A Time Series Approach to Modeling Potential Solar Output

Nathaniel Reimer, Kyle Suelflow

2023-12-18

Replacing carbon-emitting energy sources with greener alternatives is critical to preventing worst-case climate change scenarios. Solar power is an incredible option. To make the best use of solar power, we need to have a way to model potential solar output over long and short periods. We use time-series methods to model Global Horizontal Irradiance (GHI) at the hourly and daily levels. Using a sine function to model hourly GHI, we find a correlation between data taken less than four hours apart. After removing multiple seasonal relationships in the aggregated daily data, we find a correlation between data two to three days apart. We also show that each location we modeled required unique considerations and approaches.

Introduction

Greenhouse gas emissions have exploded in recent decades, causing substantial damage to the livelihoods of many across the world. One of the main culprits of these emissions is the energy sector. Using non-renewables such as Coal, Oil, and natural gas to create electricity causes carbon and other GHGs to be emmitted. To combat this, renewable energy has been brought forward as the leading solution to this crisis. Solar and wind energy, specifically, have the most potential to create lasting change in regards to our energy usage. The reality of this situation has brought us to think about modeling different solar variables. Being able to model Global Horizontal Irradiance, for example, allows places around the globe to gain a better understanding of whether or not they would be suitable for solar energy. Qualitative measures such as economic feasibility, the physical geography of places, and political and regulatory hurdles are other things that places can take into account in order to determine if solar is feasible for them, in addition to considering this analysis. This leads us into our two research questions: How well can we model Global Horizontal Irradiance using a time series approach for different situations across the world? And, Are the models we create consistent across the 25 year period (1998-2023) for which we have data from?

(Crook et al. 2011; Wild, Folini, and Henschel 2017; Boland 2020; Shahsavari and Akbari 2018)

Data Context

This project focuses on Global Horizontal Irradiance (GHI), a measure of the power of the power of sunlight that reaches the horizontal ground (Stein, Hansen, and Reno 2012). GHI can then be used to estimate the output of a hypothetical solar installation placed on the ground. GHI data used in this report is part of a publicly available GHI data set from SolarAnywhere. SolarAnywhere uses data from geosynchronous satellites to estimate GHI rather than measuring it directly with ground stations ("Historical Data Validation" 2023). Their model is calibrated with data from ground stations. Annual data is accurate within $\pm$ 4.2% based on this validation. Errors increase for shorter time periods: monthly data has an RMSE of 3.9% and hourly data has an RMSE of 20.6%.

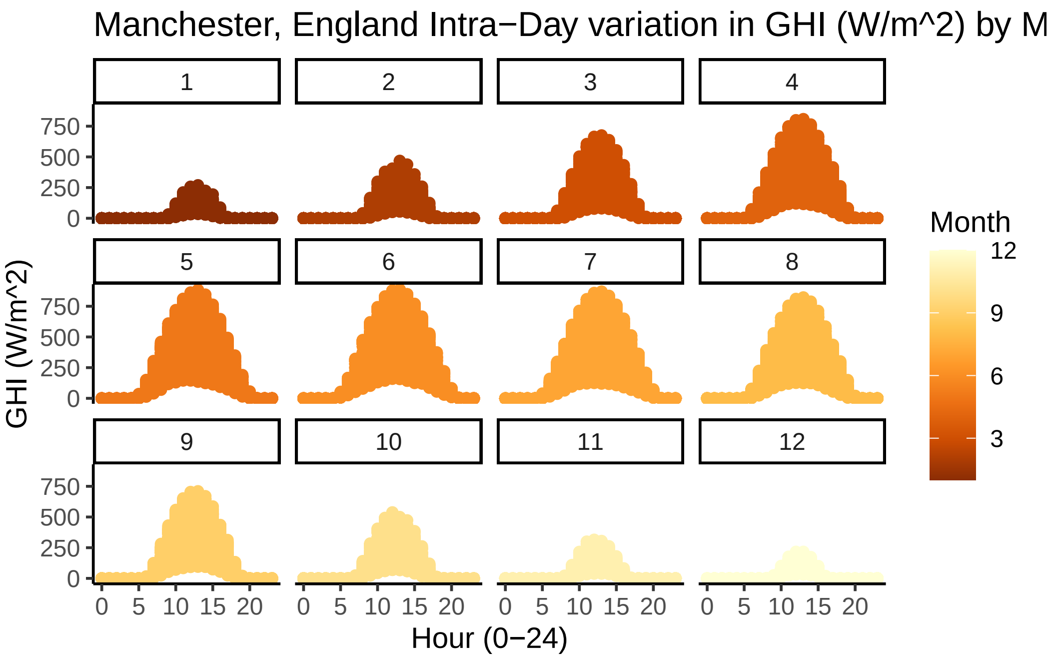
We focus our analysis on Victoria, Seychelles, with comparisons to Nairobi, Kenya, and Manchester, England. Nairobi and Victoria are at similar latitudes and so receive a similar amount of sun throughout the year but have different climates. Manchester's sunlight is more seasonal because of its latitude and it has a very different climate.

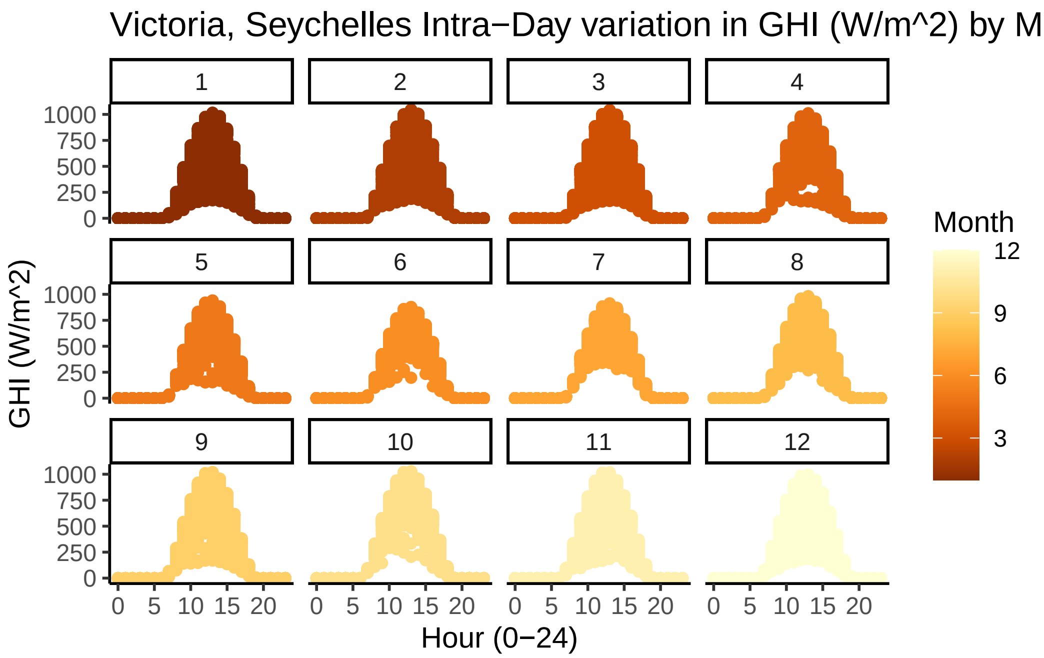
We use daily Solar Insolation data in our analysis of daily GHI. Solar Insolation in our data is the amount of downward solar energy incident at the top of the atmosphere (Hartmann 2016). It varies by latitude and time of year. We generate daily insolation data for each of our selected locations using the Python package climlab (Rose et al. 2022). The insolation data has units of $\frac{W}{m^2}$ per hour. Hourly GHI data has also has units of $\frac{W}{m^2}$ per hour. The difference is that GHI is measured on the ground and solar insolation is measured at the top of the atmosphere.

