

Hourly Linear Model Summaries

Hourly Model 1 Usage Baseline 2017-2018:

Predictors	Estimate	Std. Error	t value	p-value
Intercept	251.7409	2.9452	85.475	<2e-16 ***
HDD	14.5734	0.2213	65.843	<2e-16 ***
CDD	-1.0054	0.4809	-2.091	0.0366 *

Note. HDD = Hourly value of heating degree days, CDD = Hourly value of cooling degree days. * indicates statistical significance. Model is fit to predict hourly sum Usage (kWh).

Residual standard error: 152.4 on 8757 degrees of freedom

Multiple R-squared: 0.4148, Adjusted R-squared: 0.4147

F-statistic: 3104 on 2 and 8757 DF, p-value: < 2.2e-16

Hourly Model 2 Usage Baseline 2021-2022:

Predictors	Estimate	Std. Error	t value	p-value
Intercept	219.5138	4.6743	46.962	< 2e-16 ***
HDD	17.4565	0.3657	47.731	< 2e-16 ***
CDD	2.1579	0.7324	2.946	0.00323 **

Note. HDD = Hourly value of heating degree days, CDD = Hourly value of cooling degree days. * indicates statistical significance. Model is fit to predict hourly sum Usage (kWh).

Residual standard error: 237.4 on 8709 degrees of freedom

Multiple R-squared: 0.2569, Adjusted R-squared: 0.2567

F-statistic: 1505 on 2 and 8709 DF, p-value: < 2.2e-16

Hourly Model 3 Demand Baseline 2017-2018:

Predictors	Estimate	Std. Error	t value	p-value
Intercept	270.78243	3.13800	86.291	<2e-16 ***
HDD	16.30835	0.23582	69.155	<2e-16 ***
CDD	-0.01732	0.51234	-0.034	0.973

Note. HDD = Hourly value of heating degree days, CDD = Hourly value of cooling degree days. * indicates statistical significance. Model is fit to predict hourly peak Demand (kW).

Residual standard error: 162.4 on 8757 degrees of freedom

Multiple R-squared: 0.4306, Adjusted R-squared: 0.4305

F-statistic: 3312 on 2 and 8757 DF, p-value: < 2.2e-16

Hourly Model 4 Demand Baseline 2021-2022

Predictors	Estimate	Std. Error	t value	p-value
Intercept	239.0961	4.9505	48.30	< 2e-16 ***
HDD	19.4348	0.3873	50.17	< 2e-16 ***
CDD	3.4054	0.7757	4.39	1.15e-05 ***

Note. HDD = Hourly value of heating degree days, CDD = Hourly value of cooling degree days. * indicates statistical significance. Model is fit to predict hourly peak Demand (kW).

Residual standard error: 251.5 on 8709 degrees of freedom

Multiple R-squared: 0.2713, Adjusted R-squared: 0.2711

F-statistic: 1621 on 2 and 8709 DF, p-value: < 2.2e-16

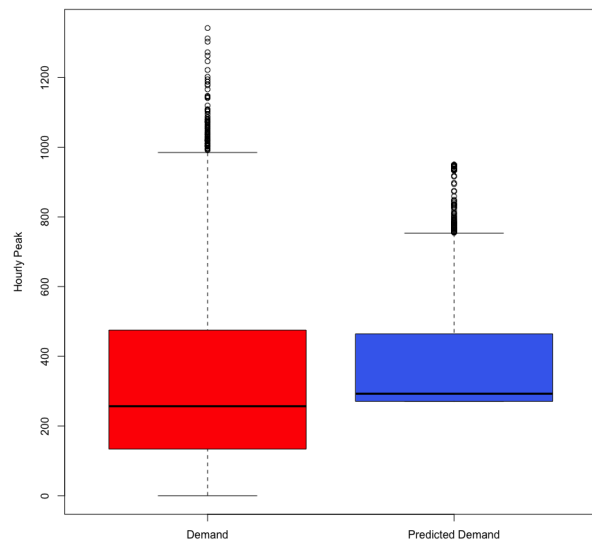
Takeaways: By evaluating these models on correlation coefficient or “Multiple R-squared”, we see that none display an adequate fit. The 2017-2018 baseline period models (both Usage and Demand) boast much higher R values of 0.4148 and 0.4306 respectively, compared the 2021-2022 Usage and Demand model coefficients of 0.2569 and 0.2713. However, even the much higher 2017-2018 models fall far short of our desired level. These lower R-squared values mean that a lot of the variation in the data cannot be explained by the models.

