Evaluation of the 2019 Public Service Company of New Mexico Energy Efficiency and Demand Response Programs

FINAL Report - Appendices

April 13, 2020

Table of Contents

APPENDIX A – COMMERCIAL COMPREHENSIVE PARTICIPANT SURVEY INSTRUMENT 1

APPENDIX B – CONTRACTOR INTERVIEW GUIDE 22

APPENDIX C – POWER SAVER DETAILED EVALUATION METHODS AND FINDINGS 26

APPENDIX D – PEAK SAVER DETAILED EVALUATION METHODS AND FINDINGS 74

APPENDIX E – COMMERCIAL COMPREHENSIVE DESK REVIEW RESULTS SUMMARY 97

Appendix A – Commercial Comprehensive Participant Survey Instrument

Hello, my name is (*YOUR NAME*) from Research & Polling, Inc. I am calling on behalf of PNM. May I please speak with ________________?

A. (Once correct respondent is reached) Hello, my name is (*YOUR NAME*) from Research & Polling, Inc. I am calling on behalf of PNM.

I'm calling because our records show that you recently completed an energy efficiency project where you installed [MEASURE_1] at your business located at [SITE_ADDRESS] and received a rebate through the PNM [REBATE PROGRAM] program. I'd like to ask a short set of questions about your experience with the [REBATE PROGRAM] program. Your time will help us improve this program for other customers like you. Are you the best person to talk to about the/these energy efficiency upgrade(s) and energy use at your firm?

1. Yes
2. No (Ask, Who would be the best person to talk to about the [MEASURE(S)] installed and energy use at your business? (REPEAT INTRO WHEN CORRECT PERSON COMES ON LINE; ARRANGE CALLBACK IF NECESSARY)
3. Never installed *(VOLUNTEERED SKIP TO Q.5)*

(IF NEEDED) PNM would like to better understand how businesses like yours think about and manage their energy use. The [REBATE_PROGRAM] program is designed to help firms with energy saving efforts. Your input is very important to help PNM improve its energy rebate programs.

<u>SECTION A [MEASURE _1]</u>

1. (A 1) Our records show in 2019 your business got a rebate through PNM for installing [MEASURE_1]. Are you familiar with this project?

1. Yes
2. No *(SKIP TO Q.2)*
3. Never installed *(VOLUNTEERED) (SKIP TO Q.5)*
4. Don't know *(SKIP TO Q.2)*

1a. Our records show it was installed at [SITE_ADDRESS] in [SITE_CITY]. Is that correct?

1. Yes *(SKIP TO Q. 3)*
2. No *(GO TO Q. 1b)*
3. Never installed *(VOLUNTEERED) (SKIP TO Q.5)*
4. Don't know *(SKIP TO Q.2)*

1b. Where was [MEASURE_1] installed? *(RECORD LOCATION)*

____*(SKIP TO Q. 3)*

99. Never installed *(SKIP TO Q. 5)*

2. (A 1a) Is there someone else in your company who would know about buying the [MEASURE_1]?

1. Yes (Ask to be transferred to better contact and go back to intro)
2. Yes (Unable to be transferred, record contact's and number to call back)
3. No **(THANK AND TERMINATE)**
4. Don't know **(THANK AND TERMINATE)**

3. (A 2) Thinking about the [MEASURE_1] for which you received a rebate, is the [MEASURE_1] still installed in your facility?

1. Yes *(SKIP TO Q. 6)*
2. No
3. Prefer not to answer *(SKIP TO Q. 6)*
4. Don't know *(SKIP TO Q. 6)*

4a. (A 3) Was the [MEASURE_1] removed?

01. Yes, it was removed *(SKIP TO Q.5)*
02 No *(CONTINUE TO Q.4b)*
03. Prefer not to answer *(DO NOT READ) (SKIP TO Q.7)*
99. Don't know *(DO NOT READ) (SKIP TO Q.7)*

Other *(SPECIFY)* ____

4b. (A 3) Was the [MEASURE_1] never installed?

01. Yes, never installed
02. Prefer not to answer *(DO NOT READ) (SKIP TO Q.7)*
99. Don't know *(DO NOT READ) (SKIP TO Q.7)*

Other *(SPECIFY)* ____

5. (A3a) Why was the [MEASURE_1] removed/never installed? *(OPEN VERBATIM)*

____

____

____

(SKIP TO SECTION A [MEASURE_2]

6. (A 4) Is the [MEASURE_1] still functioning as intended?

1. Yes

2. No
3. Prefer not to answer *(DO NOT READ)*
4. Don't know *(DO NOT READ)*

7. (A 5) Did your firm use a contractor to install the [MEASURE_1] or did internal staff do the work?

01. Contractor *(SKIP TO SECTION A [MEASURE_2])*
02. Internal Staff
03. Prefer not to answer *(SKIP TO SECTION A [MEASURE_2])*
99. Don't know *(SKIP TO SECTION A [MEASURE_2])*

Other *(SPECIFY)*____
(SKIP TO SECTION A [MEASURE_2])

8. (A 6) Why did your firm choose to use internal staff instead of a contractor?

____

____

____

98. Prefer not to answer
99. Don't know

SECTION A [MEASURE_2]

1. (A 1) Our records also show in 2019 your business got a rebate through PNM for installing a (MEASURE_2]. Do you remember this?

1. Yes
2. No *(SKIP TO Q.2)*
3. Never installed *(VOLUNTEERED) (SKIP TO Q.5)*
4. Don’t know *(SKIP TO Q.2)*

1a. Our records show it was installed at [SITE_ADDRESS] in [SITE_CITY]. Is that correct?

1. Yes *(SKIP TO Q. 3)*
2. No *(GO TO Q. 1b)*
3. Never installed *(VOLUNTEERED) (SKIP TO Q.5)*
4. Don’t know *(SKIP TO Q.2)*

1b. Where was [MEASURE_2] installed? *(RECORD LOCATION)*

____(SKIP TO Q. 3)*

99. Never installed *(SKIP TO Q. 5)*

2. VACANT

3. (A 2) Thinking about the [MEASURE_2] for which you received a rebate, is the [MEASURE_2] still installed in your facility?

1. Yes *(SKIP TO Q. 6)*
2. No
3. Prefer not to answer *(SKIP TO Q. 6)*
4. Don't know *(SKIP TO Q. 6)*

4a. (A 3) Was the [MEASURE_2] removed?

01. Yes, it was removed *(SKIP TO Q.5)*
02 No *(CONTINUE TO Q.4b)*
03. Prefer not to answer *(DO NOT READ) (SKIP TO Q.7)*
99. Don't know *(DO NOT READ) (SKIP TO Q.7)*

Other *(SPECIFY)* ____

4b. (A 3) Was the [MEASURE_2] never installed?

01. Yes, never installed
02. Prefer not to answer *(DO NOT READ) (SKIP TO Q.7)*
99. Don't know *(DO NOT READ) (SKIP TO Q.7)*

Other *(SPECIFY)* ____

5. (A3a) Why was the [MEASURE_2] removed/never installed? *(OPEN VERBATIM)*

____

____

____

(SKIP TO INTRO TO Q. 10)

6. (A 4) Is the [MEASURE_2] still functioning as intended?

1. Yes
2. No
3. Prefer not to answer *(DO NOT READ)*
4. Don't know *(DO NOT READ)*

7. (A 5) Did your firm use a contractor to install the [MEASURE_2] or did internal staff do the

work?

01. Contractor *(SKIP TO Q. 9)*

02. Internal Staff

03. Prefer not to answer *(SKIP TO Q. 9)*

99. Don't know *(SKIP TO Q. 9)*

Other *(SPECIFY)*____*(SKIP TO Q. 9)*

8. (A 6) Why did your firm choose to use internal staff instead of a contractor?

____

____

____

98. Prefer not to answer

99. Don't know

9. (A 7) Was your [MEASURE_1] AND [MEASURE_2], installed/purchased together as a single project or were these done separately?

1. Together as one project

2 Separately

3. Prefer not to answer *(DO NOT READ)*

4. Don't know *(DO NOT READ)*

<u>SECTION B</u>

Now I have some questions about how your company became aware of the PNM rebate program.

10. (B 1) How did your company FIRST learn about the program?
(DO NOT READ CATEGORIES) (TAKE <u>ONE</u> RESPONSE)

01. Word of mouth (business associate, coworker)
02. Utility program staff

03. Utility website

04. Utility bill insert

05. Utility representative

06. Utility advertising

07. Email from utility

08. Contractor/distributor

09. Building audit or assessment

10. Television Advertisement – Mass Media

11. Other mass media (sign, billboard, newspaper/magazine ad)

12. Event (conference, seminar workshop)

13. Online search, web links

14. Participated or received rebate before

98. No way in particular

99. Don't know

Other *(SPECIFY)* ____

11. (B 2) What other sources did your company use to gather information about the program….Were there any others? *(DO NOT READ CATEGORIES) (TAKE UP TO THREE RESPONSES)*

01. Word of mouth (business associate, co-worker)

02. Utility program staff

03. Utility website

04. Utility bill insert

05. Utility representative

06. Utility advertising

07. Email from utility

08. Contractor/distributor

09. Building audit or assessment

10. Television Advertisement – Mass Media

11. Other mass media (sign, billboard, newspaper/magazine ad)

12. Event (conference, seminar, workshop)

13. Online search, web links

14. Participated or received rebate before

98. None *(SKIP TO POLLER NOTE BEFORE Q. 13)*

99. Don't know *(SKIP TO POLLER NOTE BEFORE Q. 13)*

Other *(SPECIFY)* ____

12. (B 3) Of all the sources you mentioned, which did you find most useful in helping you decide to participate in the program?

____

____

____

97. None in particular
98. Prefer not to answer
99. Don't know

SECTION C

POLLER NOTE:

If Respondent's answer to Q. 9 was:
***Together as one project, prefer not to answer, or don't know* then READ:**

"For the remainder of this survey we will refer to your equipment upgrades collectively as a single project.

If Respondent's answer Q. 9 was:
***Seperately*, READ:**

"For the remainder of this survey we will refer only to the project where you installed [MEASURE_1]

POLLER NOTE: WAS MEASURE INSTALLED?

1. **Yes (GO TO Q. 13a)**
2. **No (GO TO Q. 13b)**

13a. (C 1) Did the equipment that your firm installed replace existing equipment?

1. Yes (i.e. all equipment was replacing old equipment) *(SKIP TO Q. 14a)*
2. Some equipment was a replacement and some was a new addition *(SKIP TO Q. 14a)*
3. No (i.e. all equipment was an addition to existing equipment*) (SKIP TO INTRO TO Q. 17)*
4. Prefer not to answer *(SKIP TO INTRO TO Q. 17)*
5. Don't know *(SKIP TO INTRO TO Q. 17)*

13b. (C 1) Is the equipment that your firm purchased intended to replace existing equipment?

1. Yes (i.e. all equipment is replacing old equipment) *(SKIP TO Q. 14b)*
2. Some equipment is a replacement and some was a new addition *(SKIP TO Q. 14b)*
3. No (i.e. all equipment is an addition to existing equipment) *(SKIP TO INTRO TO Q. 17)*
4. Prefer not to answer *(SKIP TO INTRO TO Q. 17)*
5. Don't know *(SKIP TO INTRO TO Q. 17)*

14a. (C 2) Was the replaced equipment…*(READ CATEGORIES)*

1. Fully functional and not in need of repair? *(SKIP TO Q. 15a)*
2. Functional, but needed minor repairs? *(SKIP TO Q. 15a)*
3. Functional, but needed major repairs? *(SKIP TO Q. 15a)*

4. Not functional? *(SKIP TO INTRO TO Q. 17)*
5. Prefer not to answer *(DO NOT READ) (SKIP TO INTRO TO Q. 17)*
6. Don't know *(DO NOT READ) (SKIP TO INTRO TO Q. 17)*

14b. (C 2) Is the equipment you intend to replace*…(READ CATEGORIES)*

1. Fully functional and not in need of repair? *(SKIP TO Q. 15b)*
2. Functional, but needed minor repairs? *(SKIP TO Q. 15b)*
3. Functional, but needed major repairs? *(SKIP TO Q. 15b)*
4. Not functional? *(SKIP TO INTRO TO Q. 17)*
5. Prefer not to answer *(DO NOT READ) (SKIP TO INTRO TO Q. 17)*
6. Don't know *(DO NOT READ) (SKIP TO INTRO TO Q. 17)*

15a. (C 3) About how old, in years, was the equipment prior to replacement? (Probe if necessary: Best guess is fine.)

____ _____ _____ (Record Years)

499. Prefer not to answer

500. Don't know

ALL ANSWERS TO 15a GO TO Q. 16

15b. (C 3) About how old, in years, is the equipment you are replacing? (Probe if necessary: Best guess is fine.)

_____ _____ _____ (Record Years)

499. Prefer not to answer

500. Don't know

ALL ANSWERS TO 15b. GO TO Q.16

16. (C 2) How much longer (in years) do you think your old equipment would have lasted if you had not replaced it? (Probe if necessary: Best guess is fine.)

1. Less than a year
2. 1 – 2 years
3. 3 – 5 years
4. 6 – 10 years
5. More than 10 years
6. Prefer not to answer

7. Don't know

(C 3a-g) Next I will read a list of reasons your firm may have considered when you decided to conduct your project. For each one, please tell me if it was *not at all important, a little important, somewhat important, very important or extremely important*.

How important was… on your decision to conduct your project?

(RANDOMIZE)	*Extremely Important*	*Very Important*	*Somewhat Important*	*A little Important*	*Not important At All*	*Don't Know/ Won't Say*
17. (C5a) Reducing environmental impact of the business	5	4	3	2	1	6
18. (C5b) Upgrading out-of-date equipment	5	4	3	2	1	6
19. (C5c) Improving comfort at the business	5	4	3	2	1	6

POLLER NOTE: Was HVAC Measure installed?

1. **Yes (CONTINUE TO Q. 20)**
2. **No (SKIP to Q. 21)**

20. (C5d) Improving air quality	5	4	3	2	1	6
21. (C5e) Receiving the rebate	5	4	3	2	1	6
22. (C5f) Reducing energy bill amounts	5	4	3	2	1	6

POLLER NOTE: Did respondent answer Contractor in Q.7?

1. **Yes (CONTINUE TO Q. 23)**
2. **No (SKIP TO INTRO Q. 24)**

23. (C5g) The contractor recommendation	5	4	3	2	1	6

SECTION D (INTRO TO Q.24)

Next, I'm going to ask a few questions about your decision to participate in the program, and choose equipment that was energy efficient

(D 1A-N). I'm going to ask you to rate the importance of each of the following factors on your decision to determine how energy efficient your project would be. Please rate the importance of each of these factors in determining your project's energy efficiency level using a scale from 0 to 10, where 0 means *not at all important* and 10 means *extremely important.* Please let me know if the factor is not applicable.

First I would like to read you some factors related to the rebate program itself.

POLLER NOTE: Did respondent answer Contractor in Q.7?

1. **Yes (CONTINUE TO Q. 24)**
2. **No (CIRCLE [12 N/A] ON Q. 24 AND SKIP TO Q. 25)**

How important was (read below)…in determining how energy efficient your project would be?

