Data imputation and pre-processing for Rwanda SL

1. Four advanced streetlights have been installed in Rwanda. Analysis of the data is performed for the period between 1st July 2019 to 31st March 2020. During this period, the wiring of the streetlights was altered between 19th July – 6th Aug 2019. This only has an impact on the load calculation for this period.
2. The data for SL systems is available on Victron portal and is available up to past 6 months. The data files are hence downloaded periodically and stitched together. The weather data is downloaded from Solcast to obtain the Irradiance (and therefore Potential PV power) values. The weather data is available in hourly intervals. Note that the Solcast data is not very accurate and is the closest approximation to the true value (certain inaccuracies in Potential PV power values are particularly notices as actual PV power exceeds the potential PV power on certain days).
3. For our analysis, we extract the following variables from the stitched system and weather data and convert data into hourly means.
    a. Battery Monitor Voltage V (can only be used to calculate light load between 19th July to 6th Aug)
    b. Solar Charger Load current (can only be used to calculate light load between 19th July to 6th Aug)
    c. Battery Monitor State of Charge (found to be erroneous)
    d. Inverter output voltage and output current (to match with AC consumption)
    e. Actual PV power
    f. Solar charger Battery Power (is close to the value of PV power)
    g. System Battery Power
    h. System AC consumption (i.e. socket consumption)
    i. Potential PV power (Irradiance * area of the panel * efficiency of the panel)
4. The data obtained from the system is incomplete owing to communication failures, component failures as well as power outages (caused by insufficient charge in the battery). As a result, we have either partial missing days or completely missing days of data. Fig. 1a shows the yield for hourly PV power data for the 4 streetlights. The yield is calculated based on whether or not a finite value of PV power is available for that hour. Moreover, the daily yield for all variables is shown in fig 1b. The yield is calculated as the % hours each day for which the data is available. From this analysis, we obtain that the % of missing data for PV power (and majority variables) is ~20%, ~12%, ~11% and 8% at SL1, SL2, SL3 and SL4 respectively. However, the % of missing data for the inverter/AC consumption data is slightly higher and is ~44%, ~16%, ~19% and ~19% for SL1, SL2, SL3 and SL4 respectively.

