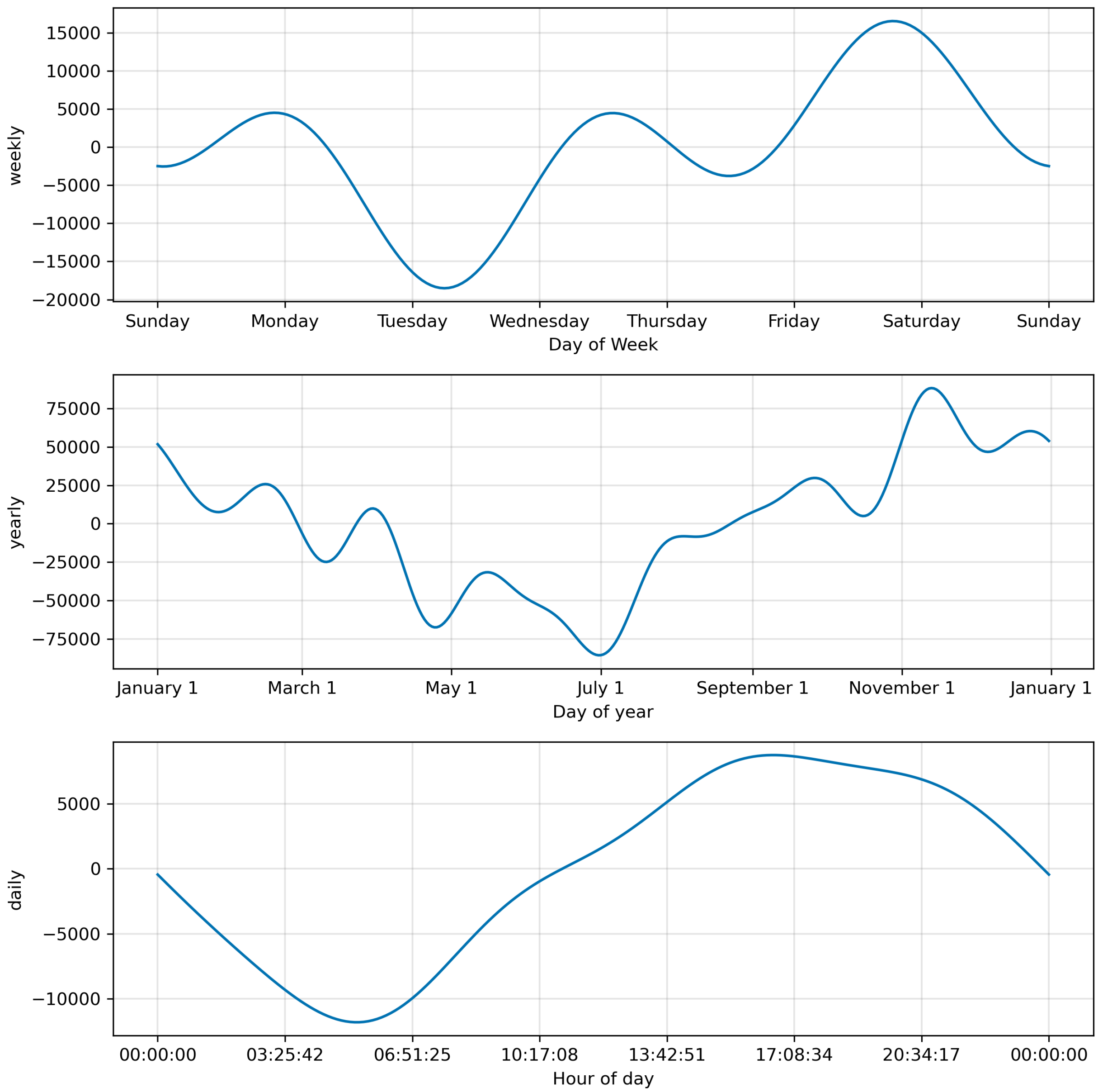

Figure 8: PWR trends

Analyzing these trends aids in optimizing energy management strategies and mitigating environmental impact in high-power computing environments. For example implementing energy optimization algorithms could reduce consumption during certain periods, resulting in negative peaks; while aiming to a better server energy efficiency can lead to improvements in server and cooling efficiency and alter the graph's shape over time, potentially reducing energy consumption peaks.