(RANDOMIZE)	*Extremely Important*										*Not at all Important*	*DK/ WS*	*N/A*
Program Factors													
24. (D1A) The contractor who performed the work	10	09	08	07	06	05	04	03	02	01	00	11	12
25. (D1B) The dollar amount of the rebate	10	09	08	07	06	05	04	03	02	01	00	11	12
26. (D1C) Technical assistance received from PNM staff	10	09	08	07	06	05	04	03	02	01	00	11	12
27. (D1D) Endorsement or recommendation by your PNM account manager or other PNM staff	10	09	08	07	06	05	04	03	02	01	00	11	12
28. (D1E) Information from PNM marketing or informational materials	10	09	08	07	06	05	04	03	02	01	00	11	12
29. (D1F) Previous participation in a PNM program	10	09	08	07	06	05	04	03	02	01	00	11	12
30. (D1G) Endorsement or recommendation by a contractor	10	09	08	07	06	05	04	03	02	01	00	11	12
31. (D1H) Endorsement or													

recommendation by a vendor or distributor..10 .... 09..... 08..... 07 .....06 .....05 .... 04..... 03.....02 ... 01 ... 00... 11   12

32. (D1I) VACANT

Now, I would like to read you some factors that are <u>not</u> related to the rebate program. Using the same scale from 0 to 10, where 0 means *not at all important* and 10 means *extremely important,* please rate the following non program factors importance in determining your project's energy efficiency.

How important was (read below)…..in determining your project's energy efficiency?

(RANDOMIZE)	*Extremely Important*										*Not at all Important*	*DK/ WS*	*N/A*

<u>Non-program Factors</u>

33. (D1J) The age or condition of the old equipment..................................10 .... 09..... 08..... 07 .....06 .....05 .... 04..... 03.....02 ... 01 ... 00... 11   12

34. (D1K) Corporate policy or guidelines ....10 .... 09..... 08..... 07 .....06 .....05 .... 04..... 03.....02 ... 01 ... 00... 11   12

35. (D1L) Minimizing operating cost………10…..09…..08…..07……06….05….04…03……02….01…00...... 11 12

36. (D1M) Scheduled time for routine maintenance ....10 .... 09..... 08..... 07 .....06 .....05 .... 04..... 03.....02 ... 01 ... 00... 11   12

37. (D2) Of the items I just asked you about, think of the program factors as relating to assistance provided by the utility, such as the rebate, marketing from PNM, recommendation by a contractor and technical assistance from PNM. I also asked you about some non-program factors, which included the age and condition of the old equipment, company policy, operating costs and routine maintenance.

If you had to divide 100% of the influence on your decision to determine how energy efficient your new equipment would be between the PNM program and non-program factors, what percent would you give to the importance of the program factors? [IF NEEDED: Again, these are things like the rebate, marketing from PNM, recommendation by a contractor and technical assistance from PNM]

____ ____ _____ % = Program Factors

499. Prefer not to answer *(SKIP TO Q.39)*

500. Don't know *(SKIP TO Q. 39)*

38. D3. And what percent would you give to the importance of the non-program factors? (IF NEEDED: These include things like the age and condition of the old equipment, company policy, operating costs and routine maintenance.)

_____ _____ _____ %= Non Program Factors

499. Prefer not to answer *(SKIP TO Q.39)*

500. Don't know *(SKIP TO Q.39)*

POLLER NOTE: INSURE ANSWERS TO Q. 37 AND Q. 38 EQUAL 100%

39. (D 5) Did you first learn about the [REBATE_PROGRAM] program BEFORE or AFTER you decided how energy efficient your equipment would be?

1. Before
2. After
3 Prefer not to answer
4. Don't know

40. (D6) Using a scale from 0 to 10, where 0 means *not at all likely* and 10 means *extremely likely*, please rate the likelihood that you would have installed the same equipment with the exact same level of energy efficiency if the [REBATE_PROGRAM] program was not available.

Extremely Likely			*Not at all Likely*	*DK/ WS*
	10 ....09 08	..... 07.......06 ....05 .... 04.........03	02 .... 01............00	 11
	GO TO Q. 41	**SKIP TO Q. 43**	**GO TO Q. 42**	**SKIP TO Q. 43**

POLLER NOTE: IF ANSWER TO Q. 40 IS 8 OR HIGHER AND ANY RESPONSE TO Q. 24-Q.32 IS 8 OR HIGHER, THEN GO TO Q. 41. IF ANSWER TO Q. 40 IS 2 OR LESS AND ANY RESPONSE TO Q.24-Q.32 IS 2 OR LESS THEN GO TO Q. 42.

41. (D7) You just rated your likelihood to install the same equipment without any assistance from the program as a(n) [RATE RESPONSE FROM Q. 40] out of 10. Earlier, when I asked you to rate the importance of each program factor on your decision, the highest rating you gave was a [HIGHEST RATING FROM Q.24-Q.32] out of 10 for the importance of [RE-READ WORDING FOR HIGHEST RESPONSES Q.24-Q.32, PAGE 10].

Can you briefly explain why you were likely to install the equipment without the program but also rated the program factors as highly influential in your decision?
(RECORD VERBATIM)

____

____

____

____

(SKIP TO Q. 43)

42. (D8) You just rated your likelihood to install the same equipment without any assistance from the program as a(n) [RATE RESPONSE FROM Q. 40] out of 10. Earlier, when I asked you to rate the importance of each program factor on your decision, the highest rating you gave was a [LOWEST RATING FROM Q.24-Q.32, Page 10] out of 10.

Can you briefly explain why you said you were not likely to install the equipment without help from the program, yet did not rate the program as highly influential in your decision? *(RECORD VERBATIM)*

____

____

____

____

43. (D 9) If the [REBATE_PROGRAM] program was not available, would you have delayed starting the project to a later date?

1. Yes
2. No *(SKIP TO Q. 46)*
3. Would not have done the project at all *(SKIP TO Q. 46)*
4. Prefer not to answer *(SKIP TO Q. 46)*
5. Don't know *(SKIP TO Q. 46)*

44. (D10) Approximately how much later would you have done the project if the [REBATE_PROGRAM] program was not available? Would it have been…
(READ CATEGORIES)

1. Within one year
2. Between 12 months and less than 2 years *(SKIP TO Q. 46)*
3. Between 2 years and 3 years *(SKIP TO Q. 46)*
4. Greater than 3 years *(SKIP TO Q. 46)*
5 Or would you not have installed the equipment at all *(SKIP TO Q. 46)*
6. Prefer not to answer *(SKIP TO Q. 46)*
7. Don't know *(SKIP TO Q. 46)*

45. (D11) Using a scale from 0 to 10, where 0 means *not at all likely* and 10 means *extremely likely*, please rate the likelihood that you would have conducted this project within 12 months of when you actually completed this project if the [REBATE_PROGRAM] program was not available.

Extremely Likely	*Not at all Likely*	*DK/ WS*

10 .... 09 .... 08 .... 07 .... 06 .... 05 .... 04 .... 03 .... 02 .... 01 ..... 00 11

SECTION E

Now I have some questions about your satisfaction with various aspects of PNM and the [REBATE_PROGRAM] program.

(E 1A-K). For each of the following, please tell me if you were *very dissatisfied, somewhat dissatisfied, neither satisfied nor dissatisfied, somewhat satisfied or very satisfied*.

46. (E1A) PNM as an energy provider

1. Very Dissatisfied
2. Somewhat Dissatisfied
3. Neither Satisfied Nor Dissatisfied *(SKIP TO Q. 48)*
4. Somewhat Satisfied *(SKIP TO Q. 48)*
5. Very Satisfied *(SKIP TO Q. 48)*
6. Not applicable *(SKIP TO Q. 48)*
7. Prefer not to answer *(SKIP TO Q. 48)*
8. Don't know *(SKIP TO Q. 48)*

47. Can you tell me why you gave that rating? *(RECORD VERBATIM)*

____

____

____

48. (E1B) The rebate program overall

1. Very Dissatisfied
2. Somewhat Dissatisfied
3. Neither Satisfied Nor Dissatisfied *(SKIP TO Q.50)*
4. Somewhat Satisfied *(SKIP TO Q.50)*
5. Very Satisfied *(SKIP TO Q.50)*
6. Not applicable *(SKIP TO Q.50)*
7. Prefer not to answer *(SKIP TO Q.50)*
8. Don't know *(SKIP TO Q.50)*

49. Can you tell me why you gave that rating? *(RECORD VERBATIM)*

____

____

50. (E1C) The equipment installed through the program

1. Very Dissatisfied
2. Somewhat Dissatisfied
3. Neither Satisfied Nor Dissatisfied *(SKIP TO Q.52)*
4. Somewhat Satisfied *(SKIP TO Q.52)*
5. Very Satisfied *(SKIP TO Q.52)*
6. Not applicable *(SKIP TO Q.52)*
7. Prefer not to answer *(SKIP TO Q.52)*
8. Don't know *(SKIP TO Q. 52)*

51. Can you tell me why you gave that rating? *(RECORD VERBATIM)*

____

____

____

POLLER NOTE: WAS INSTALLATION DONE BY A CONTRACTOR (Q.7)?

1. **Yes (CONTINUE TO Q. 52)**
2. **No (SKIP TO Q. 54)**

52. (E1D) The contractor who installed the equipment

1. Very Dissatisfied
2. Somewhat Dissatisfied
3. Neither Satisfied Nor Dissatisfied *(SKIP TO Q.56)*
4. Somewhat Satisfied *(SKIP TO Q.56)*
5. Very Satisfied *(SKIP TO Q.56)*
6. Not applicable *(SKIP TO Q.56)*
7. Prefer not to answer *(SKIP TO Q.56)*
8. Don't know *(SKIP TO Q.56)*

53. Can you tell me why you gave that rating? *(RECORD VERBATIM)*

____

____

____

54. (E1E) The overall quality of the equipment installation

1. Very Dissatisfied
2. Somewhat Dissatisfied
3. Neither Satisfied Nor Dissatisfied *(SKIP TO Q.60)*
4. Somewhat Satisfied *(SKIP TO Q.60)*
5. Very Satisfied *(SKIP TO Q.60)*
6. Not applicable *(SKIP TO Q.60)*
7. Prefer not to answer *(SKIP TO Q.60)*
8. Don't know *(SKIP TO Q.60)*

55. Can you tell me why you gave that rating? *(RECORD VERBATIM)*

____

____

____

56. (E1F) The amount of time it took to receive your rebate for your equipment

1. Very Dissatisfied
2. Somewhat Dissatisfied
3. Neither Satisfied Nor Dissatisfied *(SKIP TO Q.58)*
4. Somewhat Satisfied *(SKIP TO Q.58)*
5. Very Satisfied *(SKIP TO Q.58)*
6. Not applicable *(SKIP TO Q.58)*
7. Prefer not to answer *(SKIP TO Q.58)*
8. Don't know *(SKIP TO Q.58)*

57. Can you tell me why you gave that rating? *(RECORD VERBATIM)*

____

____

____

58. (E1G). The dollar amount of the rebate for the equipment

1. Very Dissatisfied
2. Somewhat Dissatisfied
3. Neither Satisfied Nor Dissatisfied *(SKIP TO Q.60)*
4. Somewhat Satisfied *(SKIP TO Q.60)*
5. Very Satisfied *(SKIP TO Q.60)*
6. Not applicable *(SKIP TO Q.60)*
7. Prefer not to answer *(SKIP TO Q.60)*
8. Don't know *(SKIP TO Q.60)*

59. Can you tell me why you gave that rating? *(RECORD VERBATIM)*

____

____

____

60. (E1H) Interactions with PNM

1. Very Dissatisfied
2. Somewhat Dissatisfied
3. Neither Satisfied Nor Dissatisfied *(SKIP TO Q.62)*
4. Somewhat Satisfied *(SKIP TO Q.62)*
5. Very Satisfied *(SKIP TO Q.62)*
6. Not applicable *(SKIP TO Q.62)*
7. Prefer not to answer *(SKIP TO Q.62)*
8. Don't know *(SKIP TO Q.62)*

61. Can you tell me why you gave that rating? *(RECORD VERBATIM)*

____

____

____

62. (E1I) The overall value of the equipment your company received for the price you paid

1. Very Dissatisfied
2. Somewhat Dissatisfied
3. Neither Satisfied Nor Dissatisfied *(SKIP TO Q.64)*
4. Somewhat Satisfied *(SKIP TO Q.64)*

5. Very Satisfied *(SKIP TO Q.64)*
6. Not applicable *(SKIP TO Q.64)*
7. Prefer not to answer *(SKIP TO Q.64)*
8. Don't know (SKIP TO Q.64)

63. Can you tell me why you gave that rating? *(RECORD VERBATIM)*

____

____

____

64. (E1J) The amount of time and effort required to participate in the program

1. Very Dissatisfied
2. Somewhat Dissatisfied
3. Neither Satisfied Nor Dissatisfied *(SKIP TO Q.66)*
4. Somewhat Satisfied *(SKIP TO Q.66)*
5. Very Satisfied *(SKIP TO Q.66)*
6. Not applicable *(SKIP TO Q.66)*
7. Prefer not to answer *(SKIP TO Q.66)*
8. Don't know *(SKIP TO Q.66)*

65. Can you tell me why you gave that rating? *(RECORD VERBATIM)*

____

____

____

66. (E1K) The project application process

1. Very Dissatisfied
2. Somewhat Dissatisfied
3. Neither Satisfied Nor Dissatisfied *(SKIP TO Q.68)*
4. Somewhat Satisfied *(SKIP TO Q.68)*
5. Very Satisfied *(SKIP TO Q.68)*
6. Not applicable *(SKIP TO Q.68)*
7. Prefer not to answer *(SKIP TO Q.68)*
8. Don't know *(SKIP TO Q.68)*

67. **Can you tell me why you gave that rating?** *(RECORD VERBATIM)*

____

____

____

68. **(E2) Do you have any recommendations for improving the [REBATE_PROGRAM] program?**

01. Yes *(RECORD VERBATIM)*

____

____

____

97. No
98 Prefer not to answer
99. Don’t know

SECTION: CHARACTERISTICS AND DEMOGRAPHICS

69. **(Gen 1) Finally, I have a few questions about your firm for classification purposes only. Do you own or lease your building where the project was completed?**

01. Own
02. Lease / Rent
03. Prefer not to answer *(SKIP TO Q. 71)*
99. Don't know *(SKIP TO Q. 71)*

Other *(SPECIFY)* ____

70. **(Gen1a) Does your firm pay your PNM bill, or does someone else (e.g., a landlord)?**

1. Pay own
2. Someone else pays
3. Prefer not to answer
4. Don’t know

71. **(Gen2) Approximately what is the total square footage of the building where the project was completed?** (READ CATEGORIES IF NEEDED)

1. Less than 1,000 square feet

2. Between 1,000 and 1,999 square feet
3. Between 2,000 and 4,999 square feet
4. Between 5,000 and 9,999 square feet
5. Between 10,000 and 49,999 square feet
6. Between 50,000 and 99,999 square feet
7. 100,000 square feet or more
8. Prefer not to answer *(DO NOT READ)*
9. Don't know *(DO NOT READ)*

72. (Gen3) Approximately what year was your firm's building built? (READ CATEGORIES IF NEEDED)

1. 1939 or earlier
2. 1940 to 1949
3. 1950 to 1959
4. 1960 to 1969
5. 1970 to 1979
6. 1980 to 1989
7. 1990 to 1999
8. 2000 to 2009
9. 2010 and later
10. Prefer not to answer *(DO NOT READ)*
11. Don't know *(DO NOT READ)*

73. (Gen4) Approximately, How many full-time equivalent (FTE) employees does your company currently have in the state of New Mexico?

1. Less than 5
2. 5-9
3. 10-19
4. 20 - 49
5. 50 - 99
6. 100 - 249
7. 250 - 499
8. 500 - 999
9. 1,000 - 2,500
10. More than 2,500
11. Prefer not to answer

12. Don’t know

74. (Gen5) And this is my last question. How long has your company been in business?
(Poller : Please be specific, by writing in months and years.)