Model Discussion and Justification

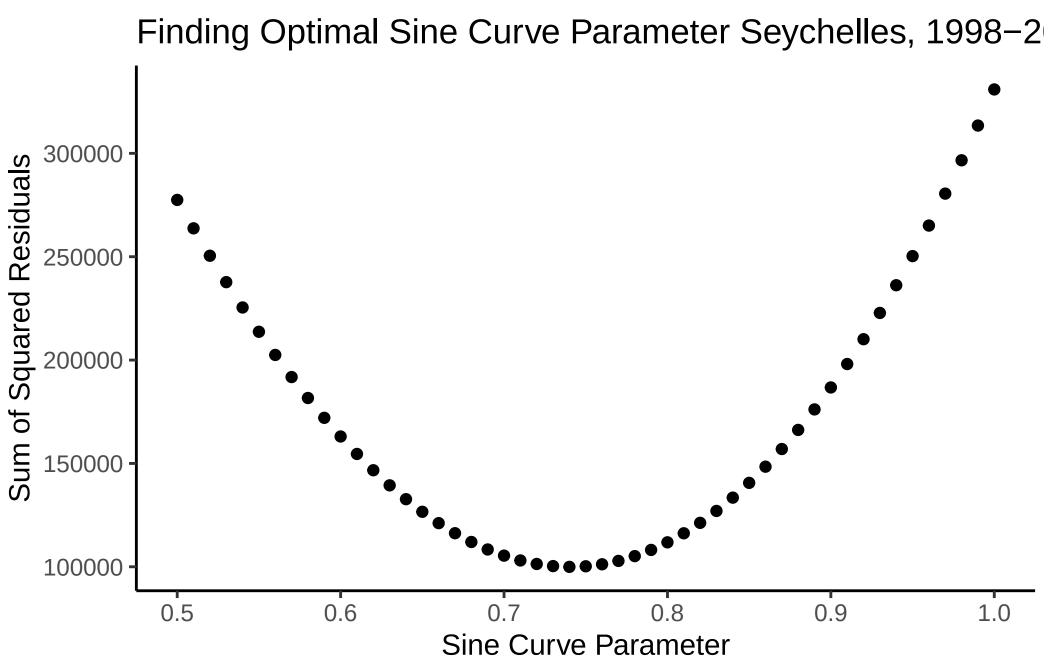
To begin, we wanted to model the intra-day variation in GHI across our 3 sites, and across the 25 year period. An overview of the process is as follows: First, we found the best fitting sine curve to model the average GHI by month for a given 5 year period. Next, we extrapolated this curve onto the plot of ALL data points (no longer mean GHI, but GHI). Then, we used a GEE model to model the residuals of our sine curve model, which determined what order AR model to use. We used 5 different 5 year periods in order to see if there was any changes in our model across time. For the purposes of this explanation, we will show what the process looked like for the years 1998-2002 across each of our 3 locations. For the first step of our process, we decided to model GHI throughout a day by using a sine curve. We made this choice based on striking similarities between the shape of our distribution and that of a sine curve. The equation for the sine curve we chose is listed below: $curveFit = \sin(\pi(time - P)/24)$ `time` is an integer measured from 0 to 24, representing an hour of a given day. `P` is some parameter used to phase shift the sine curve left or right. This equation is consistent across all three locations except for `P`. We put the results of `curveFit` into the linear model below: $MeanGHI = curveFit * Night * factor(Month)$

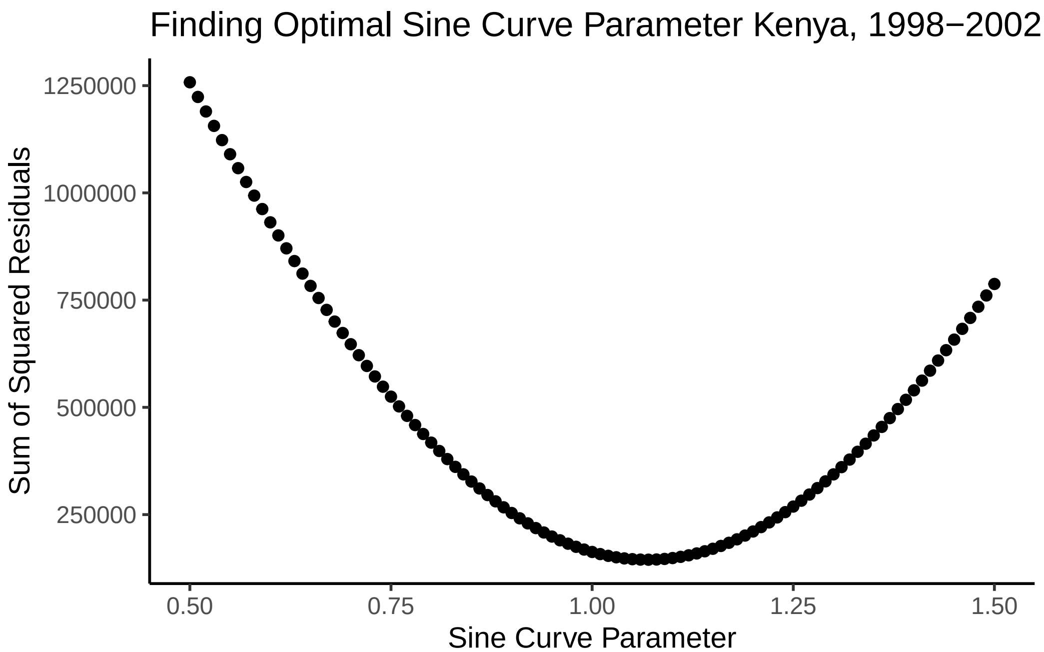
To model MeanGHI, we used 3 variables. `Month` is a categorical variable for the month of the year. `Night` is an indicator variable that says whether or not it is nighttime, i.e. the GHI is zero. `Night` is coded differently for each location. For the Seychelles, since it's latitude is only 4 degrees south of the equator, it did not have much variation in terms of the length of daylight during a day. As such, `Night` could be thought of as between 8 in the morning and 6 at night. For Nairobi, since it's latitude is very similar to Seychelles', we treated `Night` the same. But, for Manchester, as it's latitude is 53 degrees north of the equator, the length of daytime throughout the year changes drastically. Thus, we changed how we coded `Night`. For Manchester, if the minimimum GHI for a given month at a given hour is 0, meaning that atleast once during a month was the GHI zero, then we treated that hour as Nighttime. You can see the variation, or lack thereof, in GHI during a day for each month below. Nairobia is pretty similar to Seychelles, and thus is not shown.

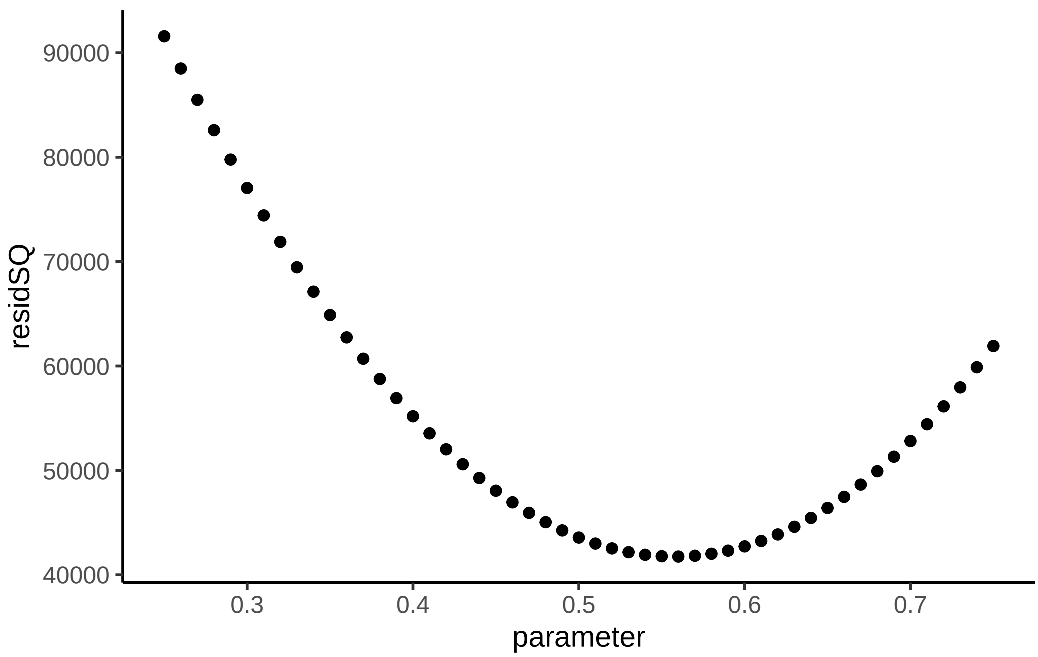




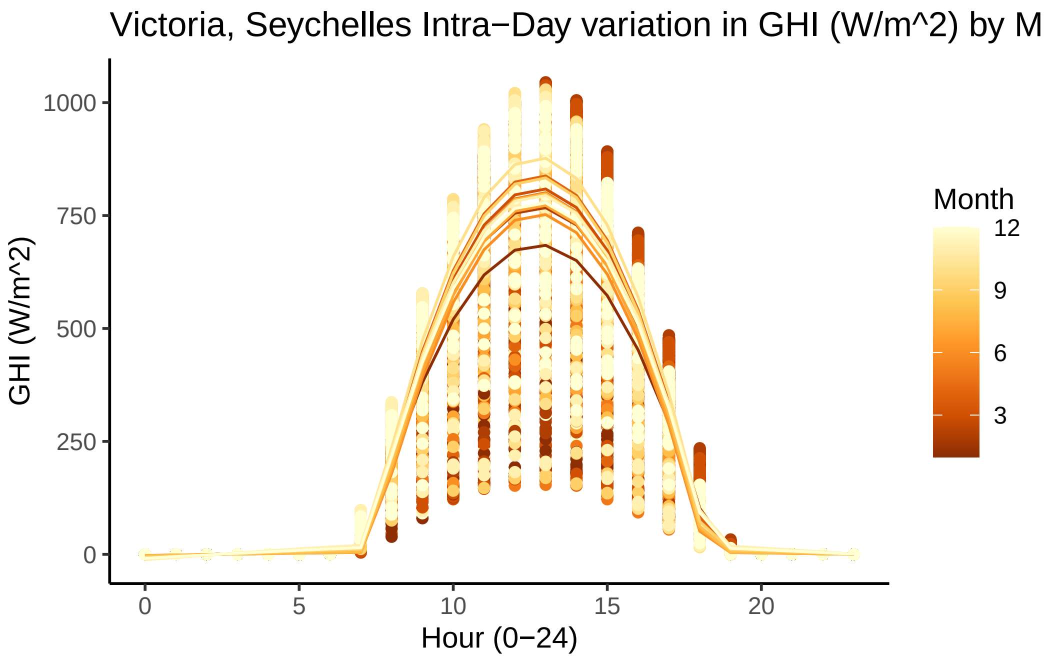
`curveFit` is the sine curve that we fit above. It has some parameter to shift the phase left or right. To determine what this parameter is, we found the optimal parameter based on the minimum of the sum of the squared residuals.



