Linear Model Summaries

Model 1

Predictors	Estimate	Std. Error	t value	p-value
Intercept	17.385606	1.310288	13.269	<2e-16 ***
GHI	0.640926	0.004908	130.575	< 2e-16 ***
Temp	-1.062301	0.048601	-21.858	< 2e-16 ***
WS	0.620057	0.302572	2.049	0.0412 *

Note. GHI = Global Horizontal Irradiance, Temp = Ambient Temperature (C), WS = Wind Speed (mph). *** indicates statistical significance. Model is fit to predict total output (kWh).

Residual standard error: 5.898 on 361 degrees of freedom
Multiple R-squared: 0.9856, Adjusted R-squared: 0.9855
F-statistic: 8241 on 3 and 361 DF, p-value: < 2.2e-16

Model 2

Predictors	Estimate	Std. Error	t value	p-value
Intercept	1.2538804	0.7641984	1.641	0.102
GHI	0.7458362	0.0063016	118.357	< 2e-16 ***
GHI:Temp	-0.0054737	0.0001757	-31.154	< 2e-16 ***

Note. GHI = Global Horizontal Irradiance, Temp = Ambient Temperature (C), : denotes interaction (product of two variables). *** indicates statistical significance. Model is fit to predict total output (kWh).

Residual standard error: 4.68 on 362 degrees of freedom
Multiple R-squared: 0.9909, Adjusted R-squared: 0.9909
F-statistic: 1.974e+04 on 2 and 362 DF, p-value: < 2.2e-16

Model 3

Predictors	Estimate	Std. Error	t value	p-value
Intercept	13.282158	1.263594	10.51	<2e-16 ***
GHI	0.571762	0.005583	102.42	< 2e-16 ***

Note. GHI = Global Horizontal Irradiance. *** indicates statistical significance. Model is fit to predict total output (kWh).

Residual standard error: 8.967 on 363 degrees of freedom
Multiple R-squared: 0.9665, Adjusted R-squared: 0.9665
F-statistic: 1.049e+04 on 1 and 363 DF, p-value: < 2.2e-16

Takeaways: Basing our decision off of Multiple R-squared (correlation coefficient) and lowest p-values, we can deduce that Model 2 with the highest R^2 of 0.9909 is the best model to use going forward. The coefficients for this model are 1.2538804 for intercept, 0.7458362 for GHI and -0.0054737 for the interaction effect of GHI and Temp. This can be written into a linear equation of y = 1.2538804 + 0.7458362(GHI) -0.0054737(GHI:Temp). Where y=kWh and GHI:Temp is the product of the two.

Summary Stats on Predicted vs Observed Values (Using Model 2)

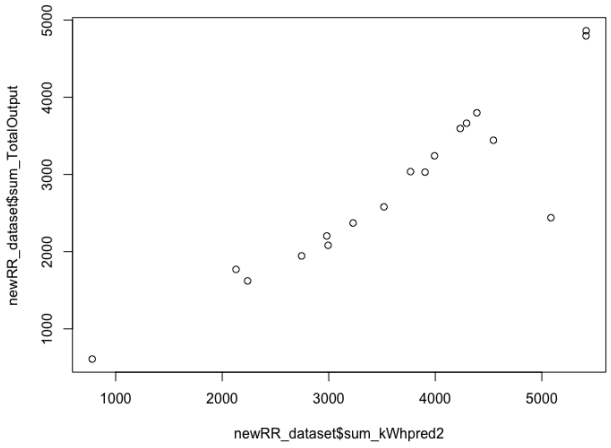
Five number summary stats:

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Energy Ratio (monthly)	0.4797	0.7337	0.7784	0.7739	0.8447	0.8978
kWh Predicted (monthly)	779.8	2984.5	3835.6	3647.6	4366.1	5415.8
kWh Observed (monthly)	608.7	2112.6	2805.2	2838.4	3557.6	4862.2

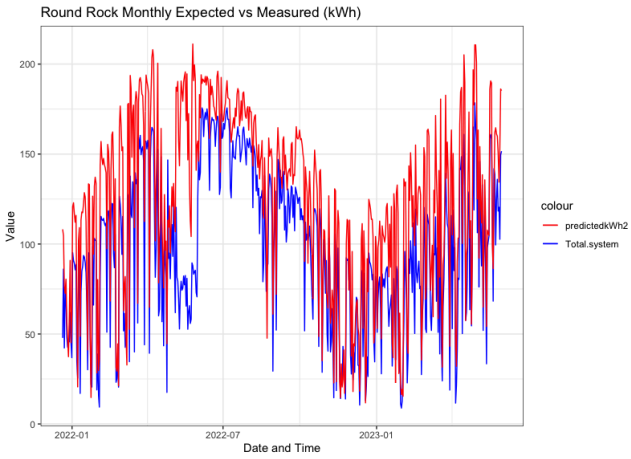
Model 2 predicted values summary stats:

	RMSE	R
Model 2	947.1741	0.9087383

Expected vs Observed (kWh)



Monthly Expected vs Measured (kWh)



Note. newRR_dataset$sum_TotalOutput = Observed/Actual Monthly sum (kWh), newRR_dataset2$sum_kWhpred2 = Monthly sum predicted kWh. Total.system = TotalOutput.

Takeaways: As we can see from the above plots and five number summary statistics, the predicted/expected kWh monthly values are higher on average than the observed ones. This is especially true during the early months of the year as we see much lower ERs in the winter. We also see a minimum ER of 0.4797 which is well below where we'd like to be, we can check the excel output chart to see which month this occurred in and investigate further. These discrepancies could be due to a difference in weather data between what Oiko models and what actually occurred, higher modeled GHI levels would lead to higher expected/predicted kWh levels and therefore these low energy ratios.