____

98. Prefer not to answer
99. Don’t know

THIS CONCLUDES OUR SURVEY. THANK YOU FOR YOUR TIME. HAVE A GOOD DAY.

NOTE TO INTERVIEWER, WAS RESPONDENT:

1. Male
2. Female

Unique ID #:_____ _____ _____

Respondent’s Phone Number:_________________________________

Interviewer’s Name:____

Interviewer’s Code:____

Appendix B – Contractor Interview Guide

Introduction

Talking points for recruitment

- Evergreen Economics is conducting an evaluation of [UTILITY's] [PROGRAM] for the New Mexico Public Regulation Commission and the state's utilities.
- We have identified selected contractors that installed equipment that received rebates from the efficiency programs in 2019 for brief telephone interviews.
- We would need about 20 minutes for the interview.
- Your responses will be anonymous, but will be very helpful in helping the state's utilities ensure their energy efficiency programs best serve their customers.
- When would be a good time to talk?

Talking points for starting the interview

- Identify self.
- This should take about 20 minutes.
- Your responses will be anonymous, so please feel free to speak candidly.
- Do you have any questions before we begin?
- Would you feel comfortable if I record this call for note taking purposes? We will not share the recording with anyone outside our company and will not attribute anything you say back to you.

Interviewee Background

Let's begin with a couple of background questions….

A1. To start, please tell me a bit about your company.
Probe to understand:

- Services offered
- Types of customers (esp. sector – residential, commercial, or both)
- Regions served
- Interviewee role

Program Awareness and Engagement

B1. Do you recall how you first learned about and got involved with the [residential/commercial] rebate programs through [UTILITY]?
Listen (and probe as needed) for:

- Any reservations about participating
- Any barriers to participating

- Whether or not they work with any other New Mexico [UTILITY] rebate programs

B2. Could you describe what involvement with New Mexico [UTILITY] rebate programs as a contractor involves?

Probe as needed:

- In what ways do you interact with New Mexico [UTILITY] or their implementers about this program?
- What information or services do you receive from New Mexico [UTILITY] (beyond the ability to offer rebates to your customers)?

B3. In what ways is the [UTILITY] program helpful to you in your business?

Probe, as needed:

- Rebate
  - Increases customer satisfaction with us
  - Increases business
  - Helps us up-sale to higher efficiency levels
- Ability to mention the connection with the [UTILITY] program
- [UTILITY] messaging to customers on benefits of [MEASURE(S)]

B4. What share of your [residential/commercial] projects within [UTILITY] territory would you estimate currently end up qualifying for and receiving a [UTILITY] rebate?

- What could [UTILITY] do to involve you more in the program?

B5. Does [UTILITY] make it clear which of your products or services are eligible for [UTILITY] rebates?

Probe as needed:

- Is there anything [UTILITY] should do to more clearly communicate that?

B6. Have the programs influenced what equipment you suggest to a customer?

B7. Do you have any suggestions for [UTILITY] contractor services and support – either overall or for the [PROGRAM] specifically?

Program Processes

C1. In what ways are you involved with the rebate portion of the program and the paperwork and process required to participate?

Probe to understand:

- Whether contractor completes the rebate application

- Time required for paperwork and whether that is a burden
- Whether the rebate goes directly to the customer or contractor (with a markdown on the charge to customer)
- Recommended improvements

C2. When and how do you bring up either [UTILITY] rebates or the equipment they rebate when talking with customers?

Listen for (and probe as needed):

- What share of customers are already aware of rebates before the contractor brings it up
- What it is the most effective sales tool or message to get customers to upgrade to high efficiency
- What role the [UTILITY] rebates play in motivating upgrades
- What particular equipment is easier or harder to get customers to upgrade to high efficiency and why

C3. Do you have any comments about the program offerings? Is there anything missing? Anything not needed? Or anything that could be better?

Market Response

D1. Overall, to what degree do you see the program increasing the interest and demand for energy efficient equipment?

Probe to understand:

- Why is that?
- Is the program having a large or small effect on the market?

D2. Are there markets that you feel [UTILITY] [residential/commercial] energy efficiency programs are reaching well? Not well?

Probe to understand:

- Suggested approaches that might expand the reach of the program into markets that may be underserved by the program.

D3. Overall, what issue(s), if any, may affect future program participation by customers? What about future program participation by contractors? [INTERVIEWER NOTE: Example issues are changes to building codes and standards being promoted and program incentive levels].

Program Satisfaction

E1. Finally, I'd like to ask about your and your customers' satisfaction with the [UTILITY] [PROGRAM]. Please rate your overall satisfaction with the program on a 1 to 5 scale where 1 is not at all satisfied, 2 is somewhat dissatisfied, 3 is neither satisfied nor dissatisfied, 4 is somewhat satisfied and 5 is very satisfied?

- What is your satisfaction?
- How do you think your customers would rate the program?

[IF RATING < 5] What could [UTILITY] do to increase your satisfaction with the program?

Probe if needed:

- What is working best?
- What is most challenging or needs improvement?

E2. Have you had any feedback from your customers about their experiences with the [PROGRAM] that you think [UTILITY] should know?

E3. Aside from anything we've already discussed, was there ever an occasion when the program didn't meet your expectations? Please explain.

Closing

F1. Is there anything else we didn't cover that you'd like to mention or discuss about your experiences with the [UTILITY] [PROGRAM]?

[THANK AND END]

Appendix C – Power Saver Detailed Evaluation Methods and Findings

Power Saver is a direct load control program offered to residential, small commercial (< 50 kW), and medium commercial (50 kW – 150 kW) Public Service New Mexico (PNM) customers. To facilitate load control, participants must have a device attached to the exterior of their air conditioning unit. This device is capable of receiving a radio signal that will turn off the unit's compressor for an interval of time. Such signals are typically sent on the hottest weekday afternoons of the summer, with the goal being to reduce peak demand. Residential and small commercial participants receive an annual $25 incentive for their participation. Medium commercial participants receive an annual incentive of $9 per ton of refrigerated air conditioning. A residential smart thermostat component was added to the program in 2018. For this component, load curtailment is achieved via communication with the WiFi-enabled thermostat.

There were five Power Saver events during the summer 2019 demand response (DR) season, which began June 1st and ended September 30th. Table 1 provides some information on these five 2019 events. All events used a 50% cycling strategy where curtailment is based on the runtime in the previous hour. The events on 8/19 and 8/26 were dispatched for just the Residential DCU M&V segment and the Small Commercial M&V segment. Note that the event start times and end times are in Mountain Daylight Time (MDT).

Table 1: 2019 Power Saver Event Summary

Date	Day of Week	Start Time (MDT)	End Time (MDT)	Daily High at KABQ (F)
7/10/2019	Wednesday	3:00 PM	7:00 PM	97.0
8/19/2019	Monday	2:00 PM	6:00 PM	93.9
8/26/2019	Monday	2:00 PM	6:00 PM	98.1
8/27/2019	Tuesday	3:00 PM	7:00 PM	95.0
9/4/2019	Wednesday	3:00 PM	7:00 PM	90.0

Shortly after the conclusion of the summer 2019 season, Itron provided the Evergreen team with a series of datasets for the evaluation. These files included:

- For Residential DCU, Small Commercial, and Medium Commercial sites, 5-minute load data from 6/1/2019 to 9/30/2019

- For Residential DCU and Small Commercial sites, an M&V list that provided the location type (residential or commercial), the group (control or curtailment), and/or the dates each load control device was active
- For Medium Commercial sites, an M&V list that provided the dates each load control device was active
- For the Two Way Smart Thermostat group, 5-minute runtime data from 6/1/2019 to 9/30/2019

The Evergreen team also received Itron's Power Saver impact evaluation report, which detailed the methods Itron employed in calculating customer baselines (CBLs) for the four different participant classes. A CBL is an estimate of what participant loads would have been absent the DR event dispatch. By customer class, the report also showed the load impact, which is the difference between the CBL and the metered load, for each 5-minute interval of each curtailment day. The key steps in the Evergreen verified savings analysis were:

1) For each customer class, reproduce the performance estimates calculated by Itron using the contractually-agreed upon CBL method.
2) Modify the CBL methodology and produce ex post estimates of what the per-device impact was during the 2019 DR season.
3) Where possible, leverage additional historical data from 2015 - 2018 to produce ex ante estimates of what the per-device impact at peaking conditions (3:00 PM at 100°F) will be in future summers.
4) Scale the per-device estimates by the number of active program devices to calculate the aggregate load impacts (MW) of the Power Saver program.

Table 2 and Table 3 summarize our findings for residential and commercial segments, respectively. The main driver in the difference between Itron and Evergreen load reduction estimates is that Itron commonly summarized impacts with the maximum (e.g., the largest 5-minute impact in a one-hour interval is the impact for that hour), whereas the Evergreen team summarized impacts with an average. Multiplying our per-device reduction estimates by the number of devices in each class (shown in Table 2) leads to a 2019 average total estimated load reduction of approximately 21.9 MW, 0.4 MW, 2.4 MW, and 0.8 MW for the Residential DCU, Two Way Smart Thermostat, Small Commercial, and Medium Commercial segments respectively. In aggregate, the average 2019 performance is 25.5 MW. This is approximately 75% of Itron's estimate (34.1 MW). After making an online adjustment for the thermostat group (85%) and an operability adjustment (86%) for the other three segments, the aggregate Evergreen-calculated impacts for 2019 are 21.9 MW (compared to 29.3 MW from Itron after adjustment).

The Evergreen team used Power Saver results from 2015-2019 to estimate the load relief capability under extreme conditions. We estimate the program is capable of delivering 32.7

MW of load reduction under planning conditions of 100°F between 5:00 PM and 6:00 PM MDT, of which 29.4 MW comes from the Residential DCU segment, 0.5 MW comes from the Two-Way Smart Thermostat segment, 1.8 MW and 1.0 MW come from small and medium commercial customers, respectively. Factoring in the operability/online adjustments, the aggregate program can provide 28.1 MW of load relief.

Table 2: High Level Results – Residential

		Unit	Residential DCU		Smart Thermostats	
			Measured	Adjusted	Measured	Adjusted
Number of Devices Installed		#	41,376	41,376	384	384
Itron	5-year Rolling Average kW Factor	kW / device[1]	0.82		0.68	
Itron	5-year Rolling Average kW Factor	Total MW	33.93		0.26	
Itron	2019 Load Reduction Estimate	kW / device	0.69	0.59	0.80	0.68
Itron	2019 Load Reduction Estimate	Total MW	28.55	24.55	0.31	0.26
Evergreen	2019 Load Reduction Estimate	kW / device	0.53	0.46	1.02	0.87
Evergreen	2019 Load Reduction Estimate	Total MW	21.93	18.86	0.39	0.33
Evergreen	Ex Ante Load Reduction Estimate[2]	kW / device	0.71	0.61	1.36	1.16
Evergreen	Ex Ante Load Reduction Estimate[2]	Total MW	29.38	25.26	0.52	0.44
Evergreen	2019 Energy Savings	kWh / device	2.43	2.09	6.55	5.57
Evergreen	2019 Energy Savings	Total MWh	100.77	86.66	2.52	2.14

[1] 2019 kW factors include a rolling average per-device result for 2015-2019. 2018 Residential DCU kW factor has an 85% operability adjustment applied. 2019 Residential DCU kW factor has an 86% operability adjustment applied. The 86% operability percentage was calculated as 85% multiplied by the number of DCU sites that have not been visited in the last two years plus 95% multiplied by the number of DCU sites that were visited in the last two years. 2019 Two-Way Smart Thermostats have an 85% offline (not operability) adjustment applied.

[2] Ex ante program capability is reported in the 5 PM – 6 PM MDT hour at 100°F.

Table 3: High Level Results – Commercial

		Unit	Small Commercial		Medium Commercial	
			Measured	Adjusted	Measured	Adjusted
Number of Devices Installed		#	3,443	3,443	2,636	2,636
Itron	5-year Rolling Average kW Factor	kW / device[3]	1.38		0.72	
		Total MW	4.75		1.90	
	2019 Load Reduction Estimate	kW / device	1.21	1.04	0.40	0.35
		Total MW	4.17	3.58	1.07	0.92
Evergreen	2019 Load Reduction Estimate	kW / device	0.69	0.59	0.30	0.26
		Total MW	2.38	2.04	0.79	0.68
	Ex Ante Load Reduction Estimate	kW / device	0.52	0.45	0.39	0.34
		Total MW	1.79	1.54	1.03	0.88
	2019 Energy Savings	kWh / device	4.37	3.76	4.70	4.04
		Total MWh	15.16	13.04	12.39	10.65

[3] 2019 kW factors include a rolling average per-device result for 2015-2019. 2019 Small Commercial and Medium Commercial have an 86% operability adjustment applied. The 86% operability percentage was calculated as 85% multiplied by the number of DCU sites that have not been visited in the last two years plus 95% multiplied by the number of DCU sites that were visited in the last two years.

1 Methodology

This section discusses the methods used to validate Itron's impact estimates and those used by the Evergreen team to provide their ex post and ex ante impact estimates.

1.1 Residential DCU & Two-Way Smart Thermostat Impact Validation

The impact evaluation for the Residential DCU class relies on an alternating treatment design. Under this approach, load in the group that was not dispatched serves as a proxy for what curtailment group load would have been if the DR event had not been initiated. Group A contained 144 devices and Group B contained 142 devices.

Impact estimates were derived using 5-minute interval kW data collected by DENT Elite Pro SP Portable Power Data Loggers and PowerCAMP and IntelliMEASURE M&V equipment. Steps taken are as follows:

1. For both the control and curtailment groups, calculate the average demand (kW) for each 5-minute interval.
2. For both the control and curtailment groups, calculate a fifteen-minute rolling average demand. Suppose the average demand for the control group is 3 kW during interval t, 4 kW during interval $t + 1$, and 5 kW during interval $t + 2$. The fifteen-minute rolling average demand for interval t would then be 4 kW.
3. For each interval, find the difference between the rolling averages for the control and curtailment groups (where difference = control – curtailment).
4. The impact for any given event hour is the maximum difference across the 12 intervals in the hour, as calculated in step 3.
5. The maximum difference across all qualified event hours[4] is the kW per device impact estimate for the 2019 DR season.
6. Adjust the residential impacts for an operability factor of 85%. The determination of the operability percentage is detailed in detail in Section 1.6.