3 Carbon Intensity Analysis

The data distribution goes from 2021-01 to 2024-01.
In this case data hasn't been manipulated in any way, as no interpolation has been applied. With carbon intensity is much easier to highlight patterns and trends; in this case we are considering the carbon intensity related to direct power and energy usage, specific for our country (Italy) and area (Emilia-Romagna).

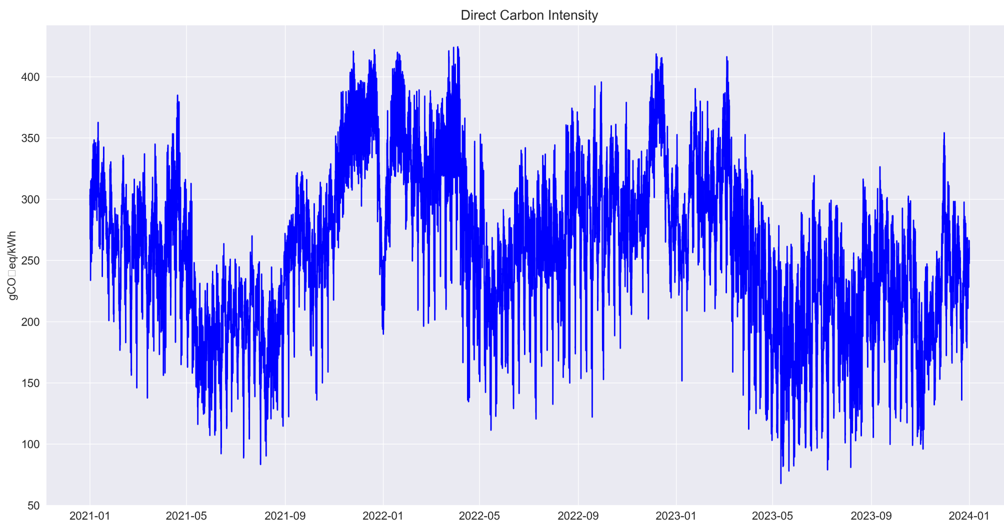

Figure 9: CI direct

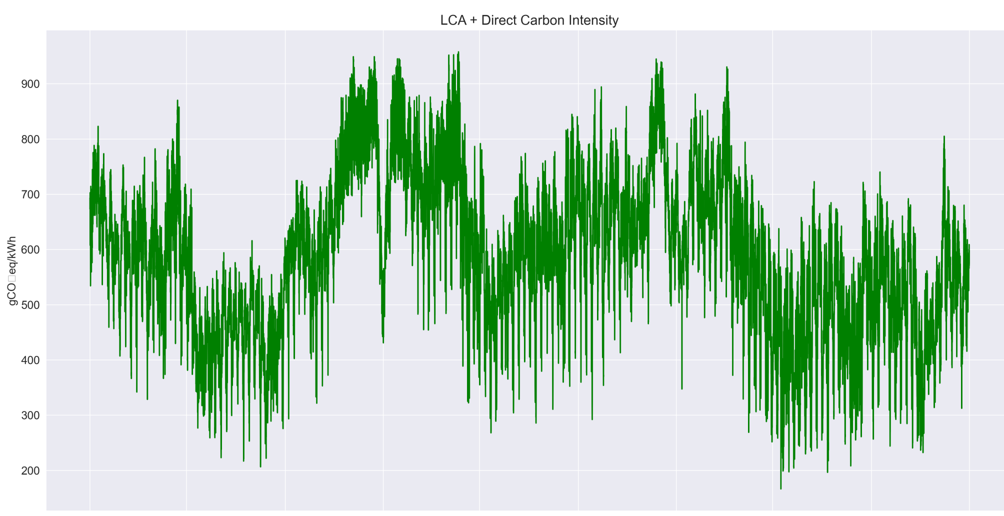

Figure 10: CI LCA + direct

count 26280.000000
mean 258.794563
std 63.450587
min 67.790000
25% 214.945000
50% 260.075000
75% 302.832500
max 424.440000

3.1 CI comparison

Both direct and indirect (LCA) carbon intensity share a similar pattern, showing equal highs and lows in time. Also the sum of the two distributions has the same behaviour.