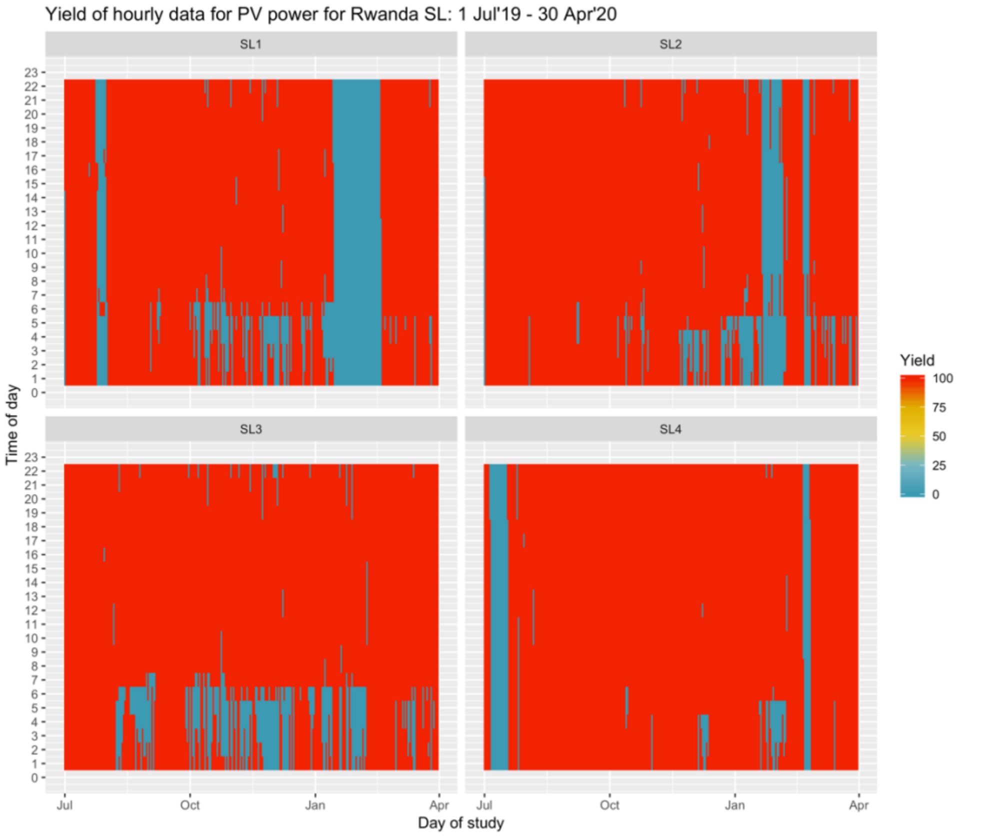


Fig. 1a: Hourly yield of PV power data for all SL

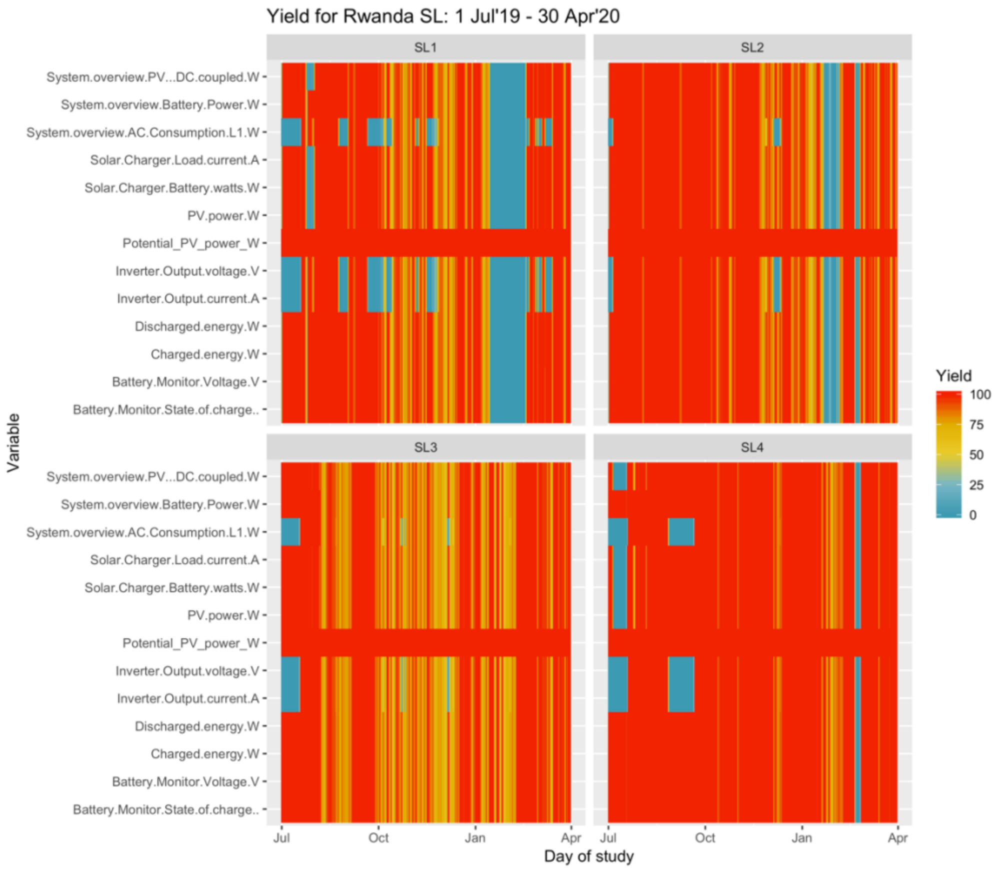


Fig.1b: Daily yield of data for all streetlights

5. In phase 1 of analysis, we only considered full data days i.e. days where data was available for all 24 hours for each SL. The hourly data was analysed to obtain daily socket and light load using the above variables, daily PV power, daily Potential PV power, and daily positive and negative Battery power (battery charge and discharge energy). It was noted that dropping incomplete data days created bias in the analysis owing to missing power outage events. This led to a battery efficiency of 120% and, thus, necessitated imputation for the missing data.
6. Numerous approaches have been proposed for imputation including hot deck imputation (using mean, median or last observed value), kNN, multiple imputation using chained equations, regression, random forests and seasonal decomposition of time series.
7. The energy data from SL was plotted to observe nature of the raw hourly data values for the duration of study to choose the appropriate imputation technique. As can be seen from fig. 2, the PV power, Potential PV power and Battery power values form a time series with a seasonal component. The series shows no trend but a seasonality of 1 day. This is expected as the variables exhibit the same behaviour each day in terms of charging of the battery in the morning and discharging in the evening. Also, the series is seen to have certain irregularities across days (mostly owing to varying Potential PV power values). The same behaviour is seen for all streetlights. Moreover, a similar behaviour is expected for AC consumption and Solar Charger Battery Power values owing to the same usage pattern of the socket and lights across days. This suggests that a time series imputation technique may be suited for this dataset. Furthermore, the symnum functionality in R was also used to understand the correlation between each variable pair above. Of all the variables, a correlation (with R value between 0.6-0.8) was obtained between Potential PV power and PV power and Potential PV power and Battery power. This is shown in fig. 3. The values suggest that Generalized Additive Models (GAM) may also be used to impute values for PV power and Battery power based on Potential PV power data. The two approaches were, therefore, evaluated.

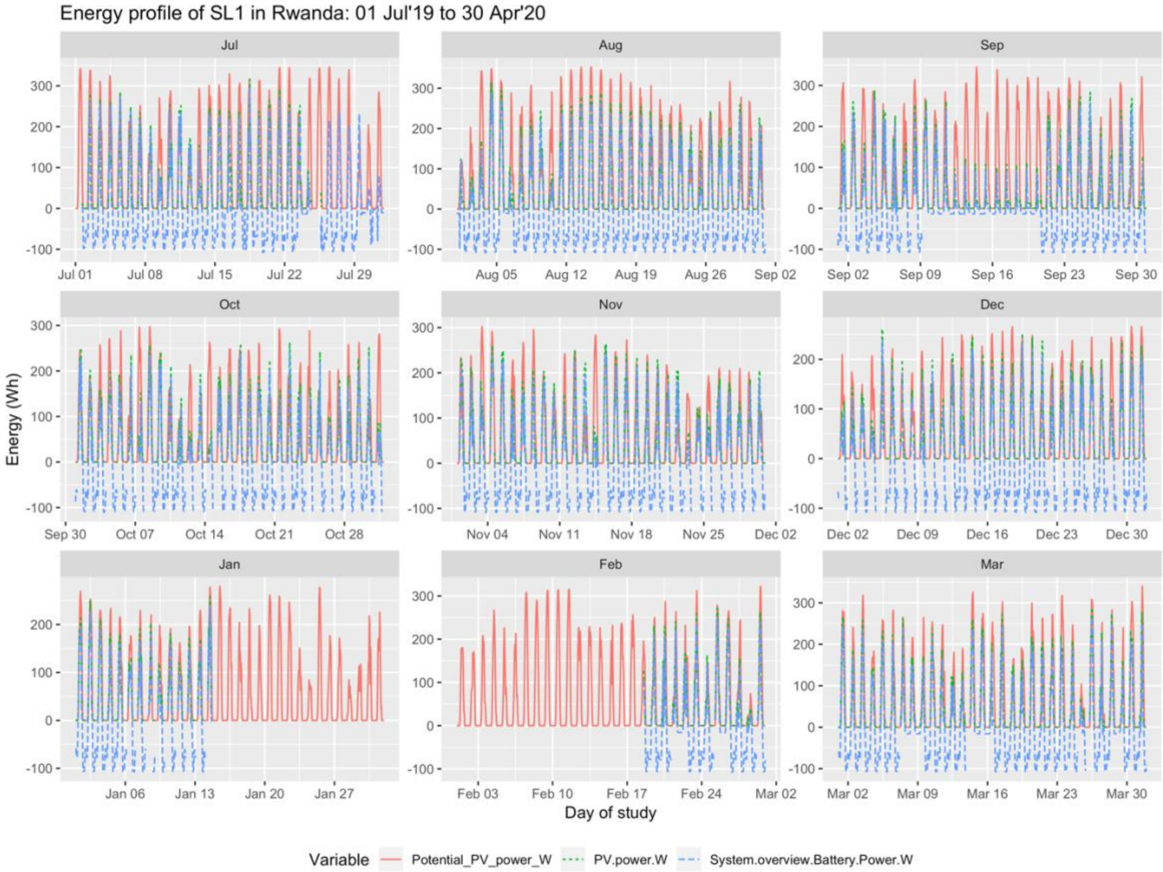


Fig. 2: Seasonality of energy data for SL1

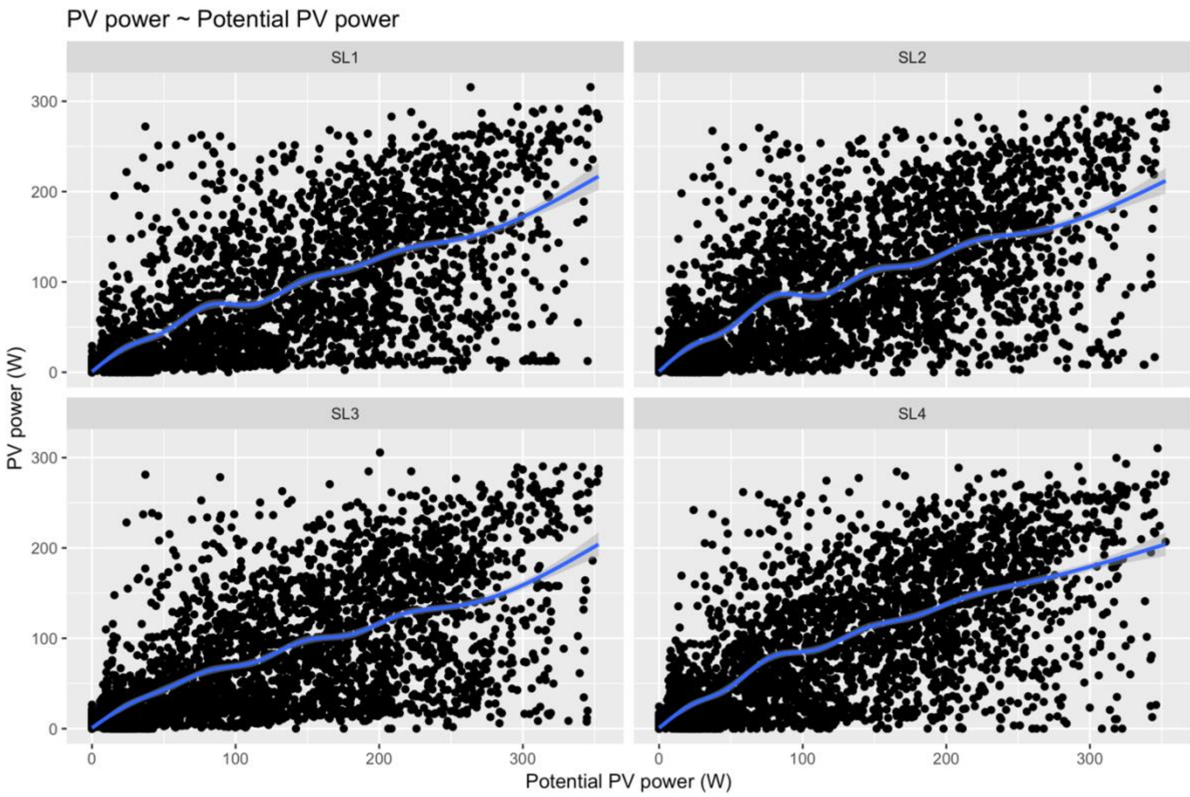


Fig. 3a: Correlation between Potential PV power and PV power. The blue curve shows the gam model