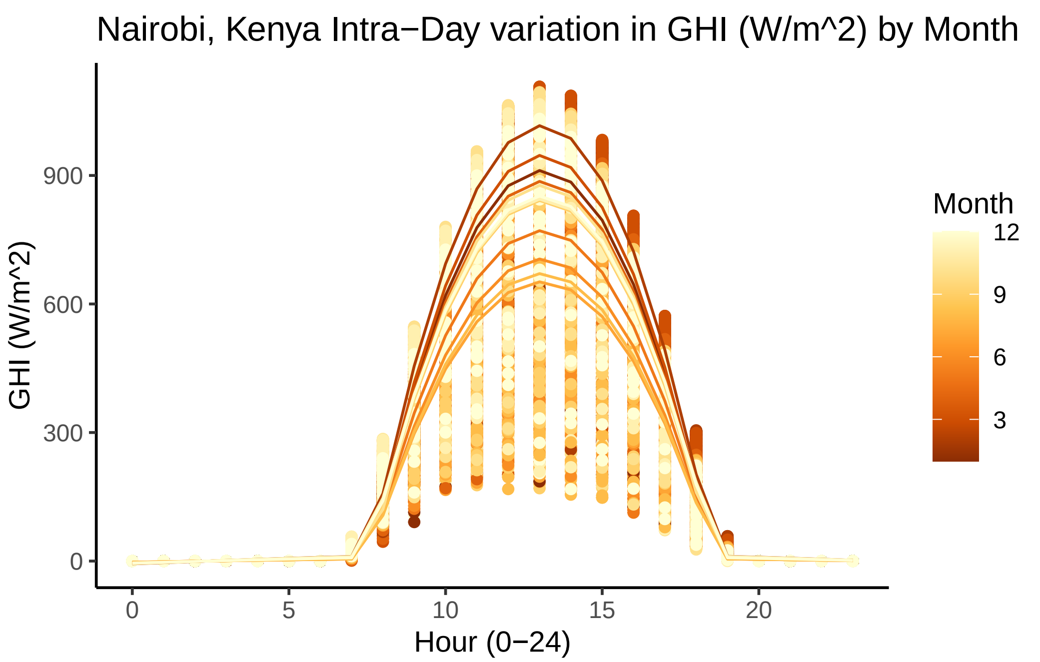




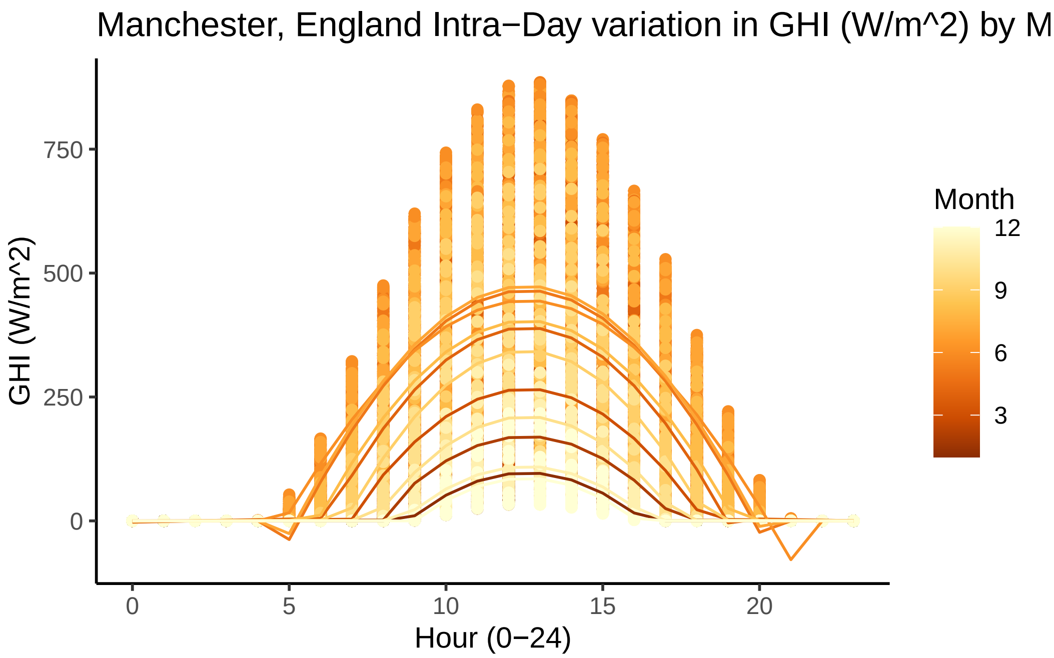
Based on these scatterplots, we chose a parameter of .74, 1.07, and .56 for Victoria, Nairobi, and Manchester, respectively. Now that we have an optimized sine curve model, we fit it over a scatterplot of GHI against Time.



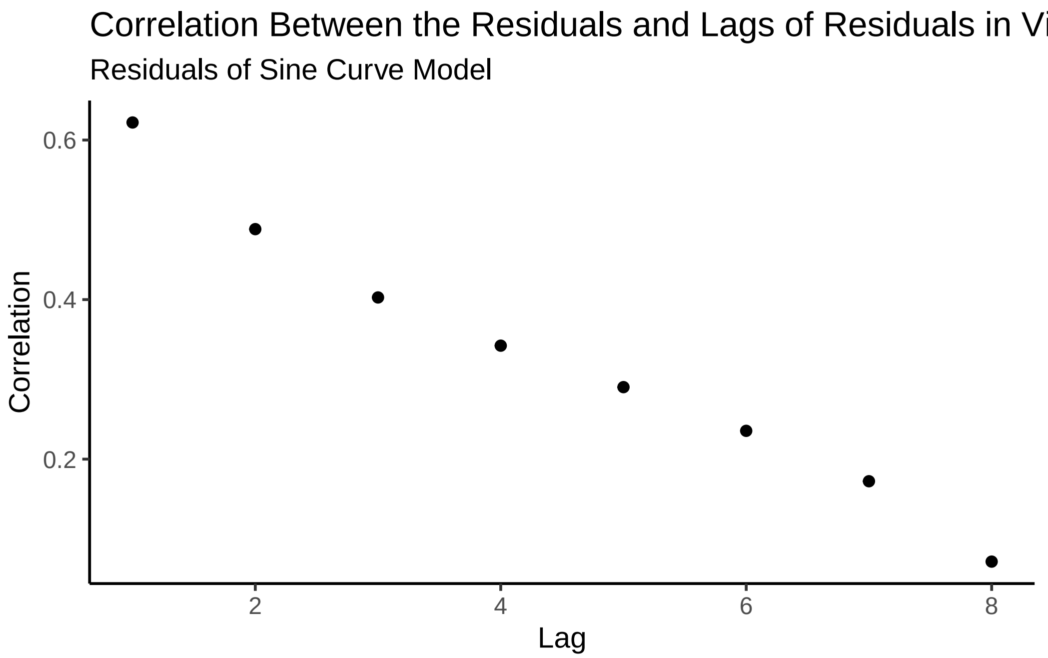
Victoria, Seychelles Intra−Day variation in GHI (W/m^2) by Mo
GHI (W/m^2)
1000
750
500
250
0
0
5
10
15
20
Hour (0−24)
Month
12
9
6
3