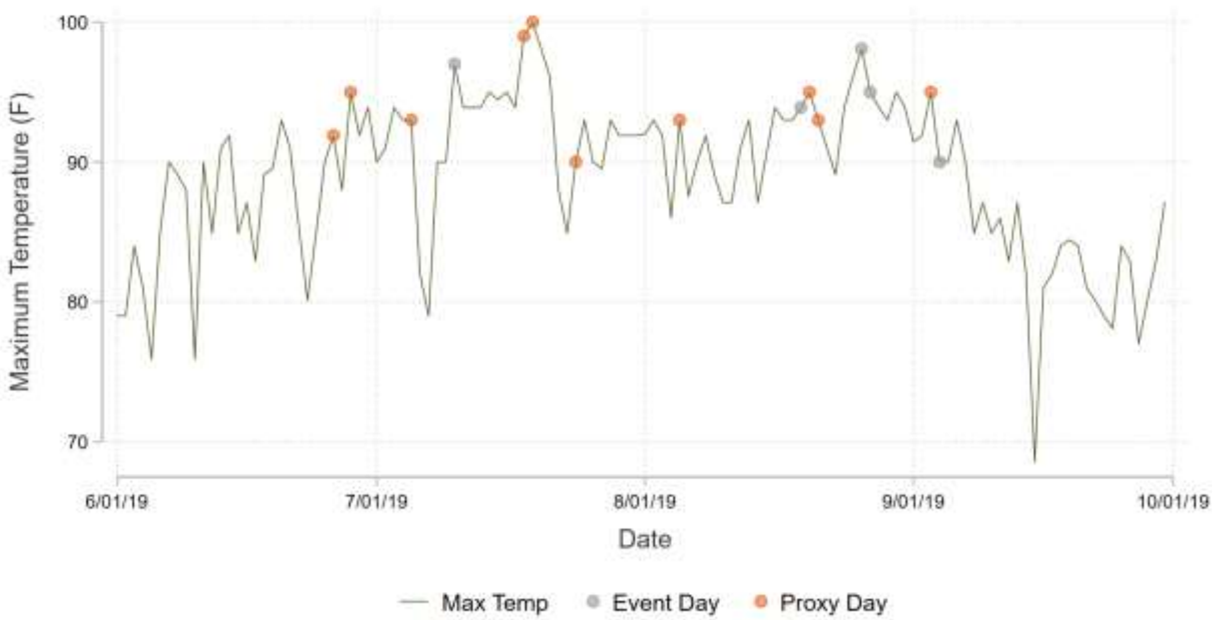
Impacts for the Two-Way Smart Thermostat segment were calculated in a similar manner. One key difference is that there was not an alternating treatment design. The same group of 50 thermostats was used as the non-curtailment group for each Power Saver demand response event. An online factor of 85% was applied to the thermostat segment in lieu of an operability factor. (These are two-way devices, so knowing which devices are active is possible for this segment.)

[4] 'Qualified' hours were defined as hours where the outdoor temperature is at least 97 degrees (F).

1.2 Evergreen Estimate of Residential DCU & Two-Way Smart Thermostat Impacts

In 2018, the Residential DCU segment of Power Saver switched to alternating dispatch between M&V groups to determine which devices were called to reduce load on event days. In theory, this means that any difference in the behavior of the two groups is removed when we look at events across the whole summer. Because dispatch alternated between the two groups, any bias in impacts should be minimal, on average. Nevertheless, to assess the differences between the groups, the Evergreen team compared the load profiles of the two groups on proxy days. Proxy days are non-event days that were chosen from non-holiday weekdays where the maximum temperature was at least as hot as the event days. From this pool, of which there were 46 available days, the top five hottest were chosen and five more were randomly selected to provide a 2:1 proxy to event day ratio.[5] Figure 1 shows the maximum temperature and distribution of proxy days throughout the summer, compared to the event days and non-event days.

Figure 1: Weather on Event and Proxy Days

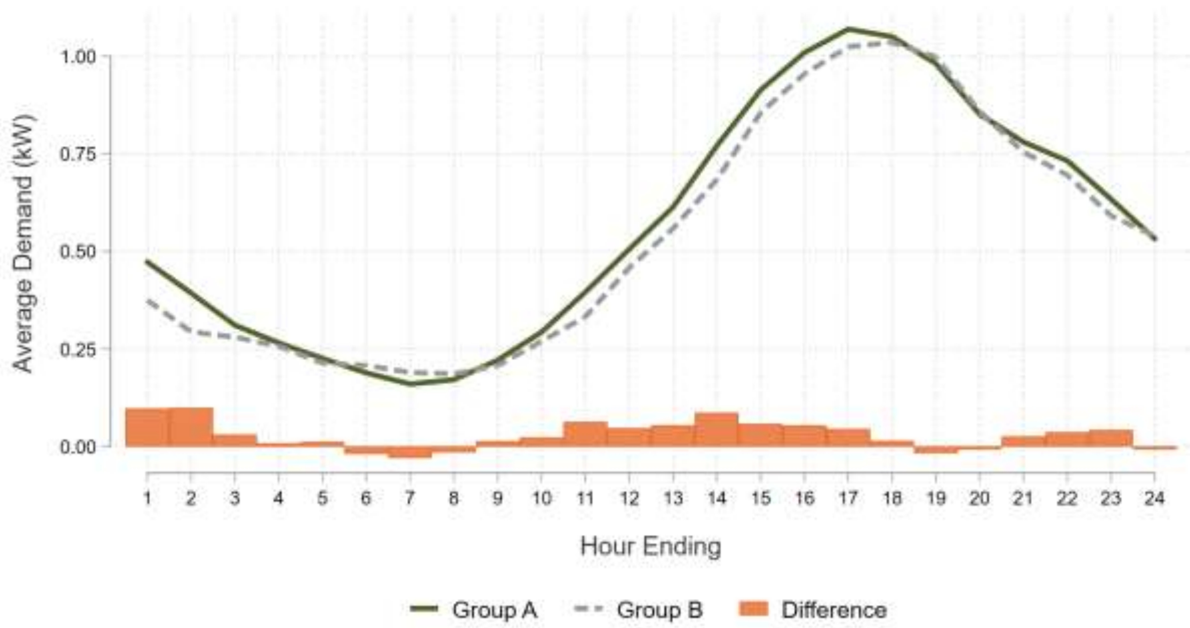


The average hourly load profiles for the two residential M&V groups, averaged across all proxy days, are shown in Figure 2. The average difference between the two groups is 0.03 kW, with a maximum difference of 0.10 kW. The average difference during event hours is 0.03 kW and the maximum is 0.06 kW.

---

[5] In order, the dates were 6/26, 6/28, 7/5, 7/18, 7/19, 7/24, 8/5, 8/20, 8/21, 9/3.

Figure 2: Residential DCU Load Shapes on Event-Like Days



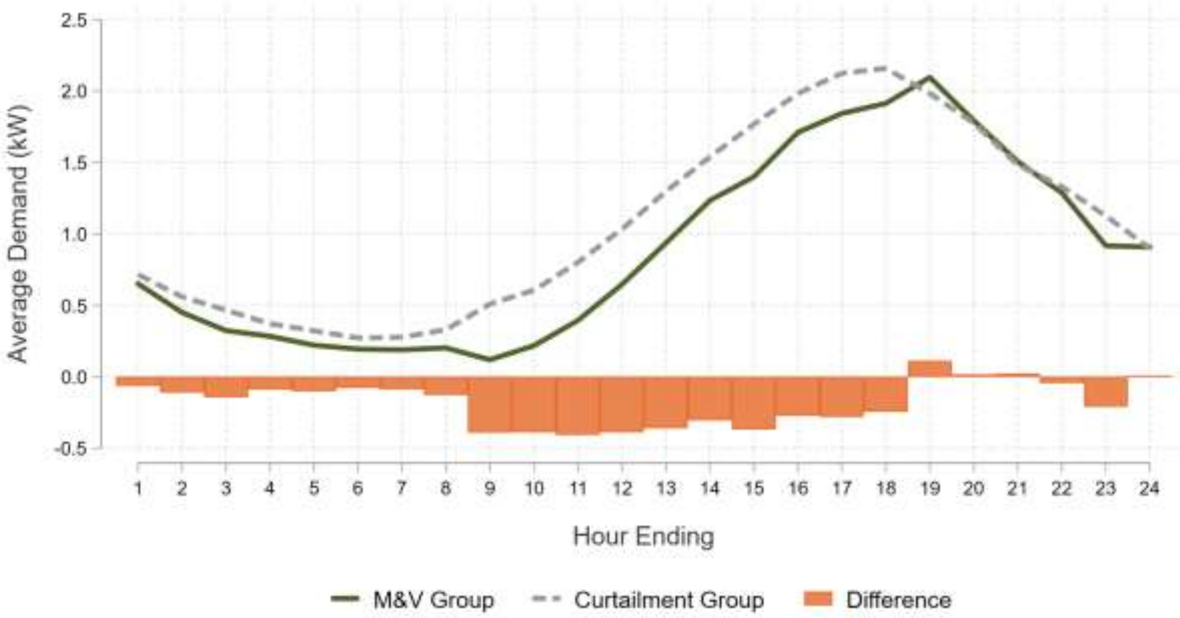
Using a t-test, the Evergreen team found the difference between average demand in Group A and Group B to be statistically significant during event hours. Therefore, we felt that taking the simple difference between the two groups would not be sufficient to calculate an unbiased ex post event impact. Instead, we used a difference-in-differences approach. Table 4 provides an illustration. In this illustration, Group A is the curtailment group. The difference-in-difference calculation nets out the proxy day difference from the event day difference.

Table 4: Difference-in-Difference Illustration

Hour Ending (MDT)	Proxy Day Difference (kW)	Event Day Difference (kW)	Difference-in-Difference (kW)
3:00 PM	0.06	0.37	0.31
4:00 PM	0.05	0.43	0.37
5:00 PM	0.05	0.46	0.41
6:00 PM	0.01	0.56	0.55

A similar method was used for the Two-Way Smart Thermostat segment, as there were statistically significant differences in average load during event hours between the M&V group and the curtailment group.

Figure 3: Smart Thermostat Two-Way Load Shapes on Event-Like Days



1.3 Small Commercial & Medium Commercial Impact Validation

The impact evaluation for the Small Commercial and Medium Commercial classes relies on a "high X of Y" customer baseline (CBL) approach with a multiplicative day-of adjustment. Under this approach, the average load for three of the previous five eligible[6] days is used as a proxy for what load would have been if the DR event had not been called. In selecting which three days to use, the criterion is greatest maximum load during the event window. For a hypothetical event that lasts from 3:00 PM until 7:00 PM, the steps to calculating the impact estimate are as follows:

1. Calculate the unadjusted baseline.
    - For each of the five eligible days prior to the event day, calculate the average demand during event hours across the entire M&V population. Select the three days with the greatest average demand (i.e., "high 3 of 5").
    - Across the three baseline days, calculate the average demand across the entire M&V population for each 5-minute interval. This essentially collapses the three baseline days into one baseline day.
    - For each 5-minute interval, calculate a 15-minute rolling average kW load. As an example, suppose the average 5-minute interval load is 10 kW at time t,

---

[6] Eligible days are weekdays that are neither holidays or DR event days.

12 kW at time $t + 1$, and 14 kW at time $t + 2$. The 15-minute rolling average kW load at time t would be (10 + 12 + 14)/3 = 12 kW. This value (12 kW) would be the unadjusted CBL at time t.

2. Calculate 15-minute rolling average demand (kW) for the entire M&V population.
   - Across the entire M&V population, calculate average demand for each 5-minute interval.
   - For each 5-minute interval, calculate a 15-minute rolling average as described above.
3. Calculate the multiplicative adjustment factor.
   - For the twelve 5-minute intervals preceding the event, sum up the 15-minute rolling average demand for the unadjusted baseline.
   - For the twelve 5-minute intervals preceding the event, sum up the 15-minute rolling average demand for the M&V population.
   - Divide the second sum by the first sum. This quotient is the adjustment factor.
4. Calculate the impact.
   - Multiply the unadjusted baseline by the adjustment factor. This yields the adjusted CBL.
   - For each 5-minute interval, subtract the 15-minute rolling average demand for the entire M&V population (as calculated in Step 2) from the adjusted baseline. Note that this yields 12 impacts in every hour.
   - For each event hour, take the maximum 5-minute impact. This value serves as the impact estimate for the event hour.
   - The maximum 5-minute impact across all qualified event hours (temperature exceeds 96°F) is the 2019 Power Saver impact estimate.

1.4 Evergreen Estimate of Small Commercial & Medium Commercial Impacts

Discussed further in the Small Commercial and Medium Commercial results sections, the Evergreen team feels that using the maximum 5-minute rolling average difference in each hour as the impact estimate overstates the capability of the program by including favorable noise. To calculate the evaluated impact estimates, Evergreen used the same baseline method as summarized in Section 1.3 is used; however the rolling 5-minute impacts are summarized by the mean rather than the maximum by hour.

1.5 Ex Ante Impacts

Of particular interest for ex ante load considerations is how sensitive the program performance is to temperature. When additional years of data are included in such an analysis, a wider range of program conditions can be investigated which leads to a more robust understanding of the capability of the program.

To produce an ex ante impact estimate for Residential DCU customers, the Evergreen team leveraged 2015-2018 summer load data in addition to the 2019 summer load data. In 2015-2017, only one of the Residential DCU M&V groups was consistently curtailed while the other group acted as a control. Because some differences exist between the two groups in terms of load profile on event-like days, the Evergreen team used a difference-in-differences impact estimation method, which was described in Section 1.2, to estimate the impacts for these earlier summers.[7] Ex post impacts in 2018 were not calculated via difference-in-differences, as statistically significant differences between the groups were not found.

To produce an ex ante impact estimate for the Small Commercial segment, the Evergreen team leveraged 2015-2018 summer load data in addition to 2019 summer load data. In prior years, impacts for the Small Commercial segment were calculated in a manner similar to the Residential DCU segment – an M&V group was split into curtailment and control groups. The control group was used as a baseline for the curtailment group. In 2019, the full M&V group was curtailed for all events, and the program implementer relied on an X-of-Y baseline method to estimate impacts (same method as the one used for the Large Commercial segment). The Evergreen team changed our approach for 2019 as well – this is discussed elsewhere in this document. The takeaway is that the ex ante estimate is a function of historical ex post estimates, and the method for developing ex post method has changed over the years.

For Medium Commercial, we leveraged 2017 and 2018 data in addition to 2019 summer load data. The same approach for estimating ex post results for the Medium Commercial segment was used in 2017, 2018, and 2019.

For the Two-Way Smart Thermostat segment, only 2019 summer data was available. Thus, our ex ante impact estimate does not leverage historical data for this segment.

Once data had been compiled for each customer segment, a regression was run that explains changes in impacts as a function of temperature and hour, which was then used to predict impacts for a range of planning scenarios. The regression equation specified was:

[7] There were not many non-event weekdays during the summer of 2015 where the maximum outdoor temperature exceeded 94 degrees (F), so a threshold of 91 degrees (F) was used for the 2015 data instead. The temperature threshold for the summer of 2016 was 94 degrees (F), just like the threshold for the summer of 2017. In 2018, the groups were similar in terms of non-event day usage, so the difference-in-differences method was not necessary.

$$\Delta kW_h = \alpha + \beta * T_t + \sum_{h=14}^{h=19} \gamma_h * I_h + \sum_{h=14}^{h=19} \delta_h * I_h * T_h + \varepsilon_h$$

Where the variables have the following interpretations:

Table 5: Ex Ante Regression Terms

Variable	Interpretation
α	Constant term
β	The incremental kW usage associated with a warming of 1 degree Fahrenheit
T_t	Outdoor air temperature in hour h
γ_h	Incremental kW usage associated with each hour
I_h	Indicator variable equal to 1 if the hour is 14, 15, 16, etc., and 0 if not
δ_h	Incremental kW usage associated with a 1 degree increase in outdoor temperature in hour h
ε_h	The error term

1.6 Operability Adjustments

To reach a true estimate of program capability, ex post and ex ante impacts in this analysis need to be adjusted for operability. In a previous evaluation, the Evergreen team recommended adjusting residential impacts by 85% based on operability inspections that occurred during Summer 2018. Our 2018 Evaluation Report covered the inspection process and key findings in detail. Itron's 2018 report adopted this recommendation. In 2019, they altered their adjustment approach. An 86% adjustment factor is used instead. This value represents a weighted average of 85% and 95%. Respectively, the weights are the number of sites that have not been visited in the past two years and the number of sites that have been visited in the past two years. We believe this approach makes sense and will adopt Itron's operability adjustment factor of 86% for the following segments: Residential DCU, Small Commercial, and Medium Commercial. Itron's report notes that an 85% online factor (not operability factor) is applied to the Two-Way Smart Thermostat group. We will adopt this adjustment as well.

Unless otherwise noted, results in this analysis are reported without the operability adjustment applied.