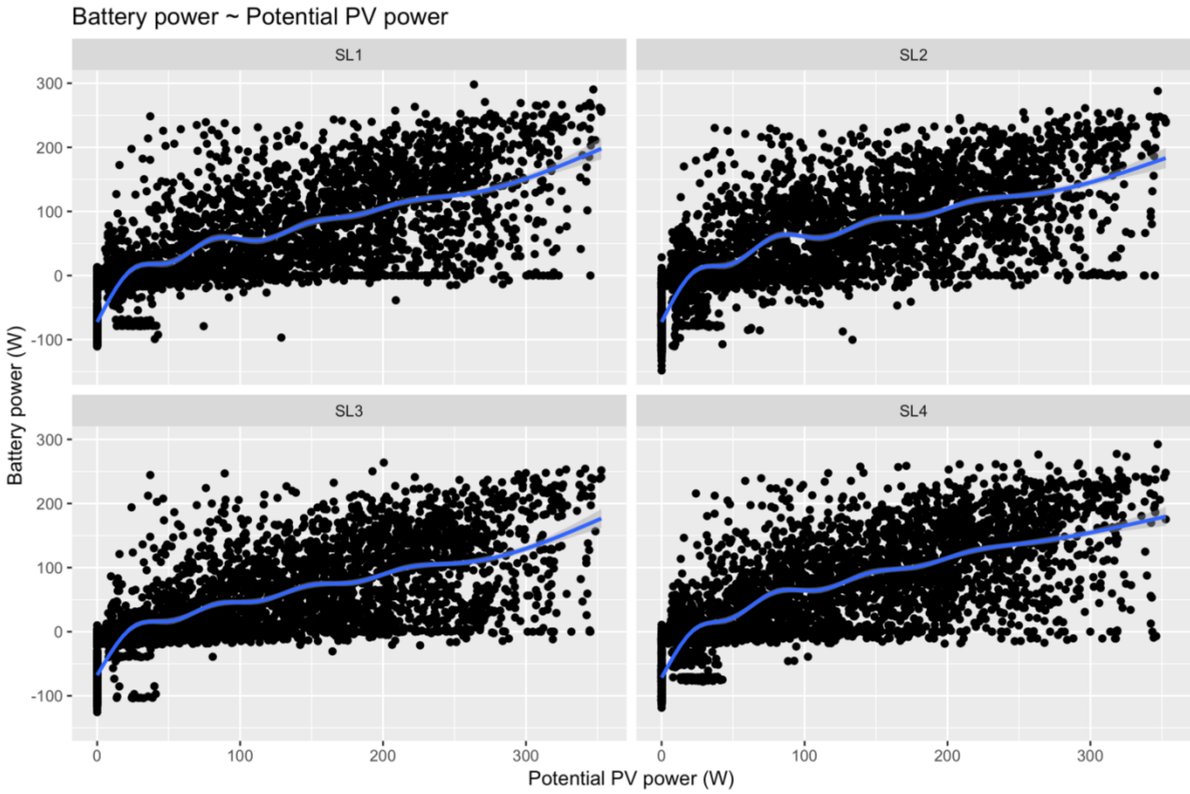
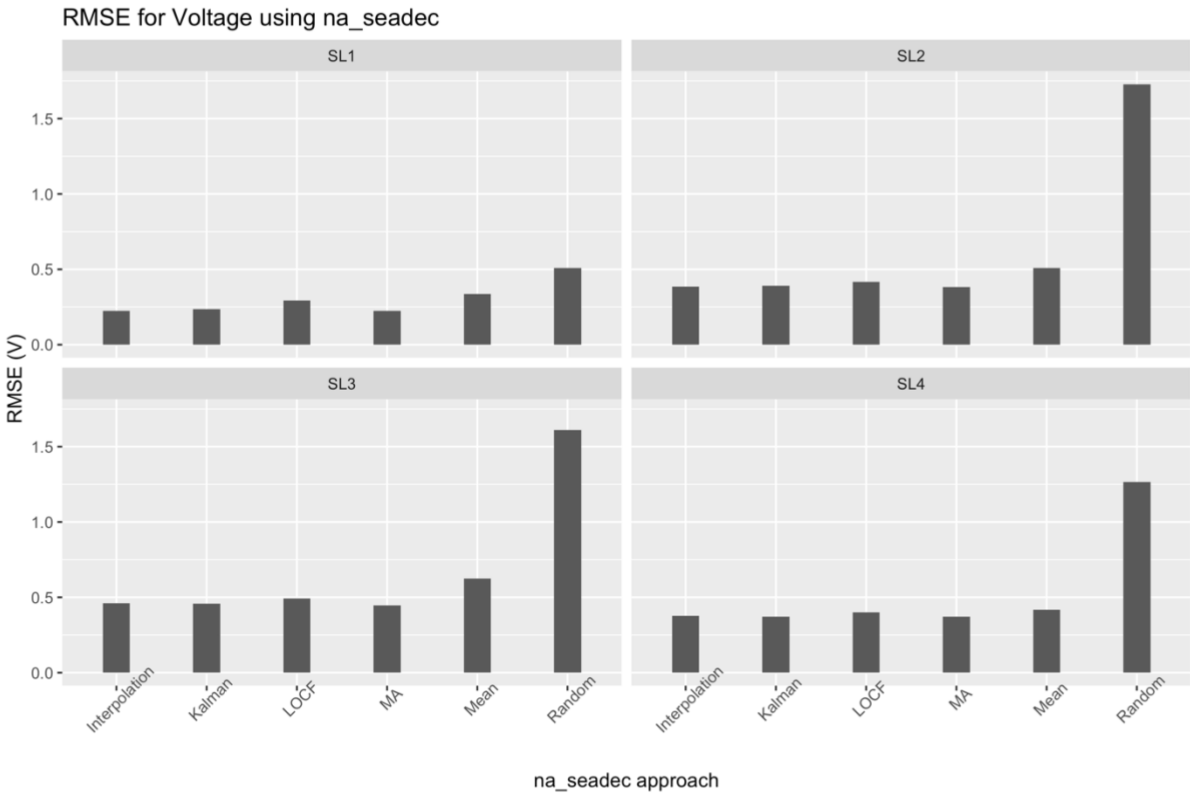


Fig. 3b: Correlation between Potential PV power and Battery power. The blue curve shows the gam model

8. Approach to test the imputation techniques: a complete hourly time series was taken using full data days such that a finite value is available for all variables. Next, we created missing values in this time series. Based on the % of missing data above, 20% data (a fifth of the rows) was selected at random and replaced with NA values. The imputation technique was then applied to get an imputed dataset. Finally, the performance was measured by calculating RMSE between actual test values and imputed values.
9. Testing seasonal decomposition technique for data imputation: Seasonal decomposition is an advanced imputation technique for univariate time series data (employs time dependencies). As the name suggests, it performs a seasonal split/decomposition operation as a pre-processing step, after which imputation is performed on the deseasonalised series. The following algorithms can be used for imputation on the decomposed the time series – interpolation, last observation carried forward, mean imputation, random sample imputation, Kalman smoothing and state space models and weighted moving average. Afterwards the seasonal component of the time series is added again. In [1], the approach is suggested to perform better than simple imputation for time series data as it considers the seasonal/time dependencies.
   Seasonal decomposition in R is implemented using the na_seadec function. We used this approach to impute data using each of the above 6 algorithms for the following variables (one at a time) – Battery voltage, PV power, Solar charger battery power, load current, AC consumption and system battery power. Once the values were

imputed, the performance of each imputation method was assessed by comparing the imputed values with the test data set.