Table 1 Predicted Usage/Demand vs Observed:

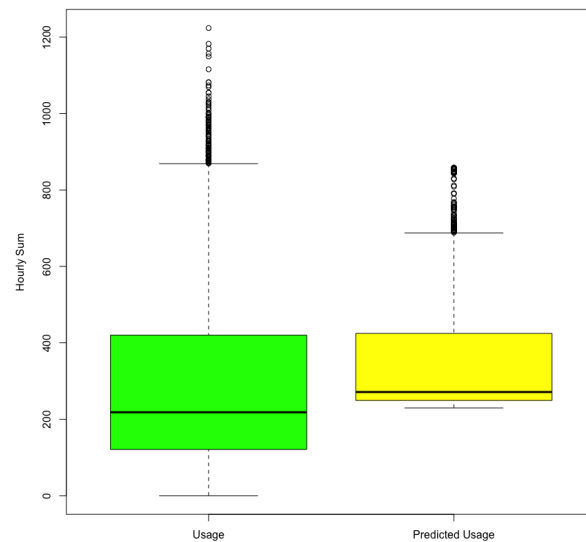
month	sum_predictedUsage	sum_Usage	AvoidedUsage	predictedDemand	Demand	AvoidedDemand
0022-07-01	180508.2828	40428.315	140079.9678	289.8632022	586.8	-296.9367978
0022-08-01	181787.748	162540.585	19247.16305	320.6859851	790.38	-469.6940149
0022-09-01	188191.8015	143834.94	44356.86155	434.029023	747.36	-313.330977
0022-10-01	239983.5924	130066.335	109917.2574	656.8010946	781.2	-124.3989054
0022-11-01	306683.6753	261417.915	45265.76025	692.0271322	1221.66	-529.6328678
0022-12-01	376866.9393	351773.505	25093.43426	904.1987758	1302.66	-398.4612242
0023-01-01	360649.5409	320979.78	39669.76087	774.5473871	1341.9	-567.3526129
0023-02-01	311020.6504	245099.115	65921.5354	764.2731261	1272.42	-508.1468739
0023-03-01	314194.2586	231119.28	83074.97859	728.557838	1074.42	-345.862162
0023-04-01	236547.4816	109159.29	127388.1916	578.0317603	662.58	-84.54823968
0023-05-01	200670.414	127842.345	72828.06897	508.558186	597.06	-88.50181397

Note. sum_predictedUsage = Monthly sums of predicted Usage, sum_Usage = Monthly sums of observed Usage, Avoided Usage = (sum_predictedUsage - sum_Usage), predictedDemand = Monthly peak predicted Demand, Demand = Monthly peak observed Demand, AvoidedDemand = (predictedDemand - Demand). Both predicted Usage and Demand are determined by 2017-2018 baseline models.

Graph 1 Observed Hourly Peak Demand vs Predicted Demand:



Graph 2 Observed Hourly Sum Usage vs Predicted Usage:



Note. Predicted Demand and Predicted Usage are determined by the 2017-2018 baseline models.

Takeaways: Our main variables of interest from Table 1 are AvoidedUsage/Demand. These variables depict the difference or savings between model predicted Usage/Demand values and observed Usage/Demand values after solar installation. We expect all Avoided values to be positive as the post installation period should lead to less energy usage than the model (which does not account for the installed system benefits) predicts. While this is the case for AvoidedUsage, AvoidedDemand has all negative values. After analyses and graphing, we can infer this difference is due to the way Demand is predicted/recorded compared to Usage. Demand is a monthly maximum value, which means outliers will become a major influence. While Usage is a monthly sum value, meaning there is less emphasis on outliers and more on mean values. As we can see from Graph 1 and Graph 2, observed demand and usage ranges from 0 to roughly 1300, while predicted demand/usage falls between roughly 250 and 900. With only a few predictor variables and a lot of variation it's hard for a regression model to accurately predict a wide range of values, this is seen from our subpar R-squared coefficients. However, both predicted demand and usage have a higher mean than observed. This is why we see higher predicted Usage values than observed (which is a sum value) but not higher predicted Demand values than observed (a maximum/outlier value). Based on the low R-squared coefficients and the negative AvoidedDemand values, we can assume the hourly models can not represent enough of the variation in the dataset and are not a good fit.