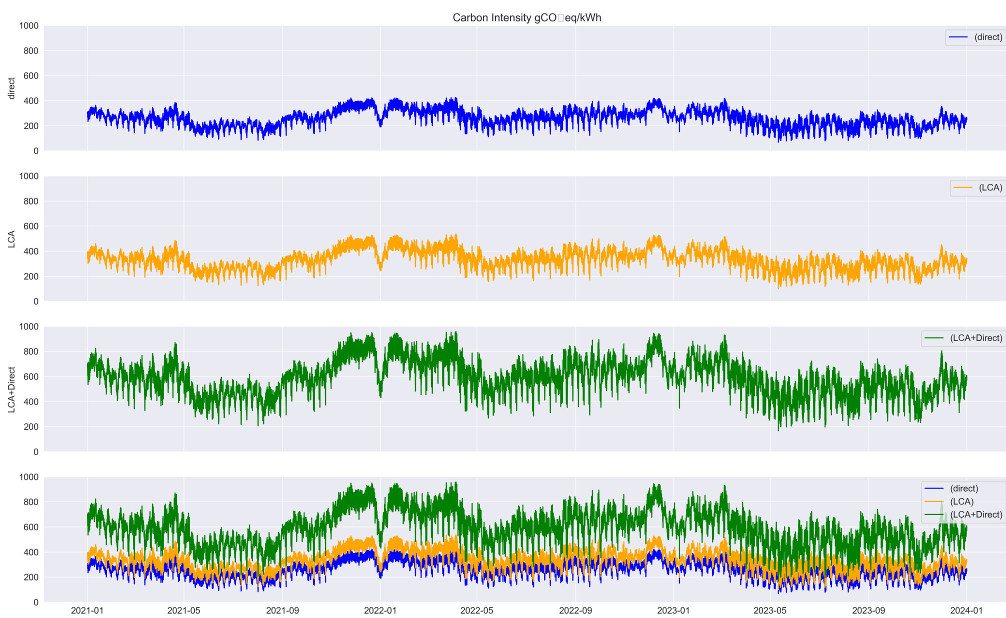

Figure 11: CI comparison

3.2 CI STL

Taking advantage of this decomposition we can easily guess the following assumptions: carbon intensity has a strong seasonality, showing a visible decrease during the warmer months and the opposite during cold ones; in addition there seems to be a strong decrease in the CI every January.