10. Testing GAM for PV power and Battery power: The GAM approach is used to fit linear or non- linear models to data. Based on the correlation between PV power and Battery power with Potential PV power values, we use both linear and non-linear (using spline-based smooths) GAM models to predict the artificially created missing values. The modelling, hence, predicts the PV power and battery power based on the potential PV for that day.
11. Performance evaluation: The performance of the 2 imputation techniques was evaluated using RMSE values (by comparing only the test data set with imputed values) and by comparing the summary values for test and imputed data (mean, median, standard deviation, skewness and kurtosis). The RMSE values obtained using the na_seadec and GAM approaches are shown in figs. 4a and 4b respectively. As can be seen the RMSE is very low for the voltage, current and AC consumption values. Moreover, the RMSE is comparable for PV power and battery power for the 2 imputation techniques. Note that the RMSE values are not calculated for the entire data set but only the test data set wherein the values are imputed. Furthermore, the summary statistics for the test and imputed data is calculated to assess how closely the imputed values map the behaviour of original data. This is shown in tables 1a and 1b. As is seen, the na_seadec approach performs better than the GAM approach and is, therefore, used for imputing missing values for all variables.



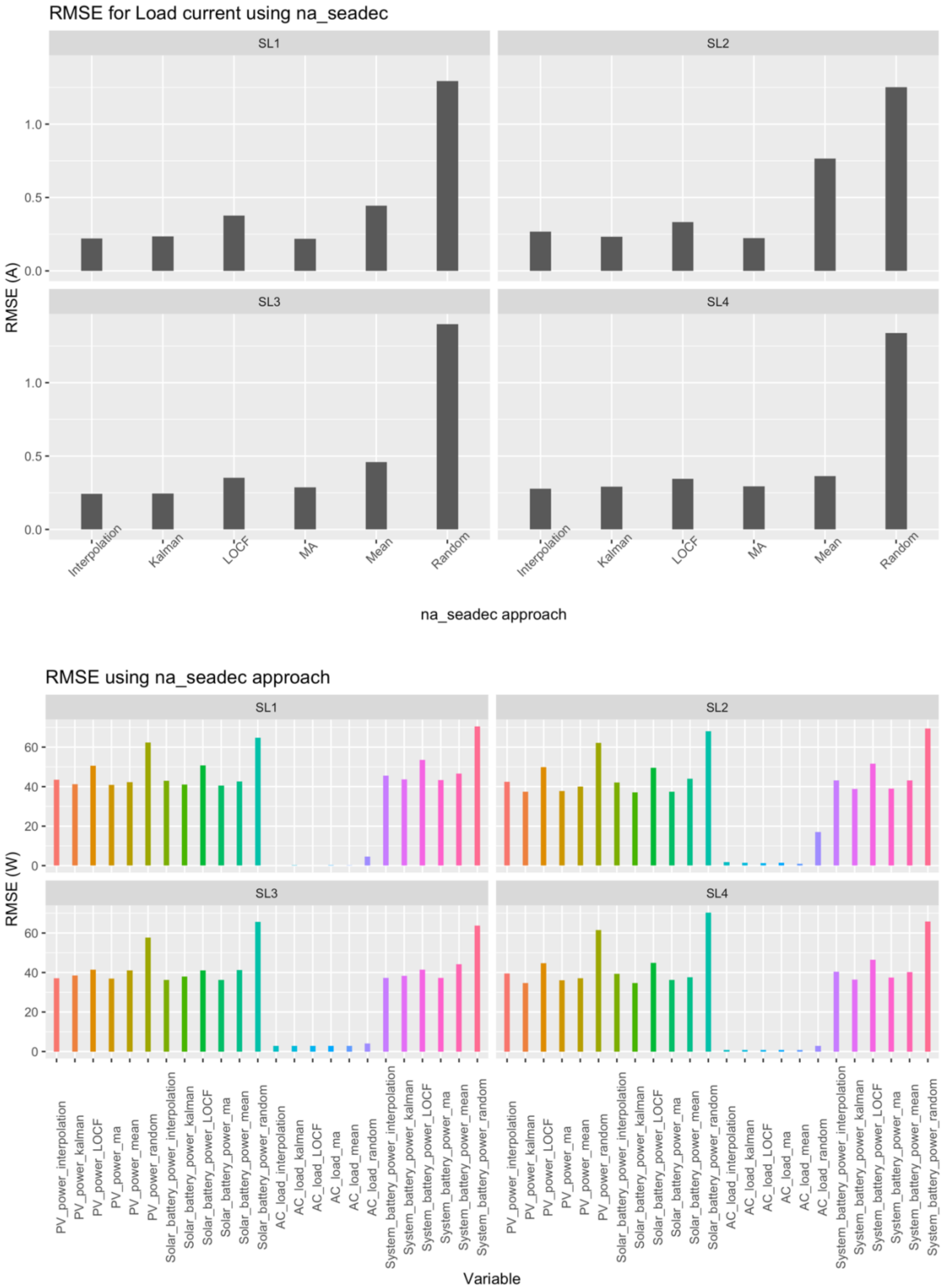


Fig. 4a: RMSE of predicting variables using na_seadec approach.