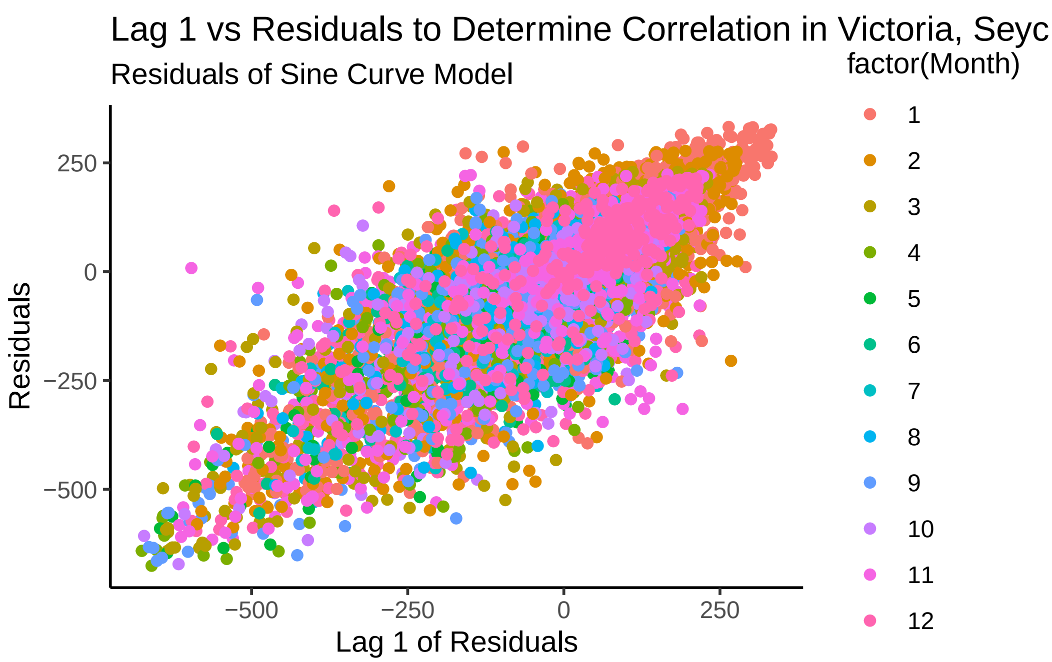


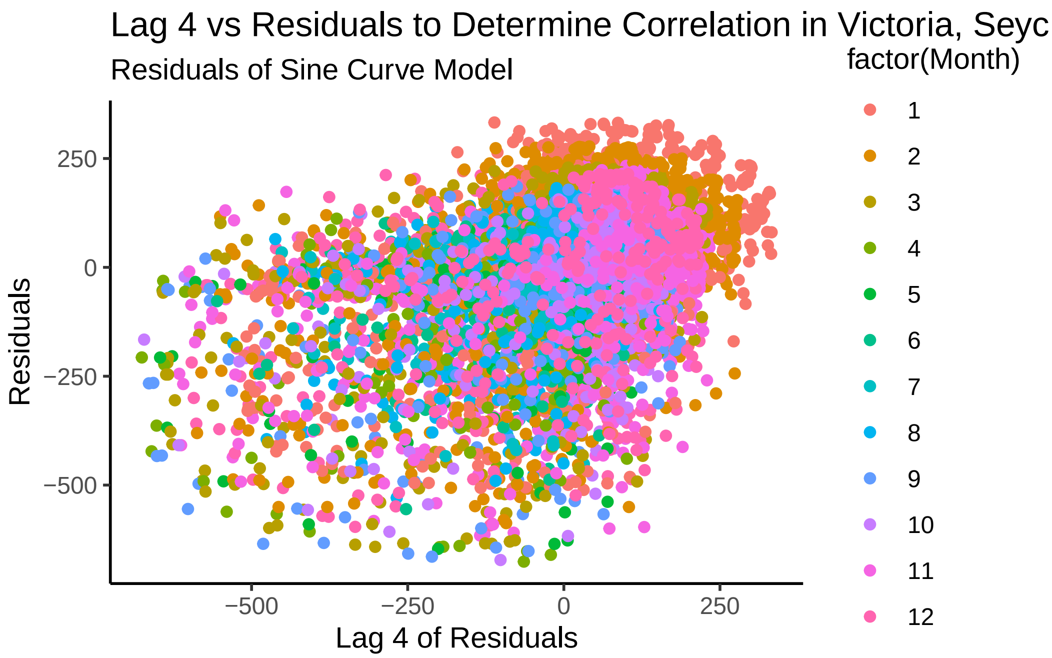
Nairobi, Kenya Intra−Day variation in GHI (W/m^2) by Month
GHI (W/m^2)
900
600
300
0
0
5
10
15
20
Hour (0−24)
Month
12
9
6
3

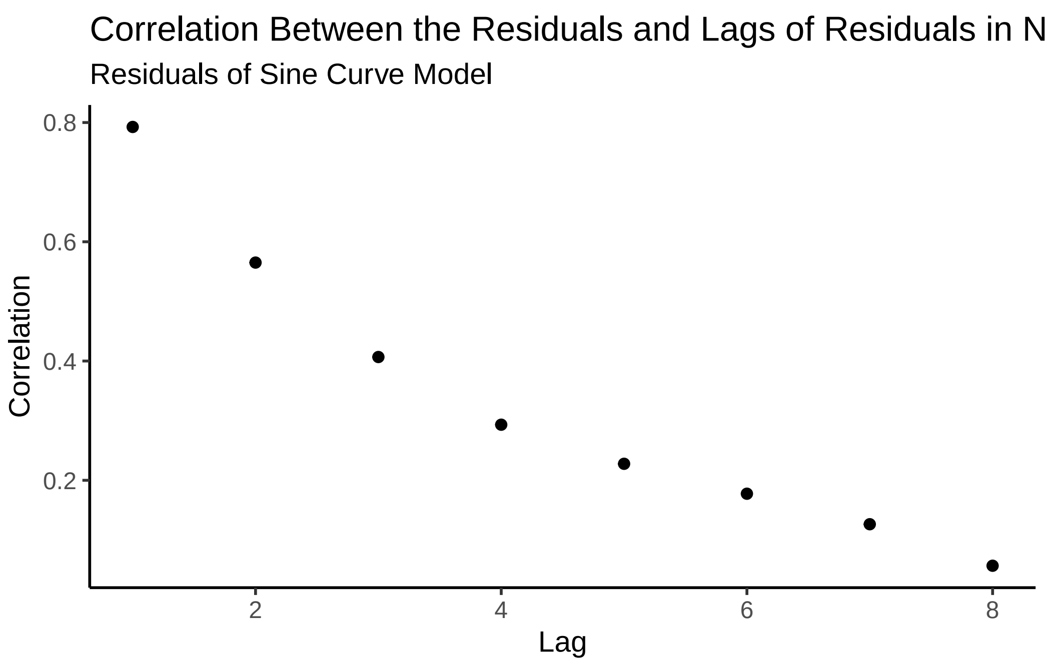


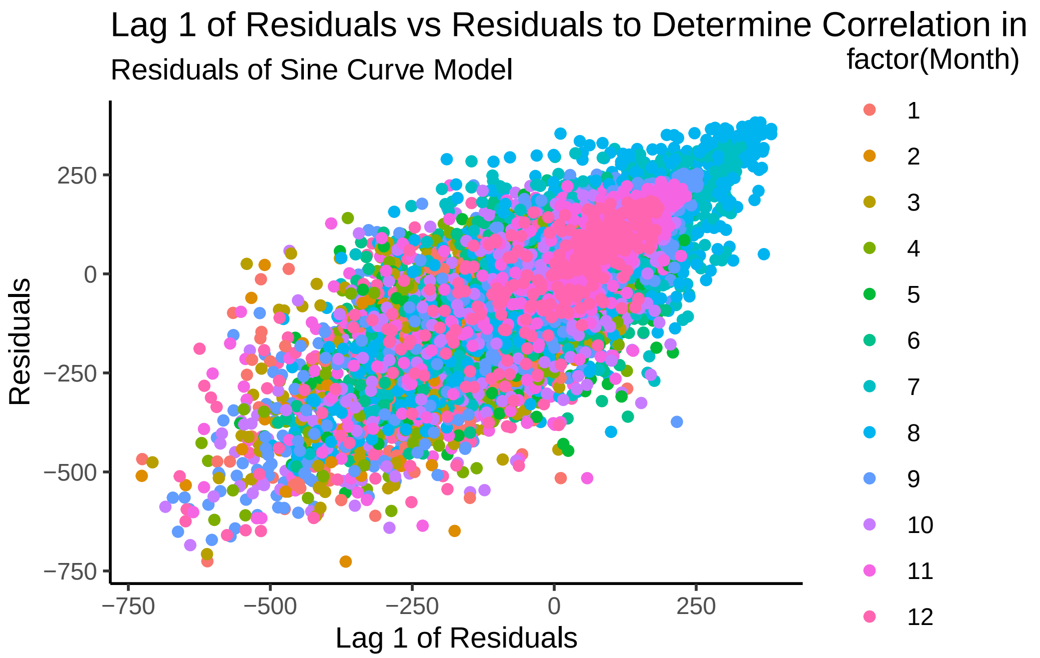
Compared to both Victoria and Nairobi, Manchester's plot looks a bit less clean. This is due to the variation in the length of nighttime across each month, as outlined above. Along the same lines, Manchester has a wider distribution due to this variation, with days in the summer months lasting much longer than those in the winter months. Additionally, Manchester's fitted lines are clustered towards the bottom of the distribution, whereas Victoria's and Nairobi's are clustered near the top, indicating the prevalence of low GHI days in Manchester, and high GHI days in Seychelles and Kenya.