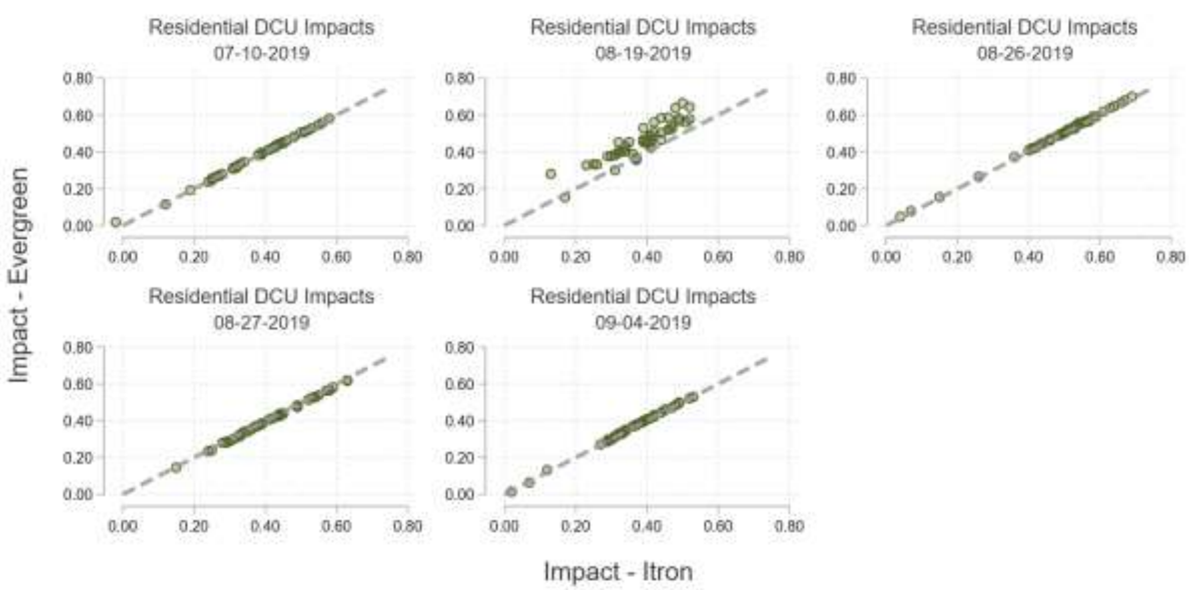
2 Residential DCU Results

This section reviews the Residential DCU impacts calculated by Itron and validated by the Evergreen team. Additionally, the team provides feedback on the evaluation approach used by Itron and provides an alternative impact analysis for summer 2019 events. Finally, ex ante impacts, combining multiple years of event history, are produced for various temperature scenarios.

2.1 Validation of Calculations

After receiving the participant load data from Itron, the Evergreen team attempted to reproduce the impacts in Itron's Power Saver impact evaluation report. For each event hour other than the hours on 8/19, the Evergreen team was able to replicate Itron's impact estimates for the Residential DCU segment. Figure 4 compares impacts as calculated by Itron and by Evergreen at the 5-minute level. For reference, Itron's Residential DCU impact estimates are shown in Table 6. Note that an asterisk (*) denotes a qualified event hour. The maximum impact during qualified event hours was 0.69 kW for the Residential DCU class without any adjustment for operability.

Figure 4: Residential DCU Impact Verification



The dotted line represents what a perfect match would look like.

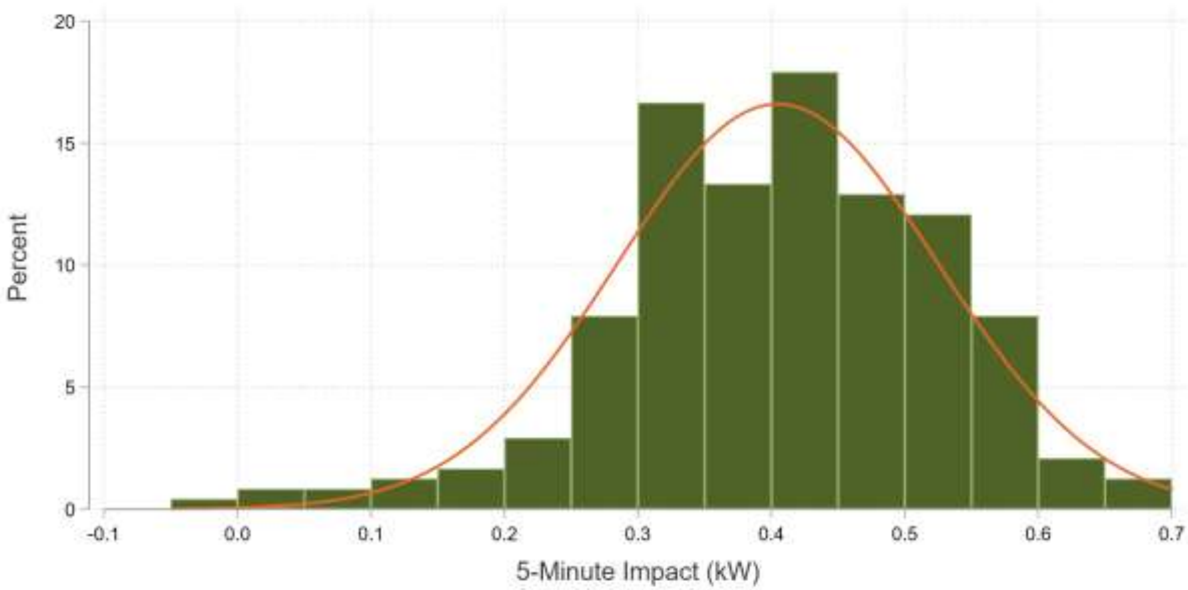
Table 6: Residential Impact Estimates (kW) by Date and Time

Date	Hour Ending (MDT)				
	3:00 PM	4:00 PM	5:00 PM	6:00 PM	7:00 PM
07/10/2019		0.42	0.51*	0.56	0.58
08/19/2019	0.44	0.51	0.49	0.52	
08/26/2019	0.50	0.56*	0.69*	0.66*	
08/27/2019		0.63	0.58	0.45	0.37
09/04/2019		0.40	0.53	0.42	0.49

2.2 Evergreen Ex Post Impacts

For the Residential DCU segment, Itron's per device kW impact estimate for the 2019 season is the maximum difference between 5-minute rolling average loads for the control and curtailment groups (0.69 kW). (See Section 1.1 for more details.) The critical word here is *maximum*. The Evergreen team feels that using the maximum difference overstates the amount of load shed produced by a typical Power Saver DR event by counting favorable noise. This is especially true from a system planning perspective, as using the maximum is a poor basis for the estimated load relief upon dispatch. Figure 5 shows the distribution of impacts at the 5-minute level – 0.69 kW clearly overstates the center of the distribution.

Figure 5: Distribution of 5-Minute Residential DCU Impacts



Respectively, the mean and median are 0.40 kW and 0.41 kW.

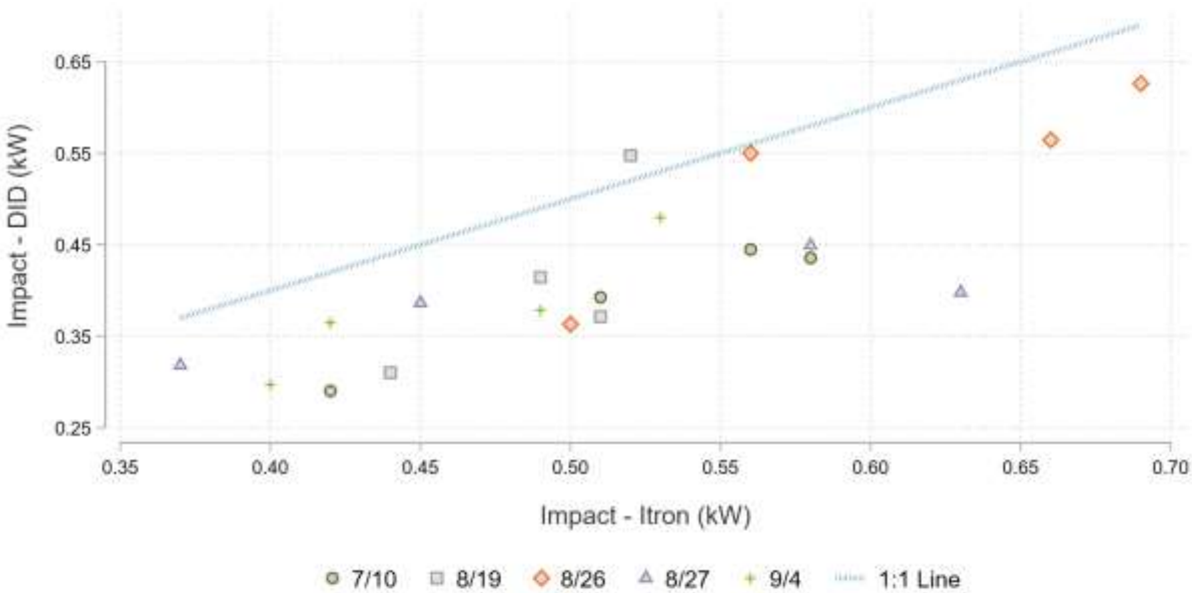
Rather than the maximum difference, the Evergreen team feels that using an average impact across an hour (rather than a maximum) returns an unbiased estimate of Power Saver program impacts during DR events. Since statistically significant differences in afternoon demand were found between the two groups (Figure 2), the Evergreen team opted for a difference-in-difference approach for estimating ex post impacts. This approach was described in Section 1.2. Results for the 2019 DR season are summarized in Table 7. Note that the curtailment group rotated between events, which is why the sign of the non-event-day difference changes from one event to the next.

Table 7: Impact Calculations

Date	# of Curtailed Devices	Hour Ending MDT	Temp. (F)	Control kW	Curtail kW	Non-Event Diff. (kW)	Impact (kW)
7/10/2019	127	16	96	0.84	0.60	-0.05	0.29
		17*	97	0.91	0.56	-0.05	0.40
		18	96	1.07	0.64	-0.01	0.44
		19	96	1.10	0.65	0.02	0.43
8/19/2019	111	15	94	0.82	0.45	0.06	0.31
		16	94	0.92	0.50	0.05	0.37
		17	94	1.03	0.57	0.05	0.41
		18	94	1.16	0.60	0.01	0.55
8/26/2019	123	15	96	0.86	0.56	-0.06	0.36
		16*	97	1.06	0.56	-0.05	0.55
		17*	98	1.22	0.64	-0.05	0.63
		18*	97	1.23	0.68	-0.01	0.56
8/27/2019	113	16	93	1.05	0.60	0.05	0.40
		17	95	1.06	0.57	0.05	0.44
		18	90	1.00	0.60	0.01	0.39
		19	87	0.94	0.63	-0.02	0.33
9/4/2019	123	16	90	0.80	0.56	-0.05	0.29
		17	90	0.95	0.51	-0.05	0.49
		18	88	0.92	0.57	-0.01	0.36
		19	88	0.97	0.57	0.02	0.38

The average impact during full event hours was 0.42 kW. This is in line with the center of the distribution shown in Figure 5. Amongst full event hours, the average impact during qualified event hours was 0.53 kW. Figure 6 compares Evergreen's ex post hourly impacts with the impacts calculated by Itron. The Evergreen impact is lower in nearly all cases, by about 0.10 kW on average.

Figure 6: Comparison of Evergreen Ex Post Impacts and Itron Impacts



The 1:1 line shows what the trend would look like if the DID and Itron impacts were identical.

2.2.1 Net Energy Savings

The Evergreen team estimated net energy impacts for the Residential DCU segment by summing ex post impacts from the onset of each event through the end of the event day. The calculation of impacts is exactly as described earlier in this section. Table 8 shows the energy savings estimates (per device) for each event day. On average, net per device savings were 0.87 kWh per event day. However, two of the events (8/19 and 8/26) were called just for the M&V group – not the full Residential DCU population. For the three population events, the average net savings per device was 0.81 kWh per day. Scaling by the number of events (three) and the number of active devices (41,376 per Itron's report) yields an aggregate savings estimate of 100.5 MWh for the Residential DCU segment. If M&V energy savings for 8/19 and 8/26 are added in, the total is 100.8 MWh.

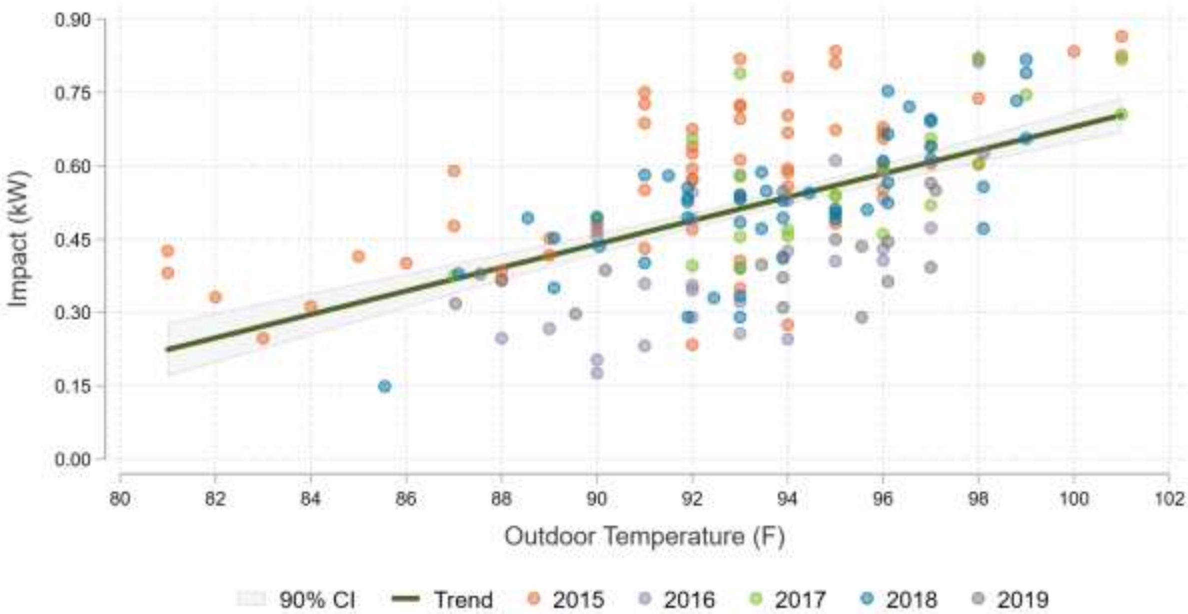
Table 8: Per Device Energy Savings by Event Day

Date	Event Start (MDT)	Event Savings (kWh)	Snapback (kWh)	Net Savings (kWh)
7/10/2019	3:00 PM	1.56	-1.10	0.46
8/19/2019	2:00 PM	1.64	-0.77	0.87
8/26/2019	2:00 PM	2.10	-1.08	1.02
8/27/2019	3:00 PM	1.55	-0.61	0.94
9/4/2019	3:00 PM	1.52	-0.49	1.03
Average		**1.68**	**-0.81**	**0.87**

2.3 Evergreen Ex Ante Impacts

Figure 7 compares 2015-2019 ex post impact estimates for each event hour with the outdoor air temperature for that hour. (Weather data comes from weather station KABQ in Albuquerque.) There is a clear trend in the figure – the hotter it is outside, the greater the impacts tend to be. To develop an ex ante impact estimate, the Evergreen team developed a regression model that estimates the ex post impact as a function of temperature and time. The specified model was shown in Section 1.5, and the results from the model are described in more detail below. Using the model, the Evergreen team predicts that the impact of a Residential DCU DR event at peaking conditions (5:00 PM – 6:00 PM MDT when outdoor temperature is 100 degrees) is 0.71 kW per device.