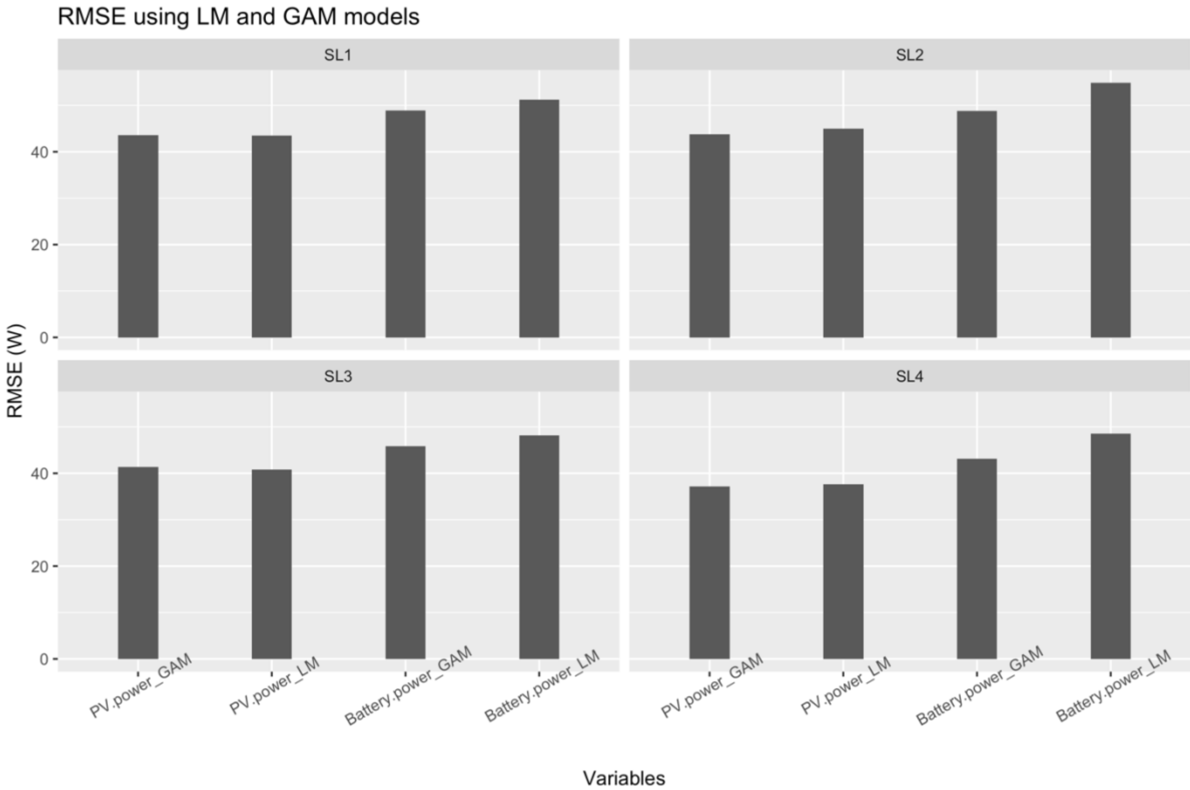


Fig. 4b: RMSE of predicting PV power and Battery power using modelling.

Table 1a: Statistical summary of original and imputed values using na_seadec approach

SL	id	mean	median	sd	skew	kurt
SL1	Battery.Voltage.V	26.43	26.39	0.57	0.23	-0.54
SL1	Battery.Voltage.V_interpolation	26.43	26.39	0.53	0.18	-0.83
SL1	Battery.Voltage.V_kalman	26.43	26.37	0.5	0.14	-1.02
SL1	Battery.Voltage.V_locf	26.41	26.37	0.57	0.18	-0.69
SL1	Battery.Voltage.V_ma	26.43	26.36	0.51	0.18	-0.96
SL1	Battery.Voltage.V_mean	26.42	26.28	0.44	0.38	-1.54
SL1	Battery.Voltage.V_random	26.41	26.4	0.61	-0.38	0.61
SL1	PV.power.W	48.71	2.89	77.7	1.58	1.21
SL1	PV.power.W_interpolation	48.69	6.34	74.23	1.43	0.72
SL1	PV.power.W_kalman	46.84	7.2	65.79	1.19	-0.02
SL1	PV.power.W_locf	48	4.81	77.49	1.63	1.39
SL1	PV.power.W_ma	47.39	6.54	69.35	1.32	0.41
SL1	PV.power.W_mean	47.49	5.49	62.57	0.88	-0.96
SL1	PV.power.W_random	48.93	5.43	74.93	1.56	1.38
SL1	Solar. Battery.watts.W	43.59	0	77.26	1.44	1.04
SL1	Solar.Battery.watts.W_interpolation	43.34	0.31	74.02	1.27	0.5
SL1	Solar.Battery.watts.W_kalman	41.53	1.46	66.23	1.03	-0.18
SL1	Solar.Battery.watts.W_locf	42.22	1.26	78.28	1.41	1.1
SL1	Solar.Battery.watts.W_ma	42.21	0.83	69.21	1.16	0.19
SL1	Solar.Battery.watts.W_mean	41.73	0.12	63.12	0.85	-1
SL1	Solar.Battery.watts.W_random	21.38	-3.93	82.71	1.03	0.49
SL1	Load.current.A	0.14	0	0.43	4.95	25.41

SL1	Load.current.A_interpolation	0.15	0	0.43	4.56	22.26
SL1	Load.current.A_kalman	0.15	0	0.4	4.29	19.31
SL1	Load.current.A_locf	0.17	0	0.54	4.75	23.5
SL1	Load.current.A_ma	0.15	0	0.4	4.24	18.6
SL1	Load.current.A_mean	0.18	0.15	0.09	0.44	-1.33
SL1	Load.current.A_random	0.98	0.55	0.93	1.13	0.27
SL1	AC.Consumption.L1.W_kalman	0.02	0	0.2	10.99	127.2
SL1	AC.Consumption.L1.W_ma	0.01	0	0.17	16.1	276.33
SL1	AC.Consumption.L1.W_mean	0.03	0	0.11	4.37	18.26
SL1	AC.Consumption.L1.W_random	0.95	0	4.55	6.64	47.64
SL1	System.Battery.Power.W	4.32	-13.27	96.8	0.95	-0.01
SL1	System.Battery.Power.W_interpolation	3.66	-13.27	93.29	0.84	-0.31
SL1	System.Battery.Power.W_kalman	1.45	-20.2	86.47	0.64	-0.8
SL1	System.Battery.Power.W_locf	2.72	-13.26	96.44	0.96	0.05
SL1	System.Battery.Power.W_ma	2.2	-15.4	89.26	0.74	-0.54
SL1	System.Battery.Power.W_mean	1.75	-22.17	83.45	0.53	-1.28
SL1	System.Battery.Power.W_random	7.02	-20.69	87.05	1.04	0.35
SL2	Battery.Voltage.V	26.3	26.33	0.66	-1.6	8.98
SL2	Battery.Voltage.V_interpolation	26.31	26.32	0.56	-0.57	2.14
SL2	Battery.Voltage.V_kalman	26.3	26.28	0.52	-0.34	0.88
SL2	Battery.Voltage.V_locf	26.32	26.32	0.56	-0.4	1.17
SL2	Battery.Voltage.V_ma	26.31	26.31	0.53	-0.45	1.71
SL2	Battery.Voltage.V_mean	26.31	26.1	0.42	0.28	-1.54
SL2	Battery.Voltage.V_random	25.6	25.99	1.68	-1.84	3.76
SL2	PV.power.W	52.41	2.51	75.18	1.3	0.44
SL2	PV.power.W_interpolation	50.71	5.6	70.43	1.23	0.31
SL2	PV.power.W_kalman	52.22	6.28	65.57	0.81	-0.92
SL2	PV.power.W_locf	51.01	4.09	73.88	1.35	0.65
SL2	PV.power.W_ma	51.07	5.21	67.26	1.03	-0.26
SL2	PV.power.W_mean	53.08	8.03	64.99	0.71	-1.2
SL2	PV.power.W_random	52.5	11.9	73.58	1.4	0.92
SL2	Solar.Battery.watts.W	44.48	0.55	80.45	0.93	0.19
SL2	Solar.Battery.watts.W_interpolation	42.87	1.12	75.99	0.84	0.06
SL2	Solar.Battery.watts.W_kalman	44.35	1.97	71.76	0.44	-0.82
SL2	Solar.Battery.watts.W_locf	42.82	0.82	79.46	0.95	0.31
SL2	Solar.Battery.watts.W_ma	43.23	0.85	73.13	0.64	-0.35
SL2	Solar.Battery.watts.W_mean	46.02	7.72	68.16	0.65	-1.27
SL2	Solar.Battery.watts.W_random	35.29	1.99	91.66	0.89	-0.19
SL2	Load.current.A	0.26	0	0.81	3.34	10.22
SL2	Load.current.A_interpolation	0.26	0	0.79	3.23	9.6
SL2	Load.current.A_kalman	0.26	0	0.78	3.26	9.84
SL2	Load.current.A_locf	0.27	0	0.83	3.18	9.18
SL2	Load.current.A_ma	0.26	0	0.78	3.26	9.81
SL2	Load.current.A_mean	0.22	0.23	0.22	0.5	-1.12
SL2	Load.current.A_random	0.82	0.27	0.97	1.28	0.81
SL2	AC.Consumption.L1.W	0.09	0	0.93	14.06	224.12