Monthly Linear Model Summaries

Monthly Model 1 Usage Baseline 2017-2018:

Predictors	Estimate	Std. Error	t value	p-value
Intercept	155,000	24,450	6.341	0.000134 ***
sum_HDD	16.70	2.536	6.585	0.000101 ***
sum_CDD	8.464	6.049	1.399	0.195246

Note. sum_HDD = Monthly sum value of heating degree days, sum_CDD = Monthly sum value of cooling degree days.

* indicates statistical significance. Model is fit to predict monthly sum Usage (kWh).

Residual standard error: 27220 on 9 degrees of freedom

Multiple R-squared: 0.9017, Adjusted R-squared: 0.8799

F-statistic: 41.29 on 2 and 9 DF, p-value: 2.924e-05

Monthly Model 2 Usage Baseline 2021-2022:

Predictors	Estimate	Std. Error	t value	p-value
Intercept	67591.899	20598.861	3.281	0.008270 **
sum_HDD	24.626	2.528	9.742	2.02e-06 ***
sum_CDD	27.987	4.958	5.645	0.000214 ***

Note. sum_HDD = Monthly sum value of heating degree days, sum_CDD = Monthly sum value of cooling degree days.

* indicates statistical significance. Model is fit to predict monthly sum Usage (kWh).

Residual standard error: 29290 on 10 degrees of freedom

Multiple R-squared: 0.9073, Adjusted R-squared: 0.8888

F-statistic: 48.94 on 2 and 10 DF, p-value: 6.844e-06

Monthly Model 3 Demand Baseline 2017-2018:

Predictors	Estimate	Std. Error	t value	p-value
Intercept	554.15137	66.89564	8.284	1.67e-05 ***
sum_HDD	0.04931	0.00694	7.105	5.63e-05 ***
sum_CDD	0.03991	0.01655	2.411	0.0392 *

Note. sum_HDD = Monthly sum value of heating degree days, sum_CDD = Monthly sum value of cooling degree days.

* indicates statistical significance. Model is fit to predict monthly peak Demand (kW).

Residual standard error: 74.48 on 9 degrees of freedom

Multiple R-squared: 0.897, Adjusted R-squared: 0.8741

F-statistic: 39.18 on 2 and 9 DF, p-value: 3.616e-05

Monthly Model 4 Demand Baseline 2021-2022

Predictors	Estimate	Std. Error	t value	p-value
Intercept	876.870497	116.787025	7.508	2.04e-05 ***
sum_HDD	0.057004	0.014331	3.978	0.00261 **
sum_CDD	-0.009355	0.028107	-0.333	0.74614

Note. sum_HDD = Monthly sum value of heating degree days, sum_CDD = Monthly sum value of cooling degree days.
 * indicates statistical significance. Model is fit to predict monthly peak Demand (kW).

Residual standard error: 166.1 on 10 degrees of freedom

Multiple R-squared: 0.7796, Adjusted R-squared: 0.7356

F-statistic: 17.69 on 2 and 10 DF, p-value: 0.0005196

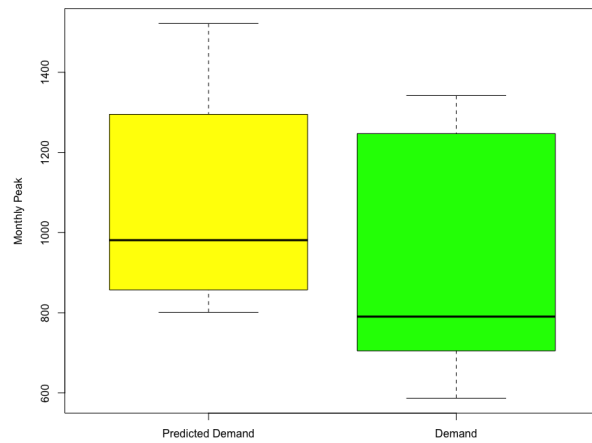
Takeaways: By evaluating these models on correlation coefficient or “Multiple R-squared”, we see that all display an adequate fit. Most models display an R-squared value around 0.9, while 2021-2022 Demand falls to 0.78. However, each model falls above our desired level of 0.75. These higher R-squared values mean that a lot of the variation in the data can be explained by the models.