Figure 7: Hourly Impacts against Outdoor Temperature (F)



The regression was run on full event hours (some events in prior summers started mid-hour) and weighted by the number of curtailed devices (each summer had slightly different numbers of dispatched devices). Regression output is shown below. In general, earlier hours corresponded to higher kW values, with a drop over time in impacts as less load was available to shed. It should be noted that Hour 20 was extremely rare; only 3 events during the past four years included a full-hour event during this period and as such, should be interpreted with care. Temperature has a positive coefficient, indicating that higher temperatures produce larger load reductions. The interaction terms, represented by δ_h, are all positive, indicating that the incremental effect of temperature in a given hour further increases the impact. Again, Hour 20 should be interpreted with caution, as only three data points were available to fit the model. Note that any coefficient with "*" next to it is statistically significant.

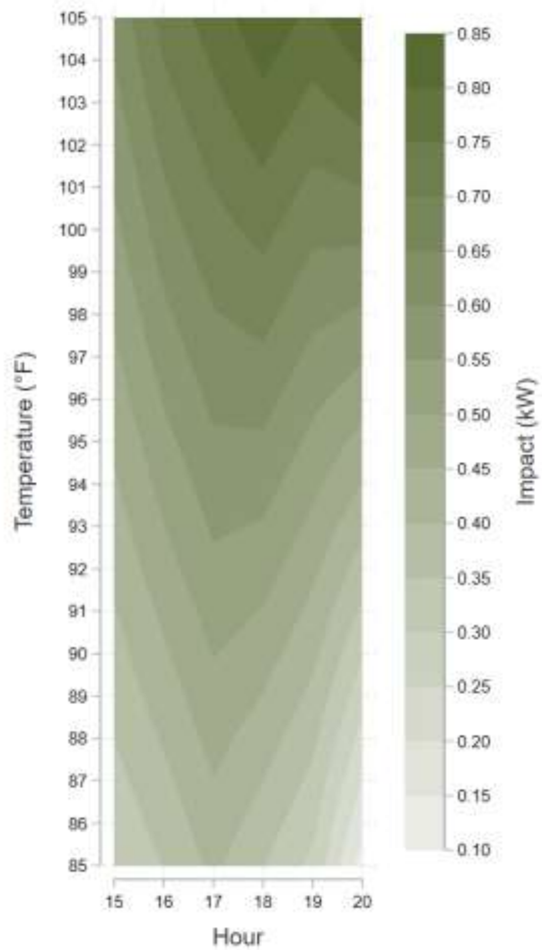
Table 9: Residential Ex Ante Regression Output

Term	Variable	Coefficient (b)	Standard Error	P-Value	95% CI
β	Temperature	0.01561*	0.00058	0.000	(0.01449, 0.01674)
γ_h	Hour 15	(base – omitted)			
	Hour 16	-0.19026*	0.07320	0.009	(-0.33373, -0.04678)
	Hour 17	-0.09878	0.06932	0.154	(-0.23465, 0.03708)
	Hour 18	-0.67727*	0.06549	0.000	(-0.80563, -0.5489)
	Hour 19	-0.81217*	0.08128	0.000	(-0.97147, -0.65286)
	Hour 20	-1.89435*	0.16316	0.000	(-2.21416, -1.57454)
δ_h	Hour_15_x_Temp	(base – omitted)			
	Hour_16_x_Temp	0.00280*	0.00080	0.000	(0.00124, 0.00436)
	Hour_17_x_Temp	0.00245*	0.00075	0.001	(0.00097, 0.00392)
	Hour_18_x_Temp	0.00855*	0.00071	0.000	(0.00715, 0.00994)
	Hour_19_x_Temp	0.00937*	0.00088	0.000	(0.00765, 0.01109)
	Hour_20_x_Temp	0.02021*	0.00179	0.000	(0.01669, 0.02372)
α	Constant	-1.02412*	0.05224	0.000	(-1.12651, -0.92173)

Using the regression coefficients shown in Table 9, the Evergreen team created a time-temperature matrix (TTM) that shows expected load reductions (per device) for different outdoor temperatures and at different times of the day. The TTM is shown in Table 10. As noted, Residential DCU Power Saver DR events have historically been infrequent during hours ending 19 and 20, so the values in those columns are informed by fewer data points.

Table 10: Residential Time-Temperature Matrix

Temp	Hour Ending MDT					
	15	16	17	18	19	20
105	0.62	0.72	0.77	0.84	0.79	0.84
104	0.60	0.70	0.76	0.81	0.76	0.81
103	0.58	0.68	0.74	0.79	0.74	0.77
102	0.57	0.66	0.72	0.76	0.71	0.74
101	0.55	0.65	0.70	0.74	0.69	0.70
100	0.54	0.63	0.68	0.71	0.66	0.66
99	0.52	0.61	0.67	0.69	0.64	0.63
98	0.51	0.59	0.65	0.67	0.61	0.59
97	0.49	0.57	0.63	0.64	0.59	0.56
96	0.47	0.55	0.61	0.62	0.56	0.52
95	0.46	0.54	0.59	0.59	0.54	0.48
94	0.44	0.52	0.57	0.57	0.51	0.45
93	0.43	0.50	0.56	0.55	0.49	0.41
92	0.41	0.48	0.54	0.52	0.46	0.38
91	0.40	0.46	0.52	0.50	0.44	0.34
90	0.38	0.44	0.50	0.47	0.41	0.31
89	0.37	0.42	0.48	0.45	0.39	0.27
88	0.35	0.41	0.47	0.42	0.36	0.23
87	0.33	0.39	0.45	0.40	0.34	0.20
86	0.32	0.37	0.43	0.38	0.31	0.16
85	0.30	0.35	0.41	0.35	0.29	0.13



To get an idea of the Residential DCU resource capability on aggregate, the number of active devices can be multiplied by the values shown in Table 10. As of the end of summer 2019, there were 41,376 active residential devices. Thus, the expected aggregate impact of an event hour ending at 6:00 PM (MDT) when the outdoor temperature is 100 degrees would be 29.4 MW. Residential results are subject to an operability adjustment to better reflect the fact that not all devices in the population will be able to curtail load when called due to damage, wiring, or connection issues. The operability adjusted aggregate load is 86% of the unadjusted load, or 25.3 MW.

3 Small Commercial Results

For the Small Commercial customer class, usage during the curtailment event is compared to usage on high load days preceding the event. This section reviews the Small Commercial impacts calculated by Itron and validated by the Evergreen team. Additionally, the team provides feedback on the evaluation approach used by Itron and provides an alternative impact analysis for summer 2019 events. Finally, ex ante impacts, combining multiple years of event history are produced for various temperature scenarios.

3.1 Validation of Calculations

After receiving the participant load data from Itron, the Evergreen team attempted to reproduce the impacts in Itron's Power Saver impact evaluation report. We were unable to replicate impacts for the events on 8/26 and 8/27, but we did replicate impacts for the other days. Figure 8 compares impacts as calculated by Itron and by Evergreen at the 5-minute level. For the events on 8/26 and 8/27, we calculated very similar impacts but could not fully replicate the Itron values. A full summary of event hour impacts, per Itron's calculation method, is shown in Table 11. Itron's per device kW impact estimate for the Small Commercial class (1.21 kW) is the maximum fifteen-minute rolling average reduction during the qualifying event hours. (See Section 1.3 for more details.)

Figure 8: Small Commercial Impact Verification

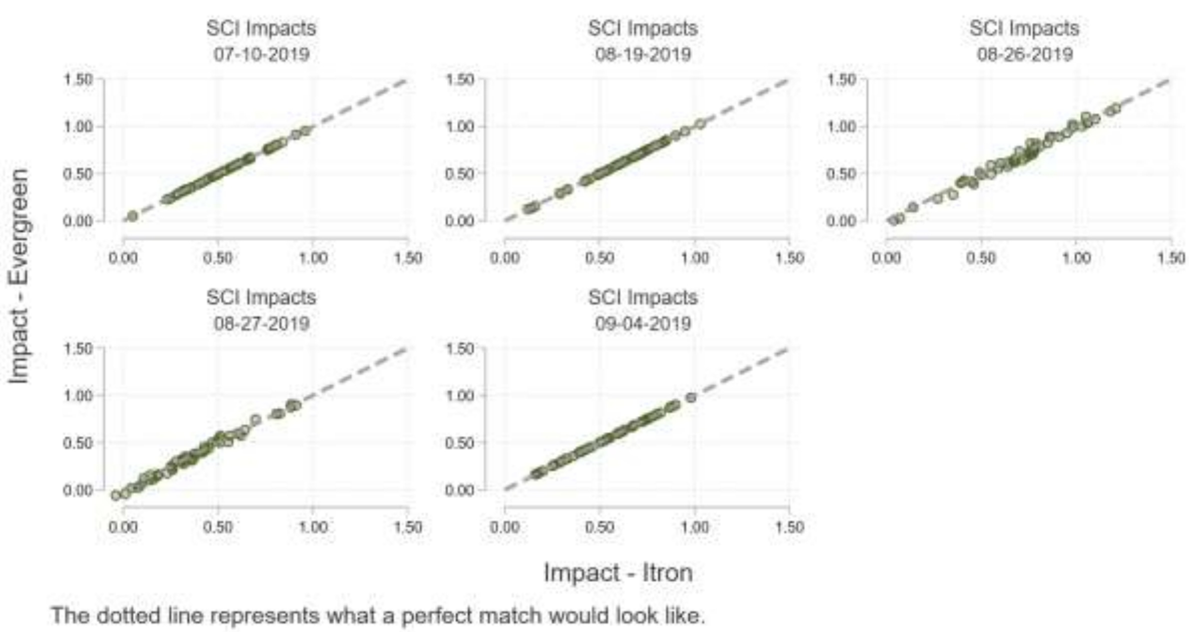


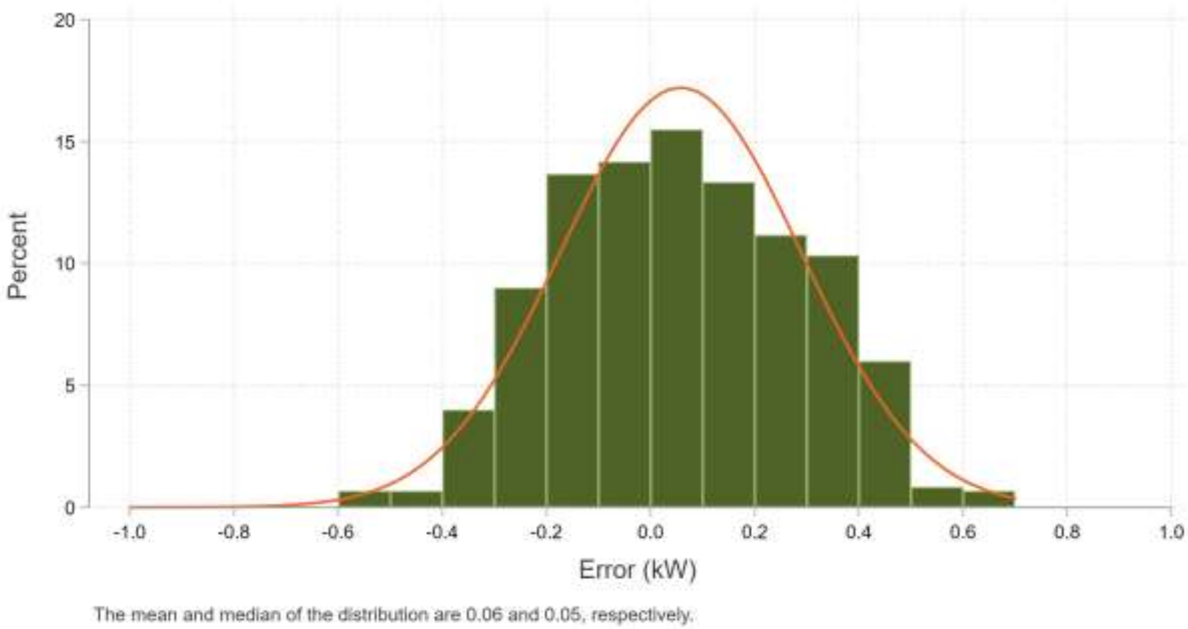
Table 11: Small Commercial Impact Estimates (kW) by Date and Time

Date	Hour Ending (MDT)				
	3:00 PM	4:00 PM	5:00 PM	6:00 PM	7:00 PM
07/10/2019		0.91	0.96*	0.78	0.76
08/19/2019	0.90	1.03	0.85	0.95	
08/26/2019	1.10	1.21*	0.91*	0.78*	
08/27/2019		0.91	0.81	0.56	0.38
09/04/2019		0.98	0.87	0.82	0.59

3.2 Baseline Accuracy

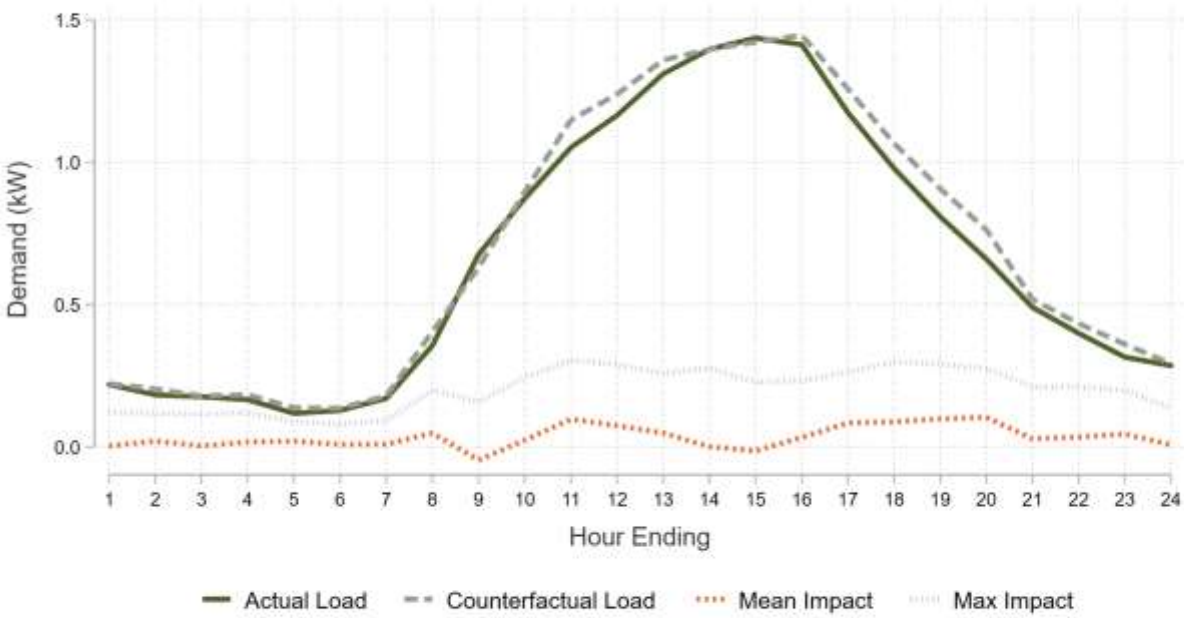
This section serves as a summary of the Evergreen team's assessment of the impact estimation methodology for the Small Commercial customer class. Specifically, we focus on the decision to use the maximum hourly impact as the per device kW impact estimate for the 2019 season by testing the accuracy of the selected CBL on non-event days. To this end, the Evergreen team used the method outlined in Section 1.3 to predict impacts during common event hours (hours ending 15-19) on event-like non-event days. Because there were no curtailment events on these event-like non-event days, the estimated baseline should mirror the actual load (or, more appropriately, the estimated 15-minute rolling baseline should mirror the 15-minute rolling average load), and the impacts should be centered around zero. Regarding the first point, the estimated load and the actual load line up well at the 5-minute interval level. Figure 9 shows the distribution of errors at the 5-minute level. The distribution is certainly centered around zero – the mean and median errors are 0.06 kW and 0.05 kW, respectively. Average demand on the proxy event days in the event window was 1.16 kW, meaning the errors are about 4-5% of demand.