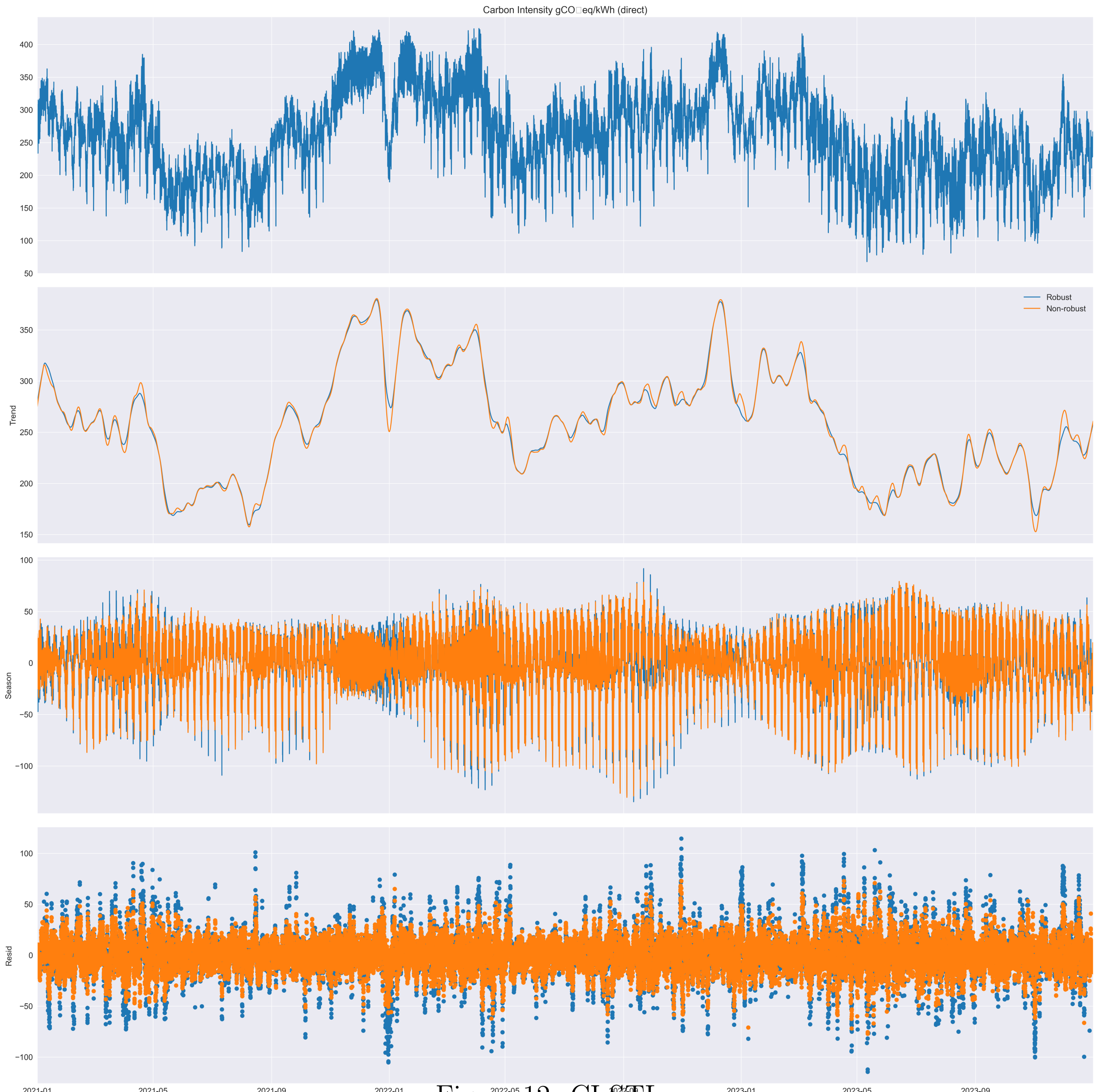

Figure 12: CI STL

3.3 CI analysis using Meta's Prophet

At deeper level (daily, weekly or monthly), much clearer patterns are visible:
During week days the carbon intensity is higher with respect to weekends, as well as the CI contribution decreases during the lunch time (noon). Finally we can better visualise what we guessed before: during the year we see a couple of peaks, near February and December,

with a lowering during January, contrary to what happens from May to September, where the general CI contribution goes down.