Table 1 Predicted Usage/Demand vs Observed:

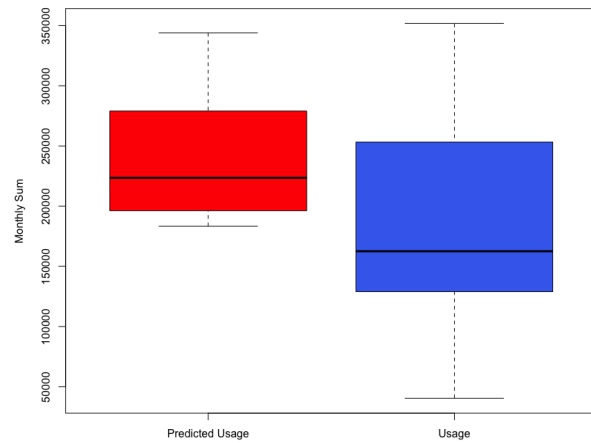
month	predictedUsage17	predictedDemand21	Demand	sum_Usage	AvoidedUsage17	AvoidedDemand21
0022-07-01	223555.8204	801.1213504	586.8	40428.315	183127.5054	214.3213504
0022-08-01	213465.7132	812.2737877	790.38	162540.585	50925.12816	21.89378768
0022-09-01	191360.0383	846.448992	747.36	143834.94	47525.09835	99.08899198
0022-10-01	198167.1858	977.2728406	781.2	130066.335	68100.85079	196.0728406
0022-11-01	272259.2868	1267.036754	1221.66	261417.915	10841.37177	45.37675351
0022-12-01	343982.7402	1521.752204	1302.66	351773.505	-7790.764811	219.0922042
0023-01-01	323135.2	1450.65487	1341.9	320979.78	2155.419993	108.7548703
0023-02-01	285817.8707	1323.22047	1272.42	245099.115	40718.75575	50.80047036
0023-03-01	262888.903	1241.98705	1074.42	231119.28	31769.62296	167.5670496
0023-04-01	194227.1189	981.1214911	662.58	109159.29	85067.82889	318.5414911
0023-05-01	183322.049	867.5865036	597.06	127842.345	55479.70399	270.5265036

Note. predictedUsage17 = Monthly sums of predicted Usage from 2017-2018 baseline model, sum_Usage = Monthly sums of observed Usage, AvoidedUsage17 = (predictedUsage17 - sum_Usage), predictedDemand21 = Monthly peak predicted Demand from 2021-2022 baseline model, Demand = Monthly peak observed Demand, AvoidedDemand21 = (predictedDemand21 - Demand). 07/22 is an incomplete month for observed usage and demand.

Graph 1 Observed Monthly Peak Demand vs Predicted Demand:



Graph 2 Observed Monthly Sum Usage vs Predicted Usage:



Takeaways: Our main variables of interest from Table 1 are AvoidedUsage/Demand. These variables depict the difference or savings between model predicted Usage/Demand values and observed Usage/Demand values after solar installation. We expect all Avoided values to be positive as the post installation period should lead to less energy usage than the model (which does not account for the installed system benefits) predicts. Unlike our hourly models, the monthly models have accurately depicted this. Every value of Avoided Usage/Demand is now positive (except 12/22 when there was an error with the system). We can see this in the boxplots as well, which align much better than the hourly ones and have higher predicted means as well as maximum values than observed. Based on the table/graph results and our high correlation coefficients, it is safe to say that these monthly models account for much more variation in the data and are good fits.

GHI Effect Monthly Linear Model Summaries

GHI Monthly Model 1 Usage 2022-2023:

Predictors	Estimate	Std. Error	t value	p-value
Intercept	26560	1.954e+04	1.359	0.21632
sum_GHI	1.802e-01	1.304e-01	1.382	0.20944
sum_HDD	2.791e+01	1.757e+00	15.883	9.51e-07 ***
sum_CDD	1.661e+01	3.745e+00	4.435	0.00303 **

Note. sum_GHI = Monthly sum value of Global Horizontal Irradiation, sum_HDD = Monthly sum value of heating degree days, sum_CDD = Monthly sum value of cooling degree days. * indicates statistical significance. Model is fit to predict monthly sum Usage (kWh).