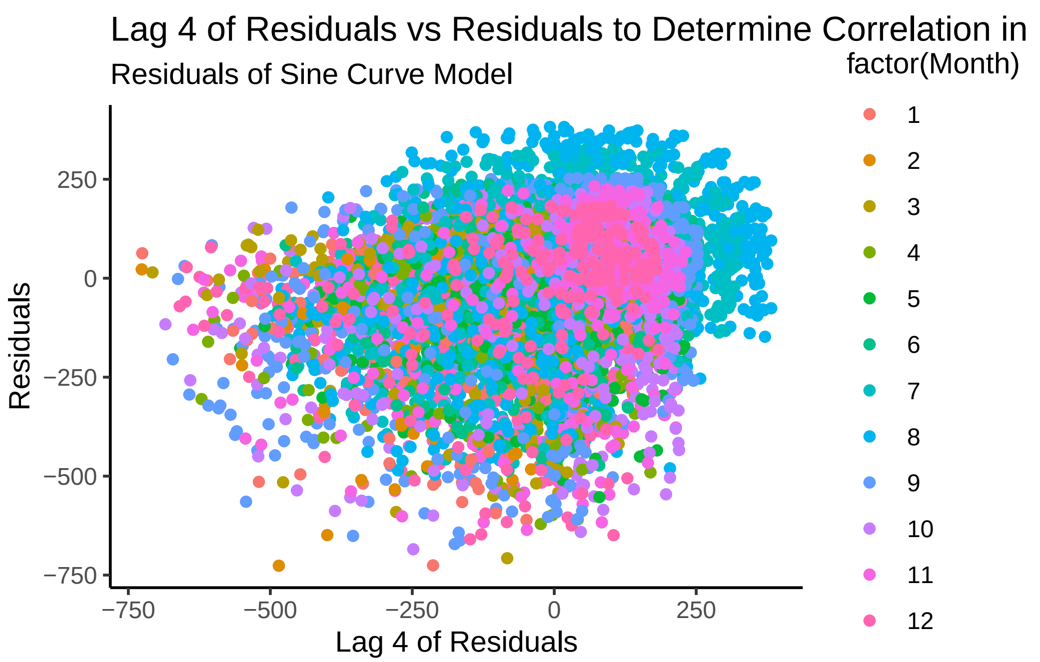

Correlation Between the Residuals and Lags of Residuals in Vi
Residuals of Sine Curve Model
Correlation
0.6
0.4
0.2
2
4
6
8
Lag

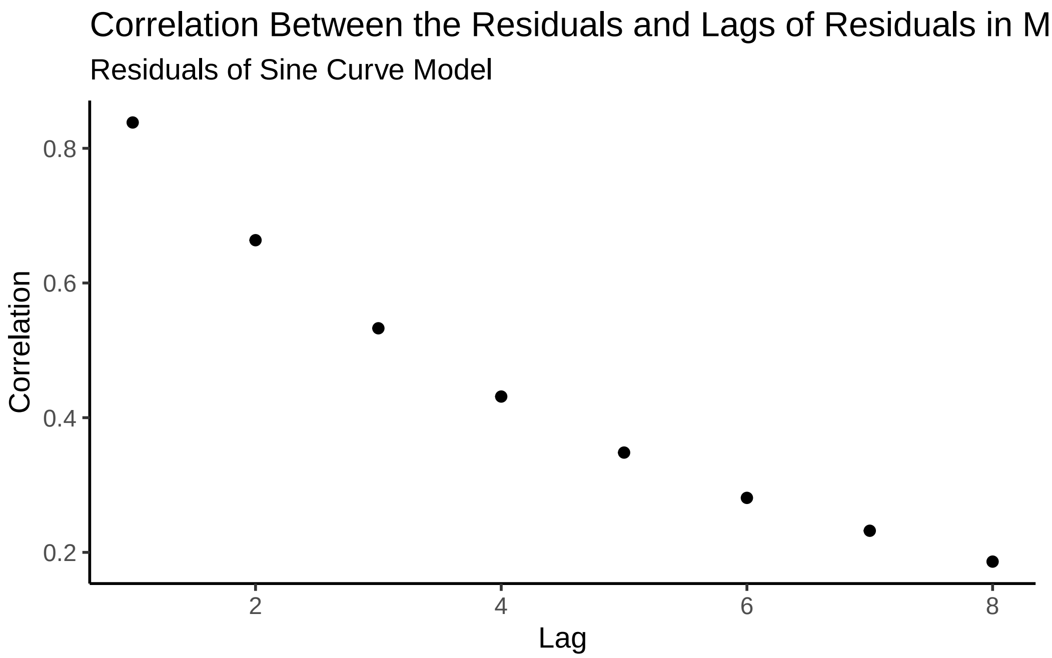

Lag 1 vs Residuals to Determine Correlation in Victoria, Seyc
Residuals of Sine Curve Model
factor(Month)
1
2
3
4
5
6
7
8
9
10
11
12
Residuals
250
0
−250
−500
−500
−250
0
250
Lag 1 of Residuals

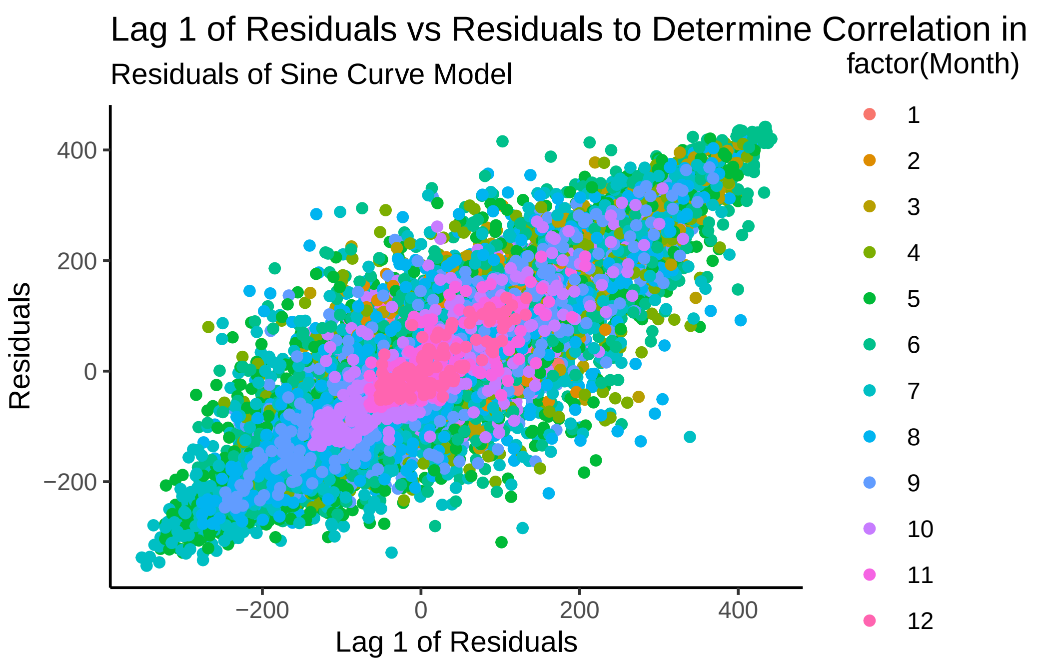

Lag 4 vs Residuals to Determine Correlation in Victoria, Seyc
Residuals of Sine Curve Model
factor(Month)
1
2
3
4
5
6
7
8
9
10
11
12
250
0
−250
−500
Residuals
−500
−250
0
250
Lag 4 of Residuals

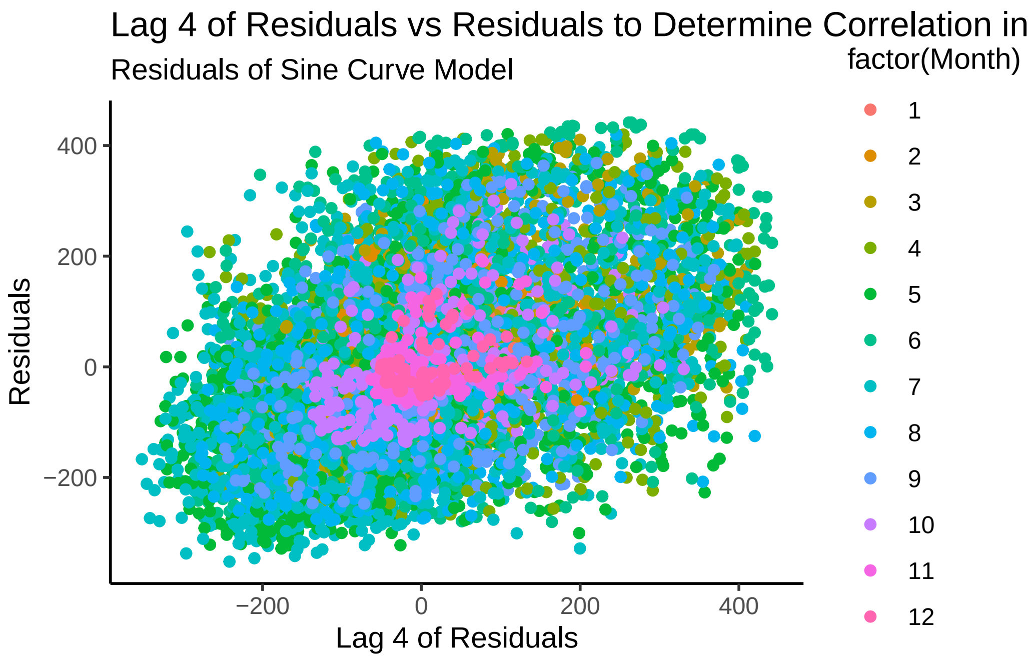

Correlation Between the Residuals and Lags of Residuals in Na
Residuals of Sine Curve Model
0.8
0.6
0.4
0.2
Correlation
2
4
6
8
Lag