SL2	AC.Consumption.L1.W_interpolation	0.14	0	1.58	16.35	303.31
SL2	AC.Consumption.L1.W_kalman	0.22	0.04	1.1	19.41	456.06
SL2	AC.Consumption.L1.W_locf	0.05	0	0.79	19.09	393.17
SL2	AC.Consumption.L1.W_ma	0.17	0	1.23	12.72	196.84
SL2	AC.Consumption.L1.W_mean	0.23	0.12	0.27	1.36	1.03
SL2	AC.Consumption.L1.W_random	9.59	4.18	14.28	2.23	4.97
SL2	System.Battery.Power.W	-0.03	-14.03	95.27	0.72	-0.55
SL2	System.Battery.Power.W_interpolation	-0.71	-15.49	90.8	0.63	-0.67
SL2	System.Battery.Power.W_kalman	0.56	-19.74	87.46	0.38	-1.27
SL2	System.Battery.Power.W_locf	-0.39	-13.32	94.38	0.72	-0.45
SL2	System.Battery.Power.W_ma	-0.45	-18.33	88.38	0.5	-0.98
SL2	System.Battery.Power.W_mean	2	-27.06	85.43	0.39	-1.43
SL2	System.Battery.Power.W_random	5.41	-22.09	97.37	0.79	-0.25
SL3	Battery.Voltage.V	26.12	26.2	0.75	-2.33	16.22
SL3	Battery.Voltage.V_interpolation	26.14	26.18	0.62	-0.81	2.66
SL3	Battery.Voltage.V_kalman	26.14	26.18	0.55	-0.17	-0.55
SL3	Battery.Voltage.V_locf	26.14	26.2	0.65	-0.89	2.04
SL3	Battery.Voltage.V_ma	26.14	26.16	0.58	-0.32	0.09
SL3	Battery.Voltage.V_mean	26.16	25.99	0.39	0.42	-1.16
SL3	Battery.Voltage.V_random	25.6	25.85	1.58	-2.13	6.08
SL3	PV.power.W	37.8	0	67.04	1.95	2.84
SL3	PV.power.W_interpolation	40.5	0	64.23	1.53	1.19
SL3	PV.power.W_kalman	40.03	0.1	58.52	1.18	-0.11
SL3	PV.power.W_locf	39.44	0	67.17	1.75	1.94
SL3	PV.power.W_ma	39.8	0.02	60.42	1.37	0.56
SL3	PV.power.W_mean	39.14	0.06	55.27	1.04	-0.59
SL3	PV.power.W_random	38.02	0.48	63.5	1.76	2.14
SL3	Solar.Battery.watts.W	26.73	-5	66.09	1.76	2.64
SL3	Solar.Battery.watts.W_interpolation	29.91	-4.99	62.51	1.39	0.97
SL3	Solar.Battery.watts.W_kalman	29.05	-4.87	56.78	1.04	-0.13
SL3	Solar.Battery.watts.W_locf	28.95	-4.95	66.06	1.56	1.64
SL3	Solar.Battery.watts.W_ma	28.92	-5	58.67	1.25	0.39
SL3	Solar.Battery.watts.W_mean	27.79	-8.68	53.59	1.01	-0.65
SL3	Solar.Battery.watts.W_random	10.97	-10.15	80.82	1.03	0.32
SL3	Load.current.A	0.34	0.3	0.46	5.96	39.53
SL3	Load.current.A_interpolation	0.33	0.3	0.35	6.47	52.73
SL3	Load.current.A_kalman	0.33	0.29	0.34	6.61	54.51
SL3	Load.current.A_locf	0.33	0.3	0.41	6.39	48.65
SL3	Load.current.A_ma	0.33	0.29	0.31	5.26	34.33
SL3	Load.current.A_mean	0.35	0.36	0.05	-0.28	-0.56
SL3	Load.current.A_random	1.14	0.59	1.07	0.96	-0.32
SL3	AC.Consumption.L1.W	0.16	0	2.83	18.57	360.5
SL3	AC.Consumption.L1.W_interpolation	0.02	0	0.4	23.15	535.01
SL3	AC.Consumption.L1.W_kalman	0.02	0	0.09	6.01	40.73
SL3	AC.Consumption.L1.W_ma	0.01	0	0.22	22.79	523.16
SL3	AC.Consumption.L1.W_mean	0.02	0	0.05	3.02	7.4