Residual standard error: 16960 on 7 degrees of freedom

Multiple R-squared: 0.9782, Adjusted R-squared: 0.9688

F-statistic: 104.6 on 3 and 7 DF, p-value: 3.54e-06

GHI Monthly Model 2 Usage 2022-2023:

Predictors	Estimate	Std. Error	t value	p-value
Intercept	1.082e+05	5.214e+04	2.075	0.07667 .
sum_GHI	-6.386e-01	3.993e-01	-1.599	0.15378
sum_GHI:sum_HDD	3.140e-04	6.338e-05	4.955	0.00165 **
sum_GHI:sum_CDD	1.546e-04	6.448e-05	2.399	0.04757 *

Note. sum_GHI = Monthly sum value of Global Horizontal Irradiation, sum_GHI:sum_HDD = interaction effect (product) of GHI and HDD, sum_GHI:sum_CDD = interaction effect (product) of GHI and CDD. * indicates statistical significance. Model is fit to predict monthly sum Usage (kWh).

Residual standard error: 49960 on 7 degrees of freedom

Multiple R-squared: 0.8107, Adjusted R-squared: 0.7295

F-statistic: 9.992 on 3 and 7 DF, p-value: 0.006346

GHI Monthly Model 3 Demand 2022-2023:

Predictors	Estimate	Std. Error	t value	p-value
Intercept	6.003e+02	1.032e+02	5.815	0.000654 ***
sum_GHI	-3.093e-04	6.886e-04	-0.449	0.666874
sum_HDD	7.632e-02	9.283e-03	8.222	7.65e-05 ***
sum_CDD	3.551e-02	1.978e-02	1.795	0.115767

Note. sum_GHI = Monthly sum value of Global Horizontal Irradiation, sum_HDD = Monthly sum value of heating degree days, sum_CDD = Monthly sum value of cooling degree days. * indicates statistical significance. Model is fit to predict monthly peak Demand (kW).

Residual standard error: 89.6 on 7 degrees of freedom

Multiple R-squared: 0.9379, Adjusted R-squared: 0.9113

F-statistic: 35.24 on 3 and 7 DF, p-value: 0.0001356

GHI Monthly Model 4 Demand 2022-2023:

Predictors	Estimate	Std. Error	t value	p-value
Intercept	7.707e+02	1.053e+02	7.320	0.000160 ***
sum_GHI	-2.575e-03	8.063e-04	-3.193	0.015205 *
sum_GHI:sum_HDD	9.591e-07	1.280e-07	7.493	0.000138 ***
sum_GHI:sum_CDD	4.294e-07	1.302e-07	3.298	0.013161 *

Note. sum_GHI = Monthly sum value of Global Horizontal Irradiation, sum_GHI:sum_HDD = interaction effect (product) of GHI and HDD, sum_GHI:sum_CDD = interaction effect (product) of GHI and CDD. * indicates statistical significance. Model is fit to predict monthly peak Demand (kW).

Residual standard error: 100.9 on 7 degrees of freedom

Multiple R-squared: 0.9212, Adjusted R-squared: 0.8875

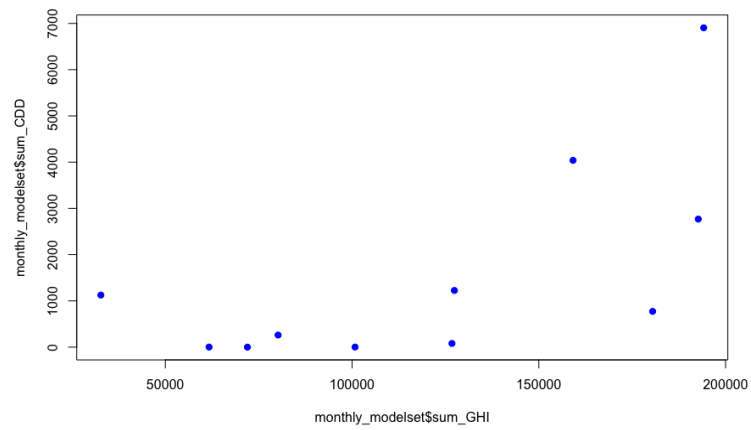
F-statistic: 27.3 on 3 and 7 DF, p-value: 0.0003092

Takeaways: By looking at the t-values and corresponding p-values we can judge the significance of each predictor term. Starting with GHI, we see this variable is non-significant in terms of predicting Demand/Usage on its own. However, GHI does have a significant interaction effect with HDD and CDD on Usage/Demand.