Figure 9: Distribution of CBL Errors at the 5-Minute Level



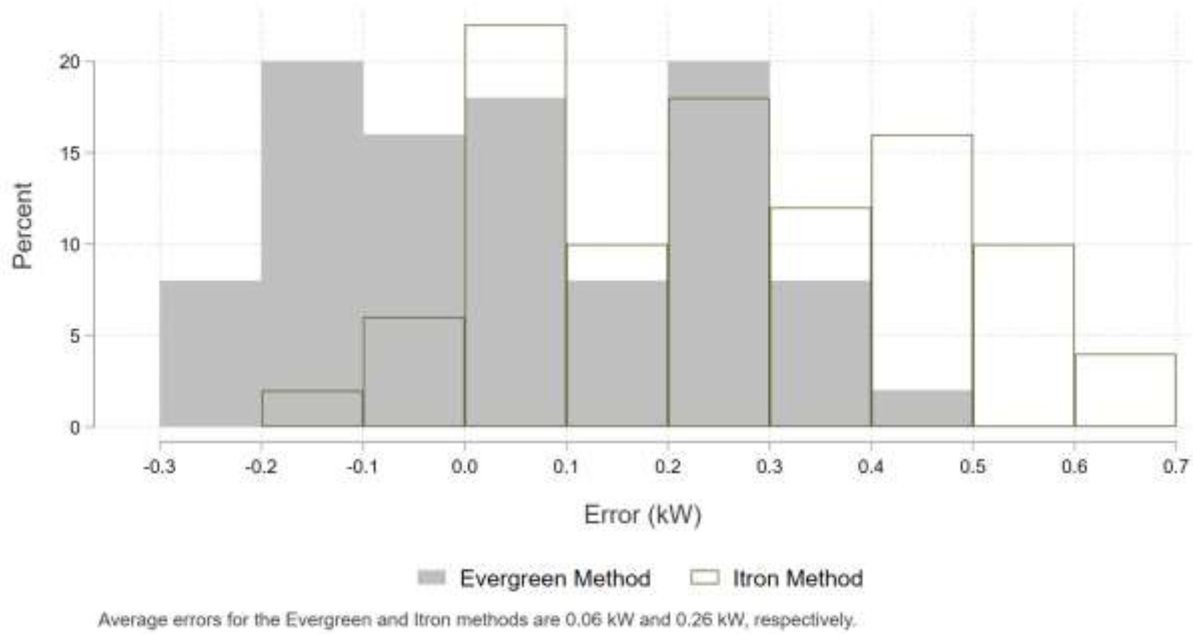
The next step in the baseline method is to select the maximum 5-minute impact per hour. This is the step that introduces significant bias into the baseline estimate. Figure 10 shows the average hourly load profile and the average hourly counterfactual across the proxy days. The figure also shows impacts under two scenarios: (1) take the average of the 5-minute impacts and (2) take the maximum of the 5-minute impacts. Though the actual and counterfactual load are quite similar, the impacts under scenario (2) range between 0.20 kW and 0.30 kW during common event hours. Relative to the actual load, this is an average upwards bias of about 25%.

Figure 10: Proxy Event Day Load and Counterfactual Load



Finally, Figure 11 shows the distribution of hourly impacts for common event hours across all proxy event days under the two scenarios. The gray bars represent impacts when the average 5-minute impact is selected; the translucent green bars represent impacts when the maximum 5-minute impact is selected. Recall that this distribution should be centered around zero. Using the maximum method, the distribution is centered around 0.26 kW – meaning that, even when no event occurred, that method will estimate an impact of 0.26 kW on average.

Figure 11: Distribution of Hourly Impacts On Proxy Event Days



The critical takeaway from this section is that the methodology Itron used in estimating impacts for the Small Commercial customer class will produce impact estimates that systematically overstate the true impact. On event-like non-event days, their method produces a load reduction estimate of 0.26 kW (which is 24% of the actual load on event-like non-event days) when in fact there was no load reduction at all. To reduce bias, the Evergreen team recommends using either the mean or the median in any place where the maximum is used.

3.3 Evergreen Ex Post Impacts

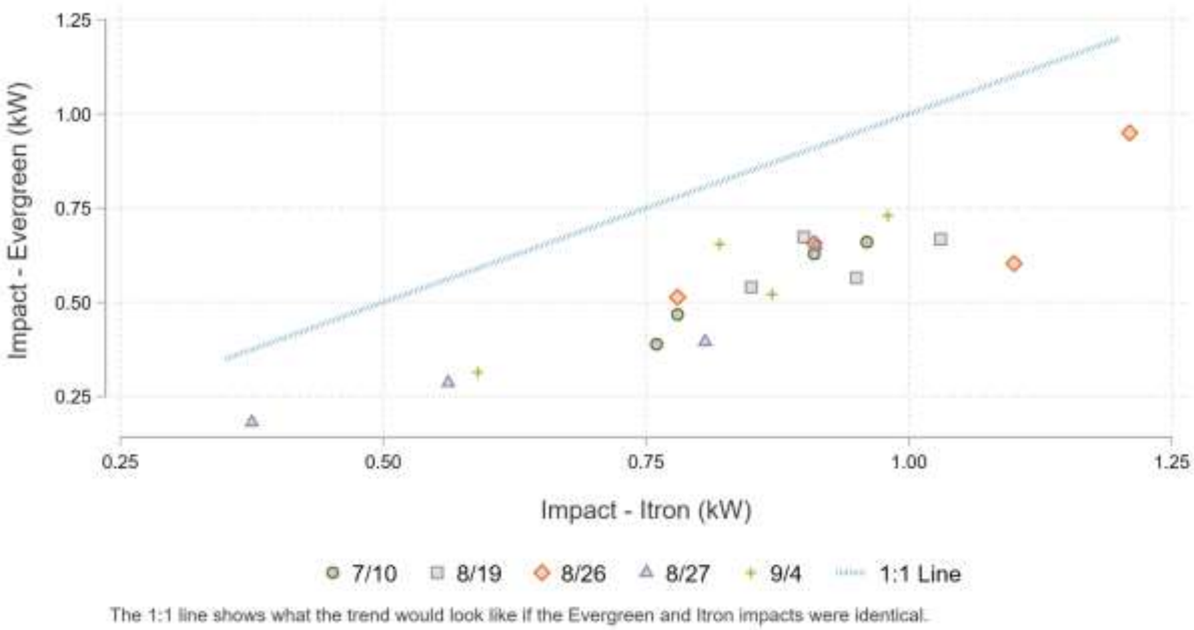
As discussed in the previous section, the Evergreen team thinks the method used to estimate impacts for the Small Commercial customer class overstates the true average impact. For each event hour during the 2019 DR season, Table 12 shows the estimates produced by the Evergreen team. Our methods differed from Itron's just slightly – in any place where a maximum was called for, we replaced it with the mean.

Table 12: Impact Calculations for the Small Commercial Segment

Date	# of Curtailed Customers	Hour Ending MDT	Temp.	CBL kW	Observed kW	Impact
7/10/2019	25	16	96	1.72	1.09	0.63
		17*	97	1.45	0.79	0.66
		18	96	1.17	0.71	0.47
		19	96	1.06	0.67	0.39
8/19/2019	24	15	94	1.56	0.89	0.67
		16	94	1.56	0.89	0.67
		17	94	1.40	0.86	0.54
		18	94	1.18	0.62	0.56
8/26/2019	23	15	96	1.69	1.09	0.60
		16*	97	1.81	0.86	0.95
		17*	98	1.51	0.85	0.66
		18*	97	1.14	0.63	0.51
8/27/2019	23	16	93	1.58	0.93	0.65
		17	95	1.31	0.92	0.40
		18	90	0.99	0.71	0.29
		19	87	0.83	0.64	0.18
9/4/2019	23	16	90	1.58	0.85	0.73
		17	90	1.36	0.83	0.52
		18	88	1.24	0.58	0.65
		19	88	0.81	0.49	0.31

The average difference during full event hours was 0.55 kW. Amongst full event hours, the average impact during qualified event hours was 0.69 kW. Figure 12 compares Evergreen's ex post hourly impacts with the impacts calculated by Itron. The Evergreen impact is lower in all cases, by about 0.30 kW on average.

Figure 12: Comparison of Evergreen Ex Post Impacts and Itron Impacts



3.3.1 Net Energy Savings

The Evergreen team estimated net energy impacts for the Small Commercial customer class by summing ex post impacts from the onset of each event through the end of the event day. The calculation of impacts is exactly as described earlier in this section. Table 13 shows the energy savings estimates (per device) for each event day. On average, net per device savings were 1.83 kWh per event day. However, two of the events (8/19 and 8/26) were called just for the M&V group – not the full Small Commercial DCU population. For the three population events, the average net savings per device was 1.46 kWh per day. Scaling by the number of events (three) and the number of active devices (3,443 per Itron's report) yields an aggregate savings estimate of 15.0 MWh for the Small Commercial DCU segment. If M&V energy savings for 8/19 and 8/26 are added in, the total is 15.2 MWh.

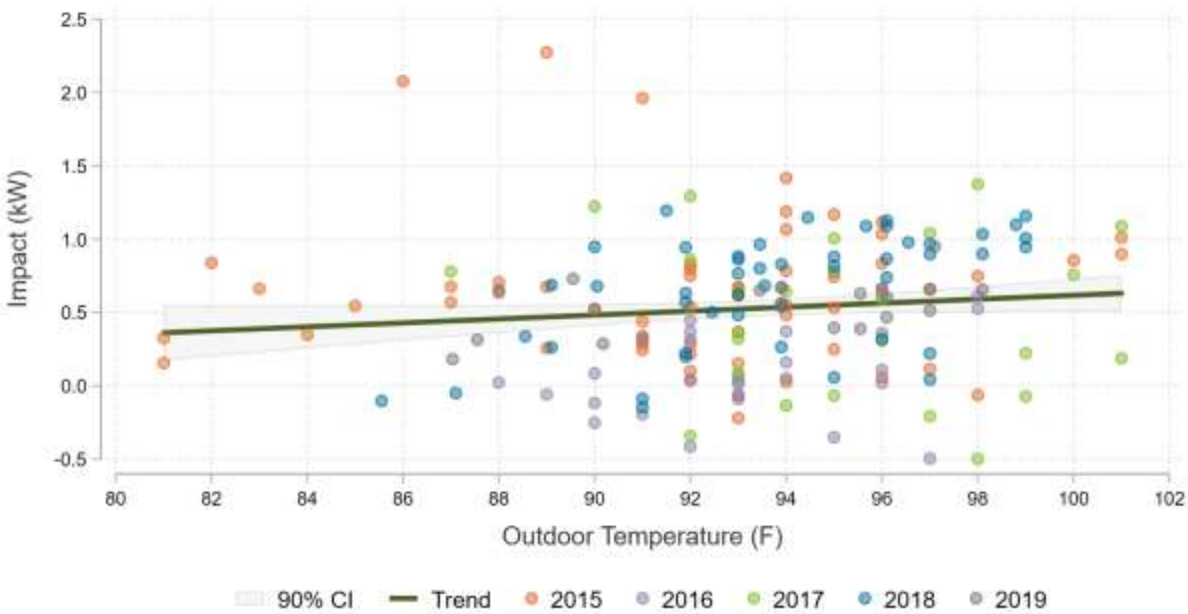
Table 13: Per Device Energy Savings by Event Day

Date	Event Start (MDT)	Event Savings (kWh)	Snapback (kWh)	Net Savings (kWh)
7/10/2019	3:00 PM	2.14	-0.78	1.36
8/19/2019	2:00 PM	2.45	0.16	2.61
8/26/2019	2:00 PM	2.72	-0.55	2.17
8/27/2019	3:00 PM	1.51	-0.42	1.09
9/4/2019	3:00 PM	2.22	-0.30	1.92
Average		**2.21**	**-0.38**	**1.83**

3.4 Evergreen Ex Ante Impacts

Figure 13 compares 2015-2019 ex post impact estimates for each event hour with the outdoor air temperature for that hour. (Weather data comes from weather station KABQ in Albuquerque.) The trend in temperature is quite subtle; there are only slight increases in impact magnitude as temperature increases. To develop an ex ante impact estimate, the Evergreen team developed a regression model that estimates the ex post impact as a function of temperature and time. The specified model was shown in Section 1.5, and the results from the model are described in more detail below. Using the model, the Evergreen team predicts that the impact of a Small Commercial DR event at peaking conditions (5:00 PM – 6:00 PM MDT when outdoor temperature is 100 degrees) is 0.52 kW per device (compared to an ex post impact of 0.69 kW).

Figure 13: Hourly Impacts against Outdoor Temperature (F)



The regression was run on full event hours (some events in prior summers started mid-hour) and weighted by the number of curtailed devices (each summer had slightly different numbers of dispatched devices). Regression output is shown below. In general, earlier hours corresponded to higher kW values, with a drop over time in impacts as less load was available to shed. It should be noted that Hour 20 was extremely rare; only 3 events during the past four years included a full-hour event during this period and as such, should be interpreted with care. Temperature has a negative coefficient, indicating that higher temperatures produce lower impacts after accounting for the hour and the interaction between temperature and time. The interaction terms, represented by δ_h, are all positive, indicating that the incremental effect of temperature in a given hour increases the impact. Again, Hour 20 should be interpreted with caution as only three data points were available to fit the model. Note that any coefficient with * next to it is statistically significant. Due to the small sample sizes and year-to-year variability, none of the estimates in this regression are statistically significant.