Lag 1 of Residuals vs Residuals to Determine Correlation in M
Residuals of Sine Curve Model
factor(Month)
1
2
3
4
5
6
7
8
9
10
11
12
250
0
−250
−500
−750
Residuals
−750
−500
−250
0
250
Lag 1 of Residuals

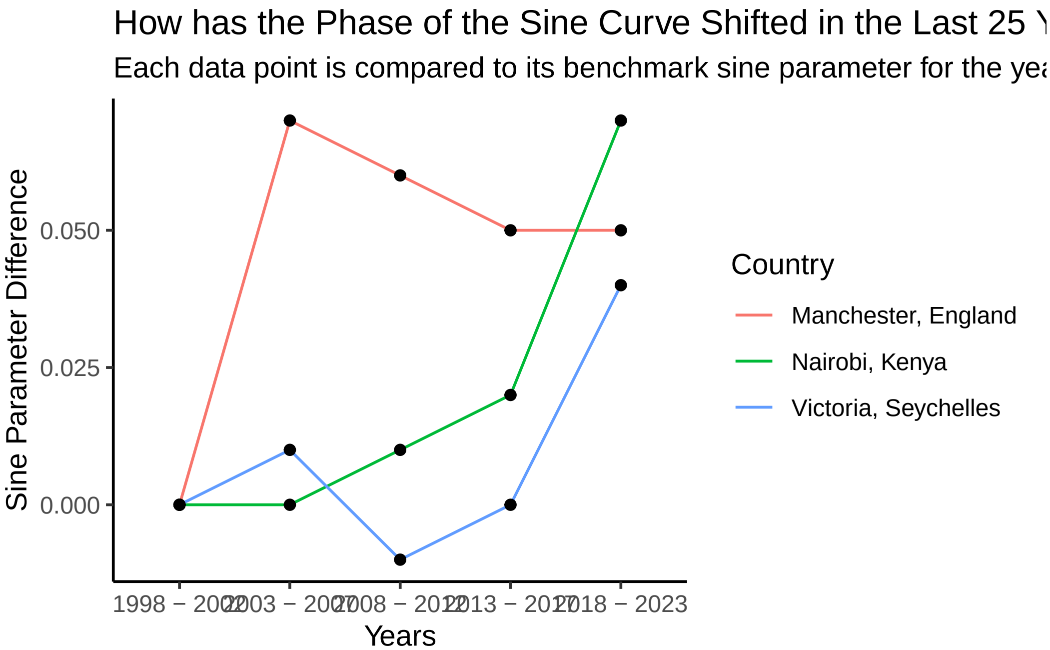

Lag 4 of Residuals vs Residuals to Determine Correlation in M
Residuals of Sine Curve Model
factor(Month)
1
2
3
4
5
6
7
8
9
10
11
12
250
0
−250
−500
−750
Residuals
−750
−500
−250
0
250
Lag 4 of Residuals


Correlation Between the Residuals and Lags of Residuals in M
Residuals of Sine Curve Model
Correlation
0.8
0.6
0.4
0.2
2
4
6
8
Lag


Lag 1 of Residuals vs Residuals to Determine Correlation in M
Residuals of Sine Curve Model
factor(Month)
1
2
3
4
5
6
7
8
9
10
11
12
Residuals
400
200
0
−200
−200
0
200
400
Lag 1 of Residuals

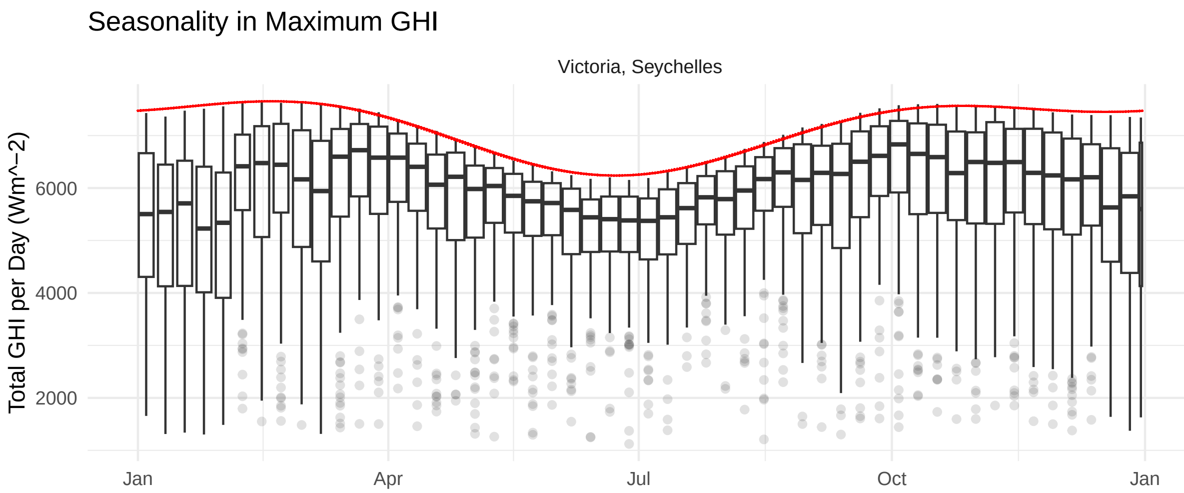


The final step in using a time series approach to model the intra-day variation in GHI is to remove the trend and seasonality. We do that by looking at the distribution of our residuals from our model, which is basically subtracting our model predictions from the actual values. The next step is to use tools like an autocorrelation function (ACF) or a partial autocorrelation function (PACF) to figure out at what lag does the correlation between a given data point and the lag X (X being what we are trying to find) of that data point drop to 0. This is what the visuals above are showing. Suppose in our PACF we saw correlation drop to 0 at lag 3. Then, we would be inclined to use an autoregressive model (AR), with an order of 3. We, however, could not use these methods in order to determine what order AR or MA (moving average) model to use because of the amount of zeros in our data. Because nighttime is an unavoidable obstacle, the PACF and ACF do not work well with finding correlation at different lags, because some of those lags might be zero, which messes up calculations. Therefore, we ran a longitudinal model called GEE. After running the GEE model with all 8 lags as explanatory variables, and the residuals as the response variable, only the first 4 lags were significant. That is why we used the GEE model; it allowed us to determine which lags were significant in predicting the residual, the data point that the lag is connected too. In essense, it is the same thing as the PACF. Whichever lag is no longer significant is the same as "dropping to 0" in a PACF. To contextualize this a bit, this is saying that, if it is 1 in the afternoon, the GHI at 9 in the morning helps to explain some of the GHI at 1. But, the GHI at 8 in the morning doesn't quite explain enough of the GHI at 1 for it to be significant enought to include in our model. From this, an AR(4) model was determined to be the most optimal for our data. Interestingly, these conclusions are similar across both Manchester, Nairobi, and Victoria. All the plots above were done on the first 5 years of data, from 1998-2002. We did this same process outlined above for each 5 year segment, for a total of 5 times. The sine curve parameter is the only difference between models across each 5 year period. Shown below is a plot of change in the sine curve parameter over time, setting the 1998-2002 parameter as a benchmark.



Interestingly, the parameter consistently increased from 1998 to 2023 across all 3 locations. Remember that this parameter deals with the shifting of the sine curve left or right, changing when the peak of the curve happens in a day. The increase in parameter means that, from 1998 to 2023, our models' predicted peak of GHI happened later in the day compared to the 5 year period of 1998 to 2002. This is an interesting observation, because one might think it be possible for the amplitude of the sine curve to increase over time, maybe due to climate change, but where the peak amount of GHI is during the day should stay the same across time. Anna Williams, an Astrology professor at Macalester College, queried that this might have to do with the tilt of the earth. Over thousands of years, the earth's tilt changes, resulting in seasonal changes in climate. On such a small scale as a 25 year period, the tilt of the earth changes ever so slightly. No research has been done to support this claim. We hope to explore this more in the future. Now that we have modeled the intra-day variation in GHI, we next need to explore the seasonal variation in GHI. How does GHI change throughout a year?

Daily GHI data for the Seychelles has multiple sources of seasonality. The maximum solar energy available changes in a regular pattern because of the rotation of the earth. Seasonal weather patterns influence the amount of solar energy reaching the ground (which is available to solar power installations). Since we have solar insolation data the seasonality in the maximum energy available is easy to account for. If we model the maximum GHI for each day of the year based on solar insolation alone we get a reasonable estimate of maximum GHI.