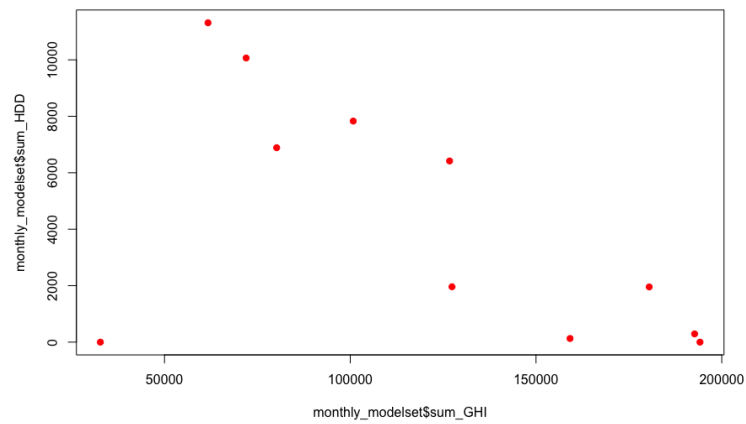
Table 2 Correlation Coefficients between Variables:

	Usage	Demand	GHI	HDD	CDD
Usage	1	0.9489734	-0.3720799	0.9352285	-0.4187249
Demand	0.9489734	1	-0.5217088	0.9529475	-0.5359837
GHI	-0.3720799	-0.5217088	1	-0.577299	0.6557167
HDD	0.9352285	0.9529475	-0.577299	1	-0.689609
CDD	-0.4187249	-0.5359837	0.6557167	-0.689609	1

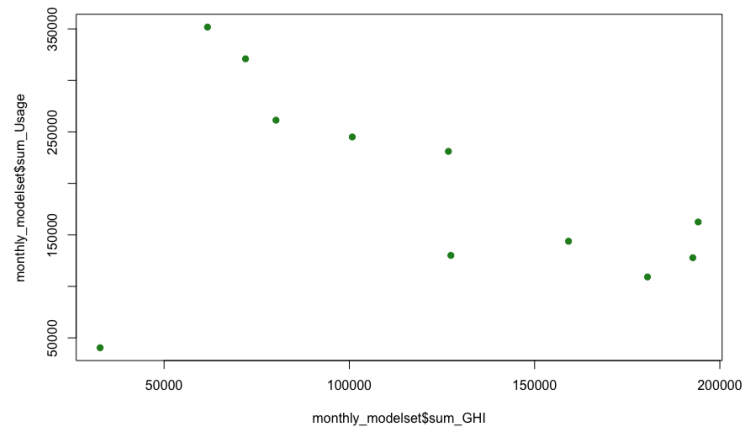
GHI Interaction Plots:



Note. Plotting the monthly sum of GHI (x-axis) and the monthly sum of CDD (y-axis).



Note. Plotting the monthly sum of GHI (x-axis) and the monthly sum of HDD (y-axis).



Note. Plotting the monthly sum of GHI (x-axis) and the monthly sum of Usage (y-axis).

Takeaways: While GHI may not have a significant direct effect on Usage or Demand, it does seem to have significant interactions with HDD and CDD on Usage/Demand. We can see from the correlation coefficient table and graph, that GHI and CDD have a positive relationship (this makes sense as solar radiance increases temperatures rise and so does CDD). We also see that GHI has a negative relationship with HDD (again as solar radiance increases temperatures rise and HDD decreases). I included a third graph showing GHI and Usage, it appears to show a fairly strong negative relationship but what I think this graph is really doing is mimicking GHI and HDD (GHI has an inverse relationship with HDD which has a strong direct relationship with Usage). So while GHI on its own may not help predict Usage/Demand its interaction/product with CDD and HDD can.