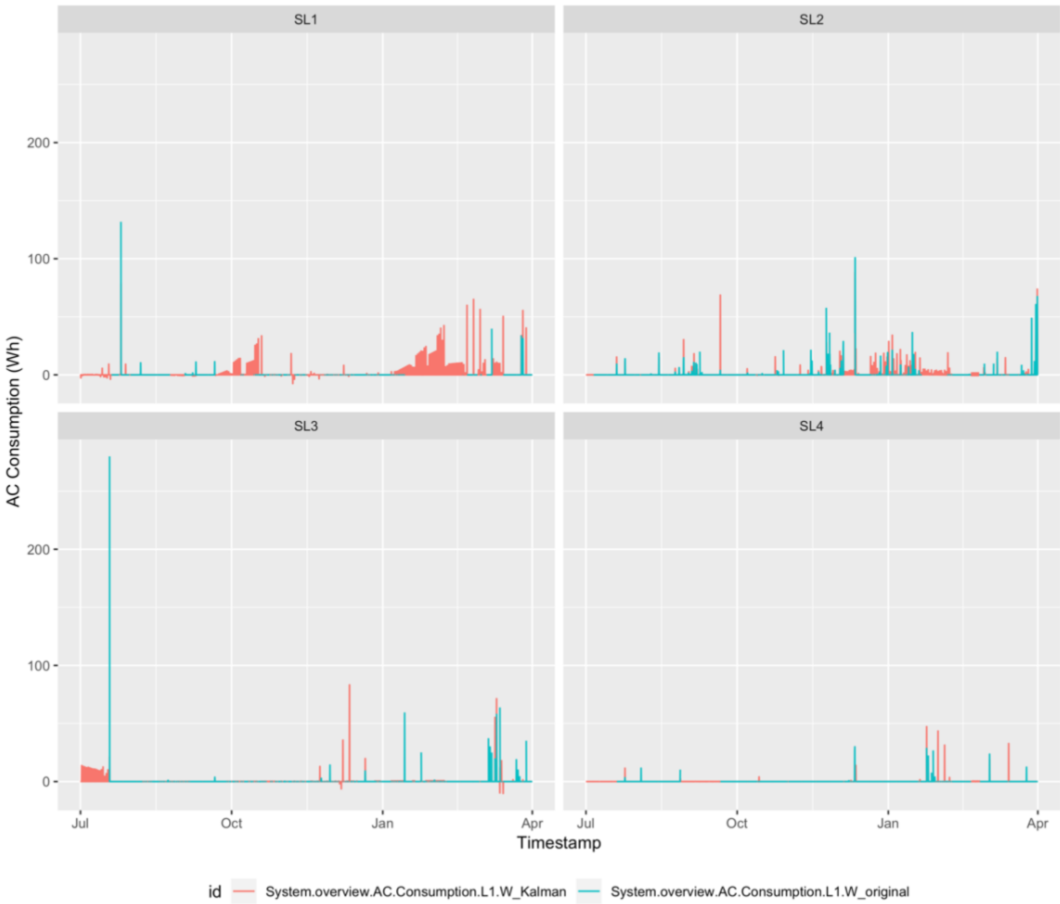
SL3	AC.Consumption.L1.W_random	0.81	0	2.95	4.05	15.83
SL3	System.Battery.Power.W	-9.1	-13.32	82.04	1.11	0.69
SL3	System.Battery.Power.W_interpolation	-6.14	-15.62	79.82	0.82	-0.25
SL3	System.Battery.Power.W_kalman	-7.07	-23.05	75.51	0.62	-0.87
SL3	System.Battery.Power.W_locf	-7.6	-13.37	82.62	0.97	0.18
SL3	System.Battery.Power.W_ma	-7.22	-23.29	76.74	0.74	-0.56
SL3	System.Battery.Power.W_mean	-7.33	-51.1	70.45	0.72	-1.05
SL3	System.Battery.Power.W_random	-7.32	-22.86	76.22	0.87	0.12
SL4	Battery.Voltage.V	26.42	26.38	0.55	-5.41	73.71
SL4	Battery.Voltage.V_interpolation	26.43	26.37	0.42	0	0.59
SL4	Battery.Voltage.V_kalman	26.43	26.36	0.39	0.24	-0.83
SL4	Battery.Voltage.V_locf	26.42	26.37	0.45	-0.62	5.24
SL4	Battery.Voltage.V_ma	26.43	26.37	0.4	0.1	-0.37
SL4	Battery.Voltage.V_mean	26.43	26.22	0.35	0.44	-1.5
SL4	Battery.Voltage.V_random	26.01	26.29	1.29	-2.86	12.08
SL4	PV.power.W	46.03	0.05	70.74	1.44	0.87
SL4	PV.power.W_interpolation	45.77	0.28	66.87	1.35	0.68
SL4	PV.power.W_kalman	46.27	0.35	62.2	0.97	-0.53
SL4	PV.power.W_locf	44.68	0.04	69.59	1.51	1.11
SL4	PV.power.W_ma	45.49	0.34	63.59	1.17	0.1
SL4	PV.power.W_mean	46	1.69	60.29	0.86	-0.93
SL4	PV.power.W_random	45.46	2.38	69.57	1.58	1.54
SL4	Solar.Battery.watts.W	40.38	0	69.91	1.34	0.85
SL4	Solar.Battery.watts.W_interpolation	40.04	0.01	65.68	1.28	0.62
SL4	Solar.Battery.watts.W_kalman	40.46	0.13	61.25	0.9	-0.54
SL4	Solar.Battery.watts.W_locf	38.91	0.02	68.64	1.4	1.04
SL4	Solar.Battery.watts.W_ma	39.82	0.03	62.54	1.11	0.04
SL4	Solar.Battery.watts.W_mean	40.13	-3.61	59.25	0.86	-0.94
SL4	Solar.Battery.watts.W_random	22.43	-8.09	87.19	0.93	-0.04
SL4	Load.current.A	0.15	0.1	0.36	7.39	61.76
SL4	Load.current.A_interpolation	0.15	0.1	0.29	6.35	50.87
SL4	Load.current.A_kalman	0.15	0.1	0.29	7.22	72.41
SL4	Load.current.A_locf	0.15	0.1	0.36	7.46	62.91
SL4	Load.current.A_ma	0.15	0.1	0.26	5.67	40.85
SL4	Load.current.A_mean	0.16	0.15	0.02	0.6	0.71
SL4	Load.current.A_random	1.02	0.6	0.97	1.01	-0.01
SL4	AC.Consumption.L1.W	0.04	0	0.86	24.49	628.85
SL4	AC.Consumption.L1.W_interpolation	0	0	0.08	24.33	620.05
SL4	AC.Consumption.L1.W_kalman	0.01	0	0.04	5.44	31.08
SL4	AC.Consumption.L1.W_ma	0	0	0.07	16.63	295.81
SL4	AC.Consumption.L1.W_mean	0.01	0	0.03	2.63	5.72
SL4	AC.Consumption.L1.W_random	0.82	0	2.59	3.79	14.93
SL4	System.Battery.Power.W	-3.71	-27.14	92.53	0.82	-0.37
SL4	System.Battery.Power.W_interpolation	-4.25	-35.65	89.14	0.74	-0.56
SL4	System.Battery.Power.W_kalman	-3.77	-37.49	85.76	0.51	-1.15
SL4	System.Battery.Power.W_locf	-5.3	-30.03	91.31	0.84	-0.3

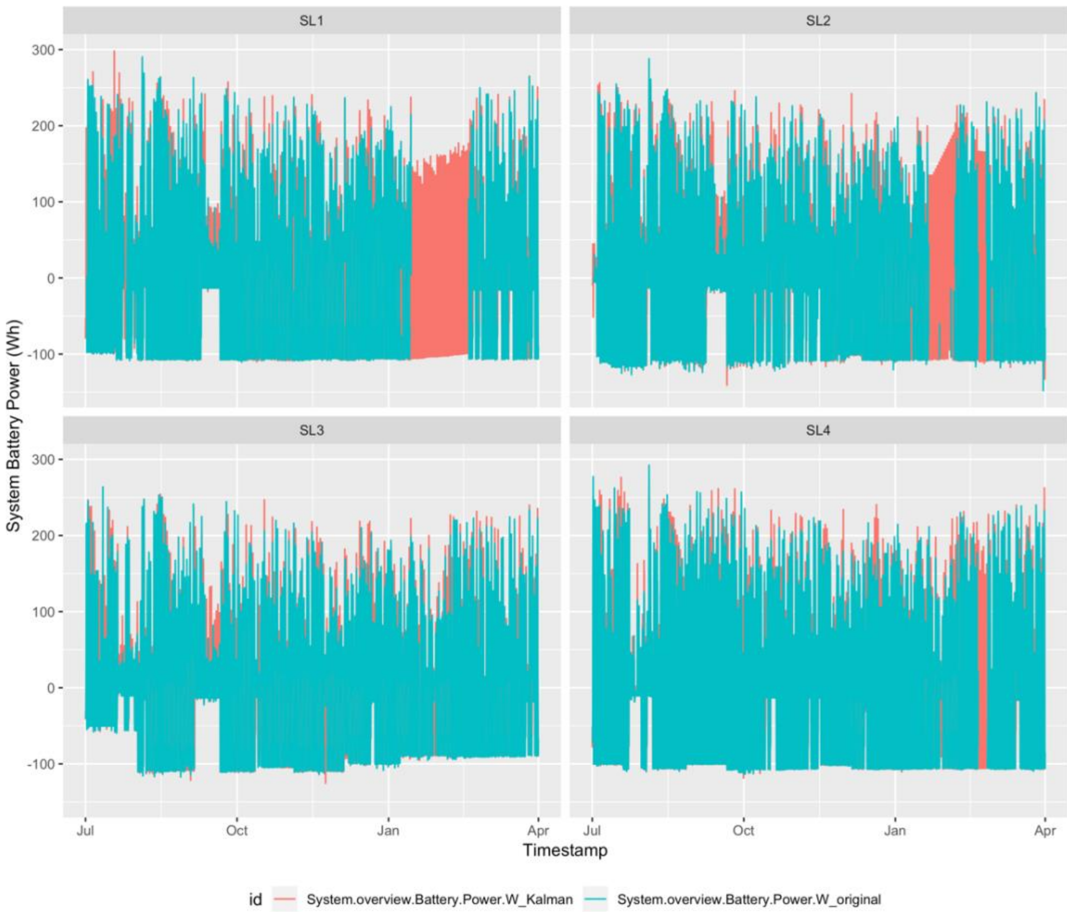
SL4	System.Battery.Power.W_ma	-4.5	-36.12	86.53	0.62	-0.87
SL4	System.Battery.Power.W_mean	-3.92	-55.58	83.24	0.52	-1.33
SL4	System.Battery.Power.W_random	5.23	-20.51	87.86	1.02	0.24