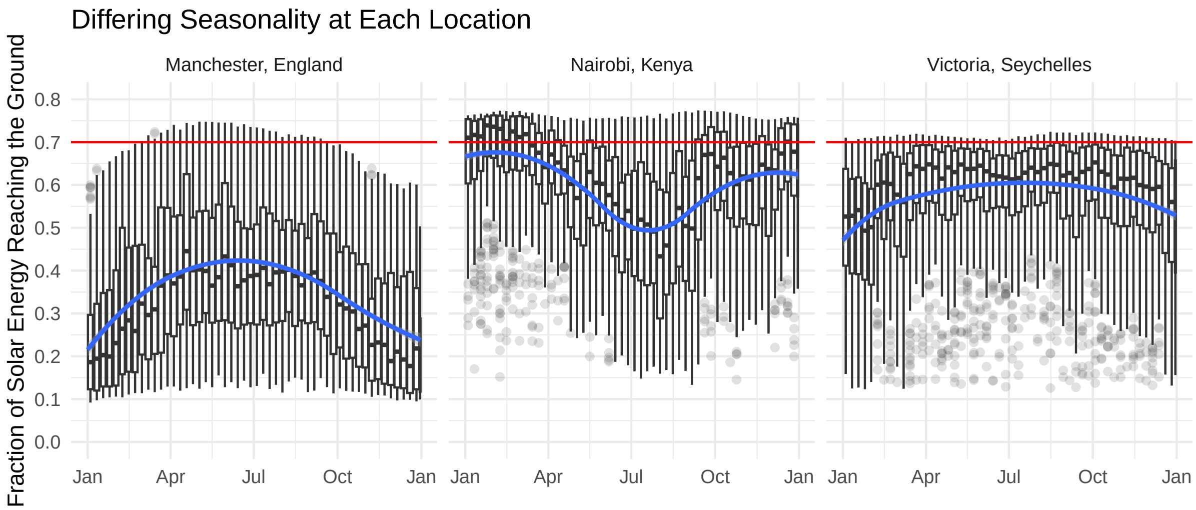
To deal with the first source of seasonality we change our approach. Instead of modeling GHI, we examine the fraction of energy that reaches the ground using the equation:

$$\text{Fraction of energy reaching ground} = 1 - \frac{\text{Solar Insolation per Day} - \text{GHI per Day}}{\text{Solar Insolation per Day}}$$

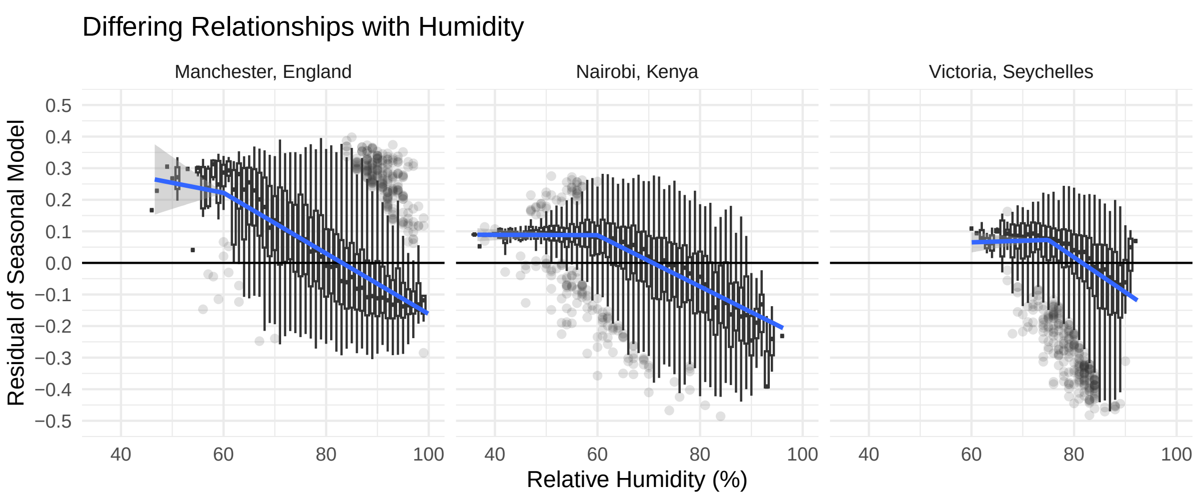
We can now see the effect of seasonal weather patterns more easily. On average, in clear skies, over the entire earth, the atmosphere reflects about 30% of incoming solar energy back into space (Rhodes 2010). In the graph, we can see that the maximum fraction of solar energy that reaches the ground is about 70%. Most days are not perfectly clear though so typically less than 70% of the energy reaches the surface. The overall trend in the fraction of energy reaching the surface is different for each location. Each year we expect the fraction of energy reaching the ground to fall by 0.00173 in Victoria, and 0.00168 in Nairobi, but rise by 0.0008 in Manchester. Solar insolation does not change significantly over these time frames so the trend we observe is likely a result of climate change and weather.

The seasonality not explained by the seasonality of insolation can also be explained by climate and weather: Seychelles receives the most precipitation in January and December, with relatively dry summers ("Climate Change Knowledge Portal" 2023). We decided to model this using a spline with two knots placed to roughly line up with the seasons, this is plotted in blue below. We add this model to the existing trend model.

Manchester also sees rainy winters and so exhibits a similar seasonality ("Climate Change Knowledge Portal" 2023). In Nairobi, the seasonality is not associated with precipitation but rather with low clouds that are present in June, July, and August (Ng'ang'a 1992). This background on the seasons informs our placement of knots for a spline model plotted in blue. The residuals of this approach have significant relationships with humidity, suggesting we should add humidity to our model.

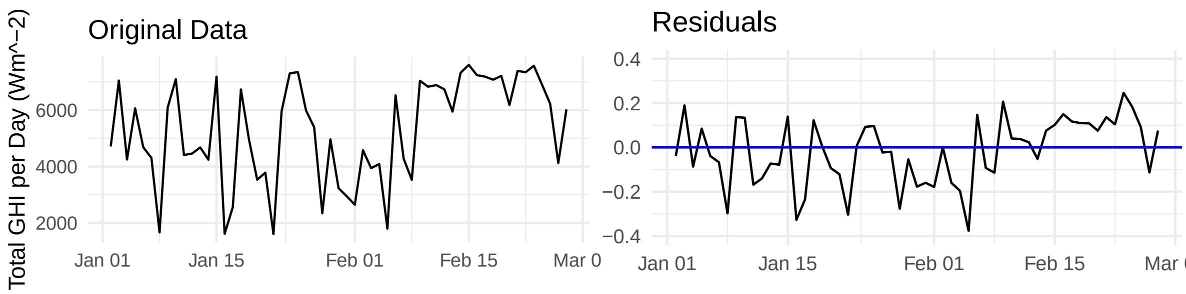


Average daily relative humidity lets us model more of the relationship between weather and the energy reaching the ground. Humidity impacts the clear sky GHI (the GHI if there were no clouds) and can give us some sense of cloud cover since the data lacks any direct measure of cloud cover. We incorporate humidity into the three seasonal models as a degree 1 spline with a single knot. This reflects the observation that the unexplained relationship with humidity is relatively flat at lower humidity but is more negative at higher humidity. Kenya and England have their knots at a relative humidity of 60%, the Seychelles has its at 75%. This greatly reduces or eliminates the relationship between the model residuals and humidity.

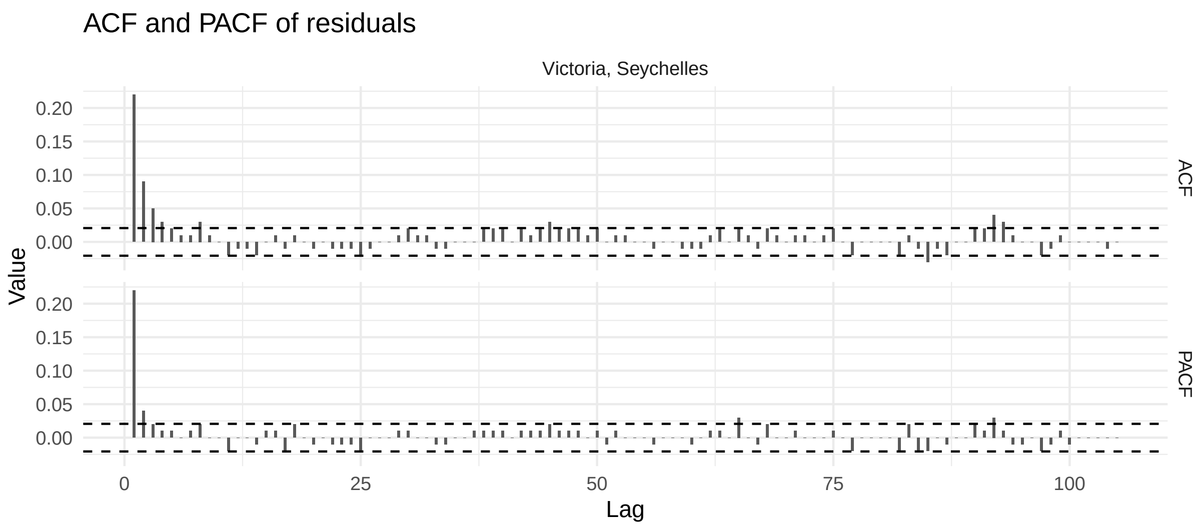


After applying the seasonal model the residual relationship with humidity depends on location

If we examine just a short period in the Seychelles, the first 2 months of 2002, it appears that many of the shorter-term fluctuations in GHI are still present after modeling. These fluctuations seem to be about 2-4 days long which is why the AR models with order 2-4 performed well on various error metrics.