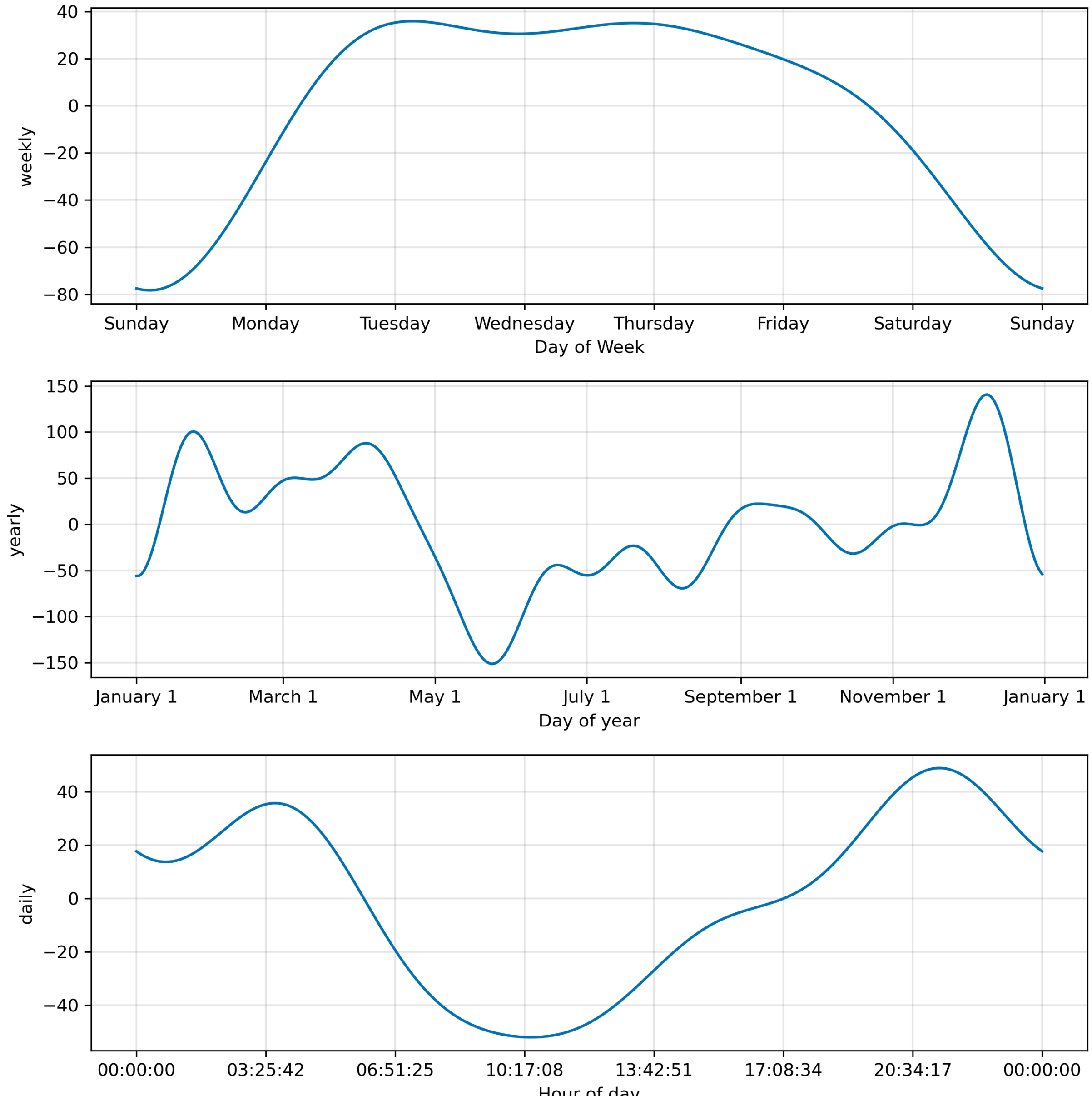

Figure 13: CI trends

4 Operational Carbon Footprint Analysis

Operation Carbon Footprint is the most important metric for the goal of our observation, since it is the result of combining energy usage (related to power consumption) and carbon intensity.
For this reason we will see a combination of the observations we did related both to power and CI.

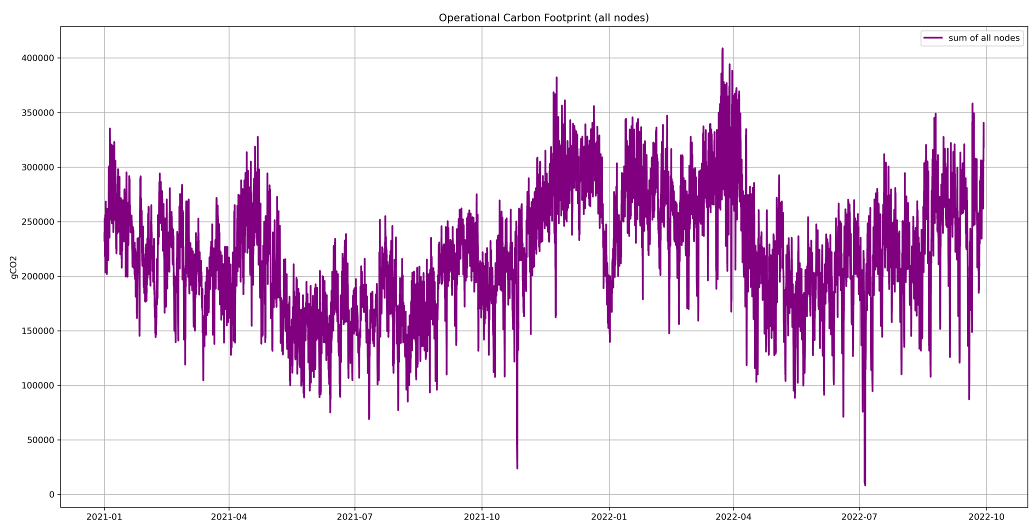

Figure 14: Cop total value (sum of all nodes in the server)

count 15263.000000
mean 220186.340067
std 54400.325364
min 8292.258962
25% 181522.725893
50% 218140.616056
75% 258846.427357
max 408868.699240