Table 1b: Statistical summary of original and imputed values using GAM approach

SL	id	mean	median	sd	skew	kurt
SL1	PV.power.W	48.71	2.89	77.7	1.58	1.21
SL1	PV.power.W_mod_gam	47	0.88	59.89	0.94	-0.52
SL1	PV.power.W_mod_lm	48.12	4.52	59.72	1.06	-0.3
SL1	System.Battery.Power.W	4.32	-13.27	96.8	0.95	-0.01
SL1	System.Battery.Power.W_mod_gam	0.34	-69.09	81.31	0.62	-1.17
SL1	System.Battery.Power.W_mod_lm	2.74	-55.61	79.91	1.06	-0.3
SL2	PV.power.W	52.41	2.51	75.18	1.3	0.44
SL2	PV.power.W_mod_gam	52.46	10.22	62.51	0.75	-0.97
SL2	PV.power.W_mod_lm	52.57	10.77	61.89	1.06	-0.31
SL2	System.Battery.Power.W	-0.03	-14.03	95.27	0.72	-0.55
SL2	System.Battery.Power.W_mod_gam	1.43	-45.8	81.86	0.49	-1.38
SL2	System.Battery.Power.W_mod_lm	1.42	-51.81	78.81	1.06	-0.31
SL3	PV.power.W	37.8	0	67.04	1.95	2.84
SL3	PV.power.W_mod_gam	38.11	0.79	52.51	1.06	-0.46
SL3	PV.power.W_mod_lm	38.58	3.99	52	1.32	0.44
SL3	System.Battery.Power.W	-9.1	-13.32	82.04	1.11	0.69
SL3	System.Battery.Power.W_mod_gam	-7.64	-60.76	67.13	0.78	-0.99
SL3	System.Battery.Power.W_mod_lm	-6.54	-49.49	64.55	1.32	0.44
SL4	PV.power.W	46.03	0.05	70.74	1.44	0.87
SL4	PV.power.W_mod_gam	46.82	0.75	60.11	0.89	-0.72
SL4	PV.power.W_mod_lm	47	6.05	58.74	1.25	0.33
SL4	System.Battery.Power.W	-3.71	-27.14	92.53	0.82	-0.37
SL4	System Battery.Power.W_mod_gam	-3.07	-72.23	82.03	0.6	-1.27
SL4	System Battery.Power.W_mod_lm	-2.61	-56.23	76.95	1.25	0.33

12. Using seasonal decomposition for imputation – Based on the above performance analysis, we chose the seasonal decomposition approach using Kalman imputation. For imputing the values, we considered the entire hourly data set and performed univariate imputation for each variable. The imputed time series for PV power, battery power and AC consumption is shown in fig. 5 below. As can be seen, the imputation of values is fairly good using the approach.


SL1
SL2
SL3
SL4
AC Consumption (Wh)
200
100
0
Jul
Oct
Jan
Apr
Timestamp
id
System.overview.AC.Consumption.L1.W_Kalman
System.overview.AC.Consumption.L1.W_original


SL1
SL2
SL3
SL4
System Battery Power (Wh)
300
200
100
0
-100
Jul
Oct
Jan
Apr
Timestamp
id
System.overview.Battery.Power.W_Kalman
System.overview.Battery.Power.W_original

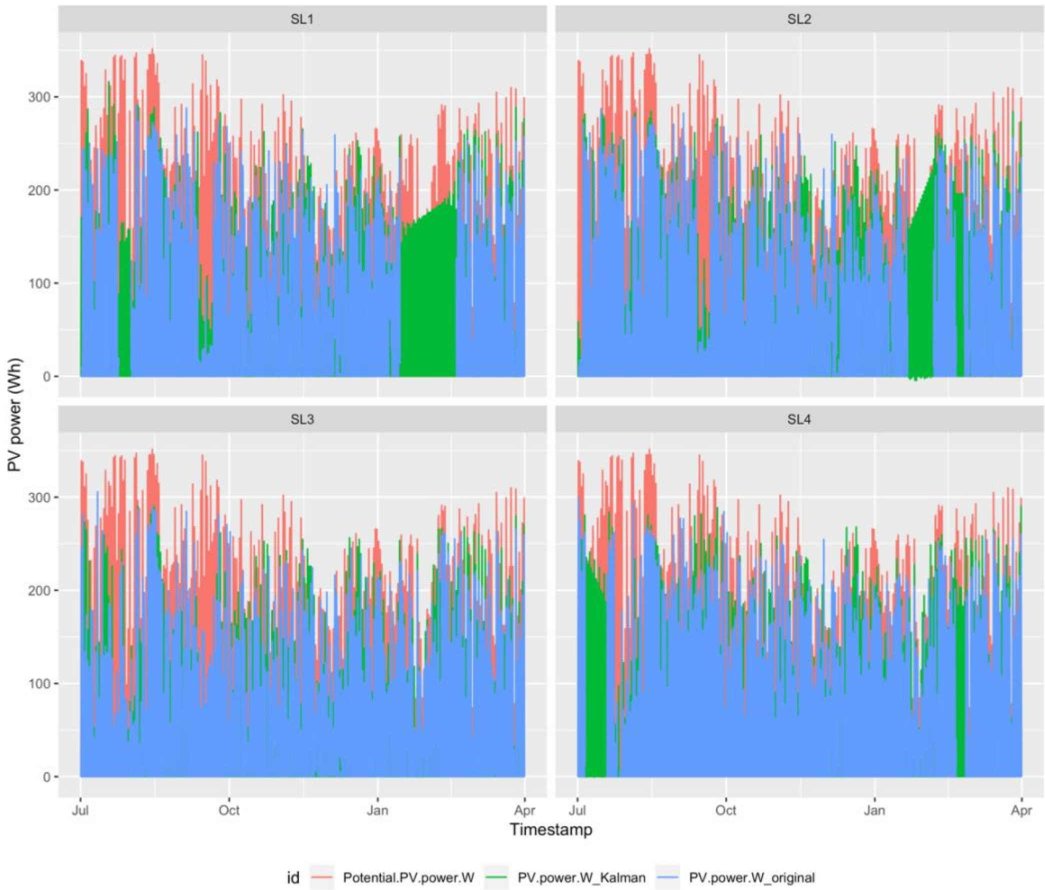


Fig. 5: Imputed time series data

13. Pre-processing data after imputation: Once the imputation is performed, we apply certain rule-based corrections to the dataset. This is done because the imputation does not consider power outage events as these are missing from the original dataset and results in certain obvious errors owing to smoothing. Accordingly, the below corrections are made.
    a. If load current or AC consumption values are <0, the values are changed to 0.
    b. State of charge is calculated using the system battery power values. We assume that the battery is fully charged at the outset i.e. 1st of July. The battery state at the 0th hour on 1st July is assumed to be 3072Wh and is then updated sequentially by adding/subtracting the hourly battery power.
    c. The system is expected to cut-off and cause a power outage if the battery state of charge drops to 20% of capacity i.e. 614Wh. For all such instances, the values for actual PV power, Solar Battery power, AC consumption and Battery power is considered 0 (due to power outage) if they were missing from the original data set.
14. Once these corrections are made, the light load is calculated using the negative battery power values and data is analysed to obtain average monthly and daily consumption.

References:

1. https://journal.r-project.org/archive/2017/RJ-2017-009/RJ-2017-009.pdf