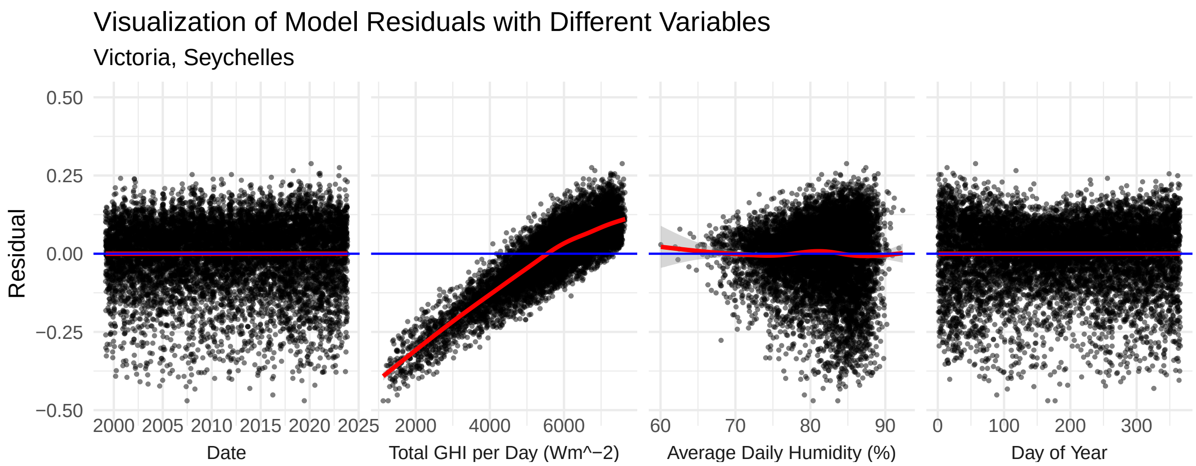


The errors of the combined model with seasonality, humidity, and the overall trend are not white noise, so we attempt to model them. The ACF plot tells us the correlation between errors that are a specific number of lags (measured in days) apart. The ACF decreases as the lags increase. This makes sense because the farther away two points are from one another in time, the weaker the correlation between those two values will be. The PACF plot tells us the correlation between errors and a specific number of lags apart after accounting for the correlation between shorter lags. Values within the dashed lines indicate independent white noise.



We come up with five candidate models after looking at the ACF and PACF: an AR3 model, an AR2 model, an AR2MA1 model, an AR3MA1 model, and an AR3MA2 model. These choices are primarily based on the slower decay of the ACF and the PACF being within the dashed lines after 3 lags. Of these, the AR2MA1 model performs the best as measured by AIC (A goodness of fit measure with a penalty for model terms). The Ljung-Box statistic has large p-values, so we fail to reject the null hypothesis that the residuals are independent There is no proof of serial correlation after fitting the model. The AR3 model performed very similarly on all metrics but has slightly worse fit as measured by AIC.

This procedure is repeated for Manchester and Nairobi with similar results. Using the same criteria as above we settle on an AR3MA1 model for Manchester and England. In England, the AR4 and AR3 models performed well. In Kenya, the AR3 model also performed well. The actual effects of these models are very slight, but they do handle the correlation present.



The Seychelles model residuals have a reasonably insignificant relationship with date, day of year, and humidity. The strong positive relationship between daily GHI and the residual suggests a major issue with the model: we underpredict the fraction of sun reaching the ground on days with higher total GHI and overpredict on days with lower total GHI.

Forecasting with this complete model is not possible without a provided forecast for humidity since we relied on having a known humidity value when creating the model.

Discussion and Conclusions

Researchers have explored many different ways to model GHI. These methods come down to ways to model weather and cloud cover. Researchers have used an autoregressive spatio-temporal model to model GHI in the short term (Dambreville et al. 2014). The forecasts resulting from that model were just 15 minutes to 1 hour. Yang et al. (2015) uses exponential smoothing to predict cloud cover and, therefore, GHI. Aside from time series and spatial approaches, neural networks have proven quite successful at accounting for the highly complex relationships that impact GHI. Zang et al. (2020) combine two types of neural networks for spatio-temporal predictions. This, of course, sacrifices interpretability.

Add discussion of intraday data/modeling.

Our approach to modeling the daily sunlight highlights the incredible variation between different locations and the difficulty of accurately forecasting weather conditions. When first converting from GHI to our metric, the fraction of sunlight reaching the ground, each location was at a different latitude and required a different set of solar insolation data. This solar insolation data simplifies the modeling process but there was still a lot to model. The overall trend in the fraction of sunlight reaching the ground, which has been roughly linear, is also different for each of the three locations. Each location experienced different climate and yearly weather patterns and required separate spline models for the time of year. Even the relationship between relative humidity and the fraction of sunlight reaching the ground was unique between the locations. Residuals after all this modeling showed correlation for data 2 to 4 days apart. That is just a result of how the weather tends to work from day to day. Each of the locations showed relatively similar results for this part of the modeling.

The overall trend we saw in daily sunlight in the Seychelles could be a part of general trends in cloud cover in the Area. Sharma et al. (2023) shows that cloud cover in the Arabian Sea (north of the Seychelles) has started to increase and predicts that it will continue to increase as climate change continues. These results are not generalizable as they find the opposite trend in the South-East Indian Ocean.

Overall, the model performed poorly. Lacking a solid proxy for cloud cover in the data, we had little way to predict the day-to-day changes in cloud cover. Of course, modeling would be very straightforward if we had a measure of cloud cover. The fraction of sunlight reaching the ground is directly linked to cloud cover. That would also force us to rely on some outside cloud cover forecast if we wanted to do any forecasting. In existing work, spatial data and spatial models play a critical role. Weather is notoriously difficult to predict and this difficulty is amplified when the data you have is for single locations hundreds of miles apart.

Acknowledgements

References

Boland, John. 2020. "Characterising Seasonality of Solar Radiation and Solar Farm Output." *Energies* 13 (2): 471. https://doi.org/10.3390/en13020471.

"Climate Change Knowledge Portal." 2023. World Bank Group. https://climateknowledgeportal.worldbank.org/.

Crook, Julia A., Laura A. Jones, Piers M. Forster, and Rolf Crook. 2011. "Climate Change Impacts on Future Photovoltaic and Concentrated Solar Power Energy Output." *Energy & Environmental Science* 4 (9): 3101. https://doi.org/10.1039/c1ee01495a.

Dambreville, Romain, Philippe Blanc, Jocelyn Chanussot, and Didier Boldo. 2014. "Very Short Term Forecasting of the Global Horizontal Irradiance Using a Spatio-Temporal Autoregressive Model." *Renewable Energy* 72 (December): 291–300. https://doi.org/10.1016/j.renene.2014.07.012.

Hartmann, Dennis L. 2016. "The Global Energy Balance." In, 25–48. Elsevier. https://doi.org/10.1016/b978-0-12-328531-7.00002-5.

"Historical Data Validation." 2023. Clean Power Researh; SolarAnywhere. https://www.solaranywhere.com/resources/validation/historical/.

Ng'ang'a, JK. 1992. "The Climate and Meteorology of Nairobi Region, Kenya." *Journal Of Social Development In Africa* 7 (1,2). http://erepository.uonbi.ac.ke:8080/xmlui/handle/123456789/33021.

Rhodes, Christopher J. 2010. "Solar Energy: Principles and Possibilities." *Science Progress* 93 (1): 37–112. https://doi.org/10.3184/003685010x12626410325807.

Rose, Brian, Andrew Williams, Moritz Kreuzer, Chris Cardinale, Sebastian Meßlinger, Xoviat, Ryan Abernathey, Filipe, and Paul Gierz. 2022. *Climlab/Climlab: Version 0.8.1*. Zenodo. https://doi.org/10.5281/ZENODO.6575235.

Shahsavari, Amir, and Morteza Akbari. 2018. "Potential of Solar Energy in Developing Countries for Reducing Energy-Related Emissions." *Renewable and Sustainable Energy Reviews* 90 (July): 275–91. https://doi.org/10.1016/j.rser.2018.03.065.

Sharma, Sahil, Kyung-Ja Ha, Ryohei Yamaguchi, Keith B. Rodgers, Axel Timmermann, and Eui-Seok Chung. 2023. "Future Indian Ocean Warming Patterns." *Nature Communications* 14 (1). https://doi.org/10.1038/s41467-023-37435-7.

Stein, Joshua, Clifford Hansen, and Matthew Reno. 2012. "Global Horizontal Irradiance Clear Sky Models : Implementation and Analysis." https://doi.org/10.2172/1039404.

Wild, Martin, Doris Folini, and Florian Henschel. 2017. "Impact of Climate Change on Future Concentrated Solar Power (CSP) Production." *AIP Conference Proceedings*. https://doi.org/10.1063/1.4975562.

Yang, Dazhi, Vishal Sharma, Zhen Ye, Lihong Idris Lim, Lu Zhao, and Aloysius W. Aryaputera. 2015. "Forecasting of Global Horizontal Irradiance by Exponential Smoothing, Using Decompositions." *Energy* 81 (March): 111–19. https://doi.org/10.1016/j.energy.2014.11.082.

Zang, Haixiang, Ling Liu, Li Sun, Lilin Cheng, Zhinong Wei, and Guoqiang Sun. 2020. "Short-Term Global Horizontal Irradiance Forecasting Based on a Hybrid CNN-LSTM Model with Spatiotemporal Correlations." *Renewable Energy* 160 (November): 26–41. https://doi.org/10.1016/j.renene.2020.05.150.