4.1 Cop r206n01 STL

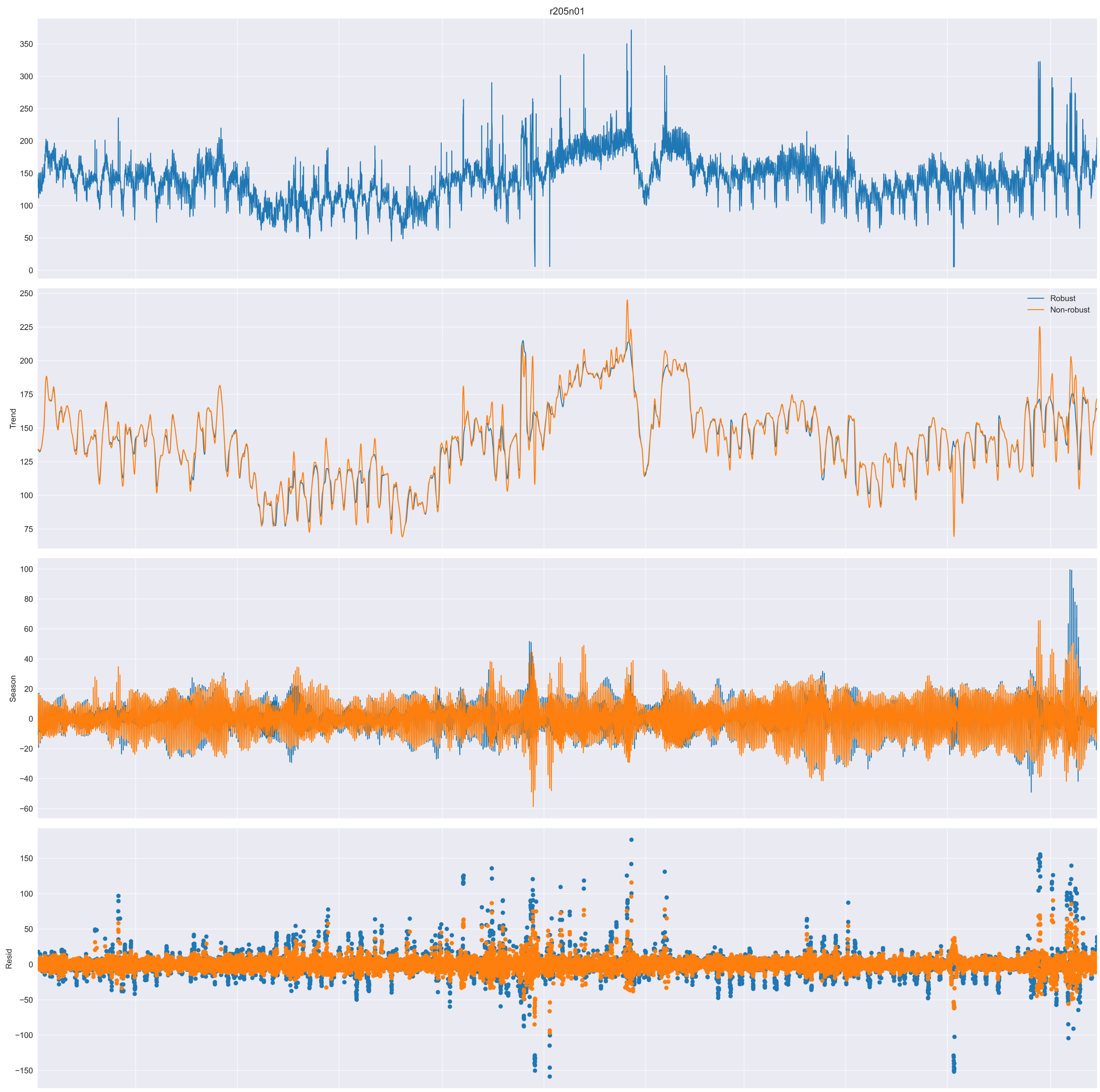

Figure 15: Cop r206n01 STL

4.2 Cop analysis using Meta's Prophet

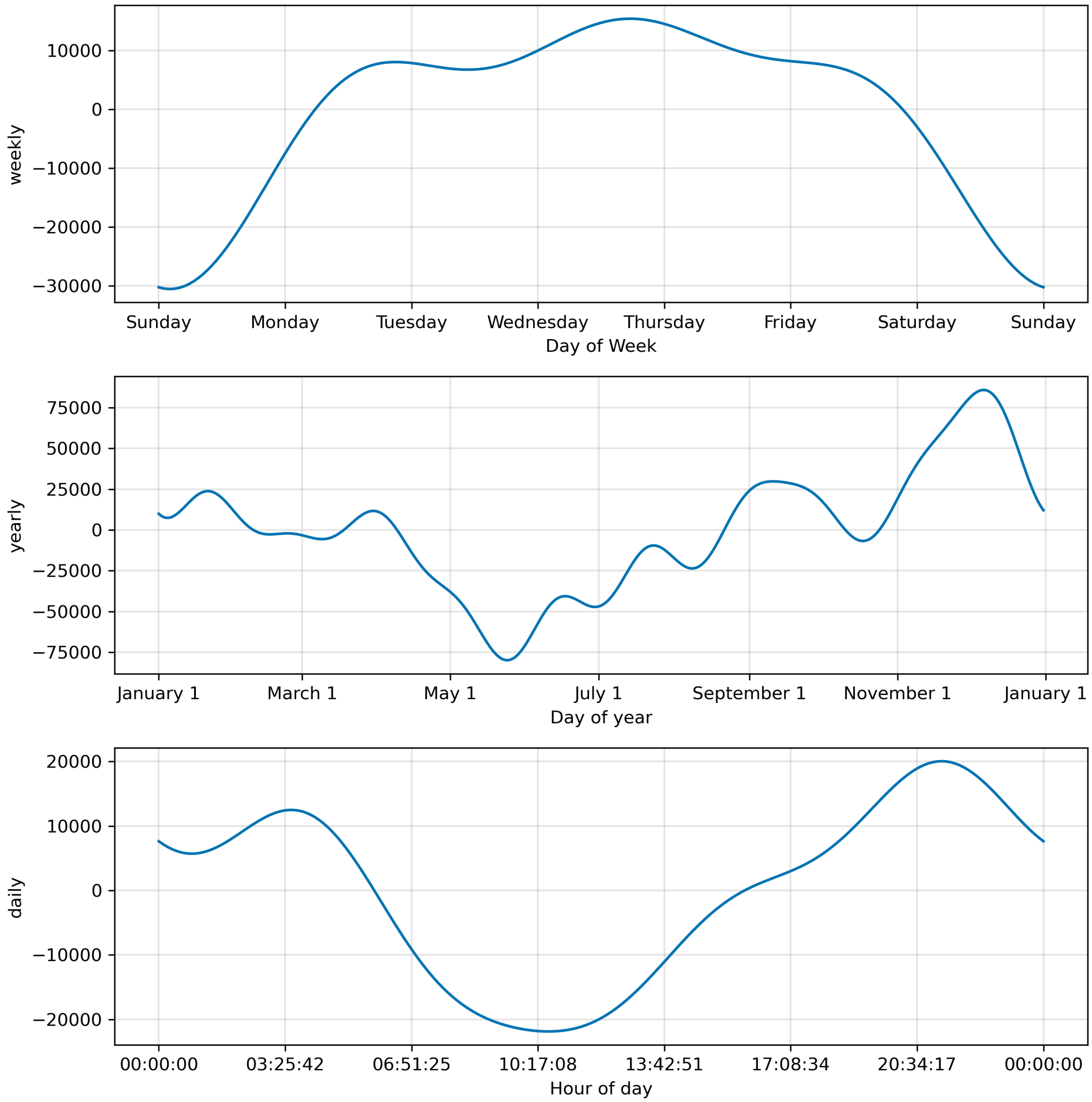

Figure 16: Cop trends