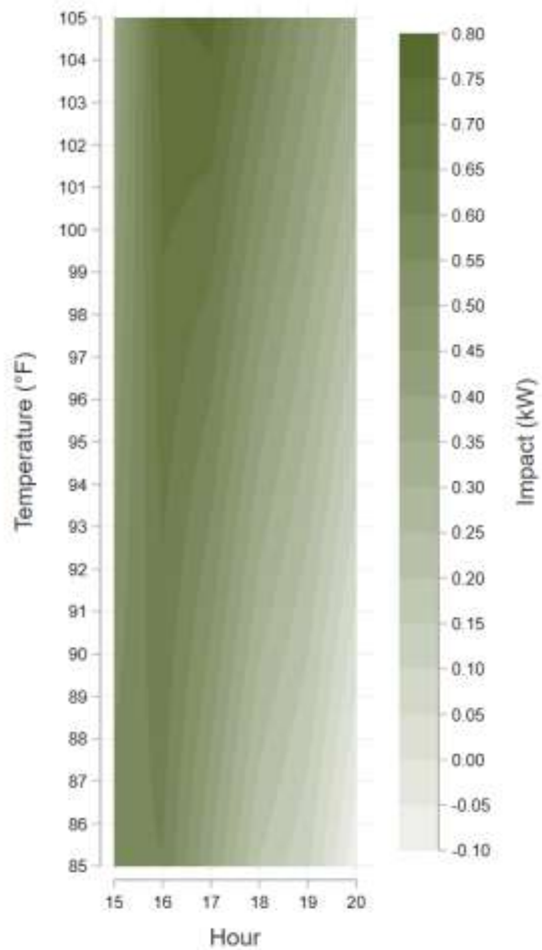
Table 14: Small Commercial Ex Ante Regression Output

Term	Variable	Coefficient (b)	Standard Error	P-Value	95% CI
β	Temperature	-0.00970	0.02253	0.667	(-0.05418, 0.03478)
γ_h	Hour 15	(base – omitted)			
	Hour 16	-1.41112	2.90998	0.628	(-7.15621, 4.33397)
	Hour 17	-2.58001	2.71863	0.344	(-7.94733, 2.78731)
	Hour 18	-3.00920	2.57421	0.244	(-8.09138, 2.07299)
	Hour 19	-2.88772	3.08474	0.351	(-8.97784, 3.20239)
	Hour 20	-3.41079	5.94628	0.567	(-15.15036, 8.32878)
δ_h	Hour_15_x_Temp	(base – omitted)			
	Hour_16_x_Temp	0.01682	0.03159	0.595	(-0.04554, 0.07918)
	Hour_17_x_Temp	0.02820	0.02948	0.340	(-0.03, 0.08639)
	Hour_18_x_Temp	0.03091	0.02795	0.270	(-0.02427, 0.08609)
	Hour_19_x_Temp	0.02842	0.03346	0.397	(-0.03764, 0.09448)
	Hour_20_x_Temp	0.03223	0.06577	0.625	(-0.09761, 0.16207)
α	Constant	1.40469	2.05086	0.494	(-2.64427, 5.45364)

Using the regression coefficients shown in Table 14, the Evergreen team created a time-temperature matrix (TTM) that shows expected load reductions (per device) for different outdoor temperatures and at different times of the day. The TTM is shown in Table 15. These results should be interpreted with caution due to their small sample sizes. The use of hour ending 18 as the "peak" is meaningful here. At hour ending 16, the kW impact at 100°F is 0.71 – slightly larger than the ex post estimate.

Table 15: Small Commercial Time-Temperature Matrix

Temp	Hour Ending MDT					
	15	16	17	18	19	20
105	0.39	0.74	0.77	0.62	0.48	0.36
104	0.40	0.73	0.75	0.60	0.46	0.34
103	0.41	0.73	0.73	0.58	0.44	0.31
102	0.42	0.72	0.71	0.56	0.43	0.29
101	0.42	0.71	0.69	0.54	0.41	0.27
100	0.43	0.71	0.67	0.52	0.39	0.25
99	0.44	0.70	0.66	0.49	0.37	0.22
98	0.45	0.69	0.64	0.47	0.35	0.20
97	0.46	0.68	0.62	0.45	0.33	0.18
96	0.47	0.68	0.60	0.43	0.31	0.16
95	0.48	0.67	0.58	0.41	0.30	0.13
94	0.49	0.66	0.56	0.39	0.28	0.11
93	0.50	0.66	0.54	0.37	0.26	0.09
92	0.51	0.65	0.53	0.35	0.24	0.07
91	0.52	0.64	0.51	0.33	0.22	0.04
90	0.53	0.63	0.49	0.30	0.20	0.02
89	0.54	0.63	0.47	0.28	0.18	0.00
88	0.55	0.62	0.45	0.26	0.16	-0.02
87	0.56	0.61	0.43	0.24	0.15	-0.05
86	0.57	0.61	0.42	0.22	0.13	-0.07
85	0.58	0.60	0.40	0.20	0.11	-0.09



To get an idea of the Small Commercial resource capability on aggregate, the number of active devices can be multiplied by the values shown in Table 15. As of the end of summer 2019, there were 3,443 active small commercial devices. Thus, the expected aggregate impact of an event hour ending at 6:00 PM (MDT) when the outdoor temperature is 100 degrees would be 1.8 MW. Adjusted for operability, the aggregate impact is 1.5 MW.

4 Medium Commercial

For the Medium Commercial customer class, usage during the curtailment event is compared to usage on high load days preceding the event. The remainder of this section provides greater detail on how the Evergreen team attempted to validate Itron's calculations, as well as a discussion of ex post and ex ante impacts and baseline accuracy.

4.1 Validation of Calculations

After receiving the participant load data from Itron, the Evergreen team attempted to reproduce the impacts in Itron's Power Saver impact evaluation report. For each event hour on 7/10 and 8/27, the Evergreen team was able to replicate Itron's impact estimates for the medium commercial customer class using the top 3/5 baseline method. We were unable to replicate impacts for the 9/4 event. Figure 14 compares impacts as calculated by Itron and by Evergreen at the 5-minute level. For reference, medium commercial impact estimates are shown in Table 16. Note that an asterisk (*) denotes a qualified event hour. The maximum impact during qualified event hours was 2.89 kW per facility for this class.

Figure 14: Medium Commercial Impact Verification

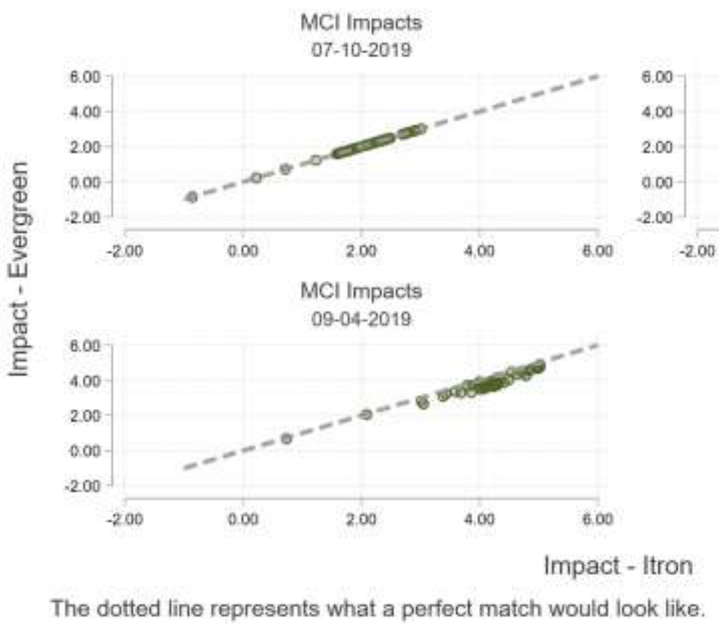


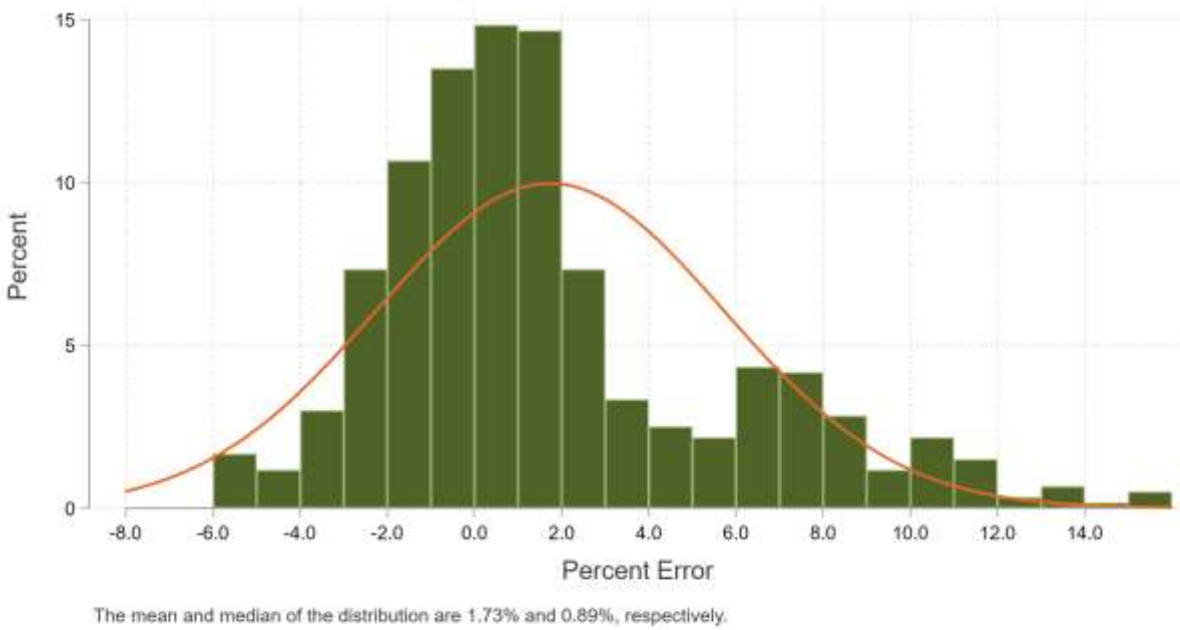
Table 16: Medium Commercial Impact Estimates (kW) by Date and Time

Date	Hour Ending (MDT)			
	4:00 PM	5:00 PM	6:00 PM	7:00 PM
07/10/2019	2.45	2.89*	3.01	1.83
08/27/2019	5.57	4.34	5.63	2.54
09/04/2019	5.02	5.02	4.78	4.32

4.2 Baseline Accuracy

This section serves as a summary of the Evergreen team's assessment of the impact estimation methodology for the Medium Commercial customer class. Specifically, we focus on the decision to use the maximum hourly impact as the per device kW impact estimate for the 2019 season by testing the accuracy of the selected CBL on non-event days. To this end, the Evergreen team used the method outlined in Section 1.3 to predict impacts during common event hours (hours ending 15-19) on event-like non-event days. Because there were no curtailment events on these event-like non-event days, the estimated baseline should mirror the actual load (or, more appropriately, the estimated 15-minute rolling baseline should mirror the 15-minute rolling average load), and the impacts should be centered around zero. Regarding the first point, the estimated load and the actual load line up well at the 5-minute interval level. Figure 15 shows the distribution of percent errors at the 5-minute level (where percent error is calculated as impact / actual load * 100%). The distribution has a slight right skew but is largely centered near zero – the mean and median are 1.73% and 0.89% respectively.

Figure 15: Distribution of Errors at the 5-Minute Level



The next step in the baseline method is to select the maximum 5-minute impact per hour. This is the step that introduces significant bias into the baseline estimate. Figure 16 shows the average hourly load profile and the average hourly counterfactual across the proxy days. The figure also shows impacts under two scenarios: (1) take the average of the 5-minute impacts and (2) take the maximum of the 5-minute impacts. Though the actual and counterfactual load are quite similar, the average impacts during common event hours under scenarios (1) and (2) are 1.00 kW and 2.12 kW. These errors translate to a 1.7% error and a 3.6% error relative to average demand (Figure 17).

Figure 16: Proxy Event Day Load and Counterfactual Load

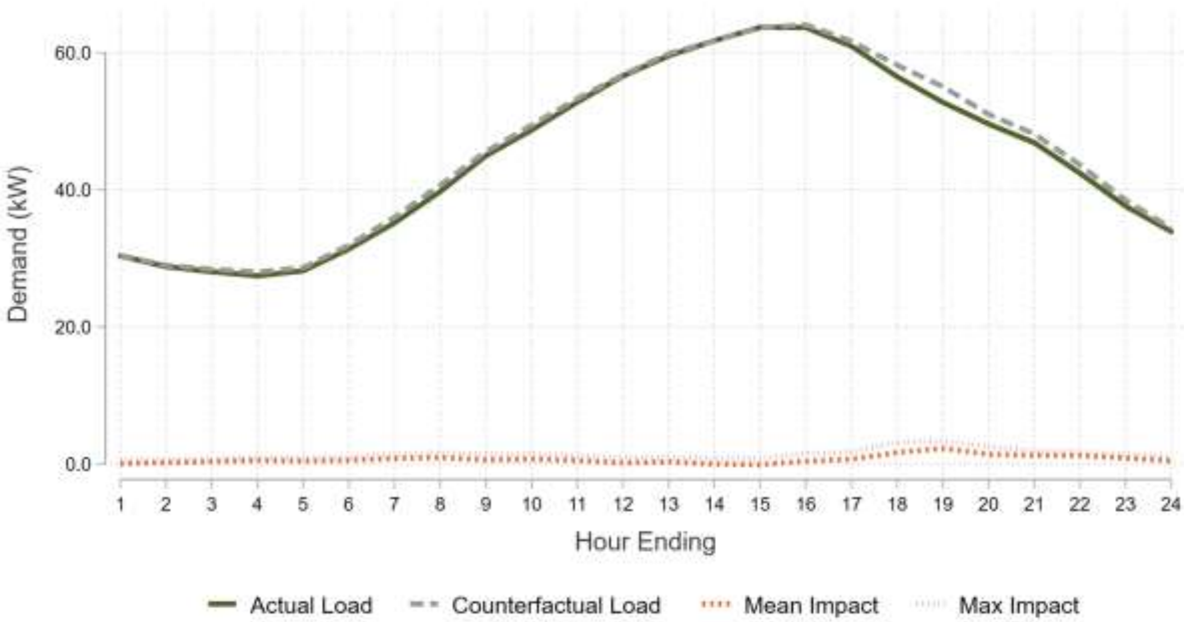
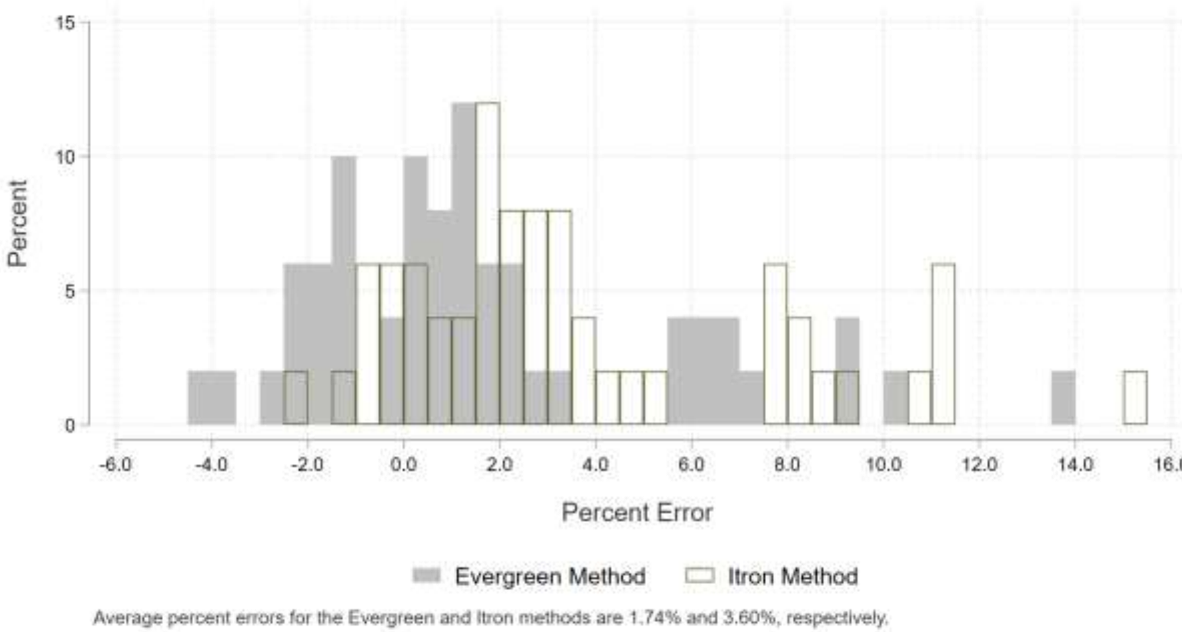


Figure 17: Distribution of Hourly Impacts On Proxy Event Days



The critical takeaway from this section is that the methodology Itron used in estimating impacts for the Medium Commercial customer class will produce impact estimates that systematically overstate the true impact. To reduce bias, the Evergreen team recommends using either the mean or the median in any place where the maximum is used.

4.3 Evergreen Ex Post Impacts

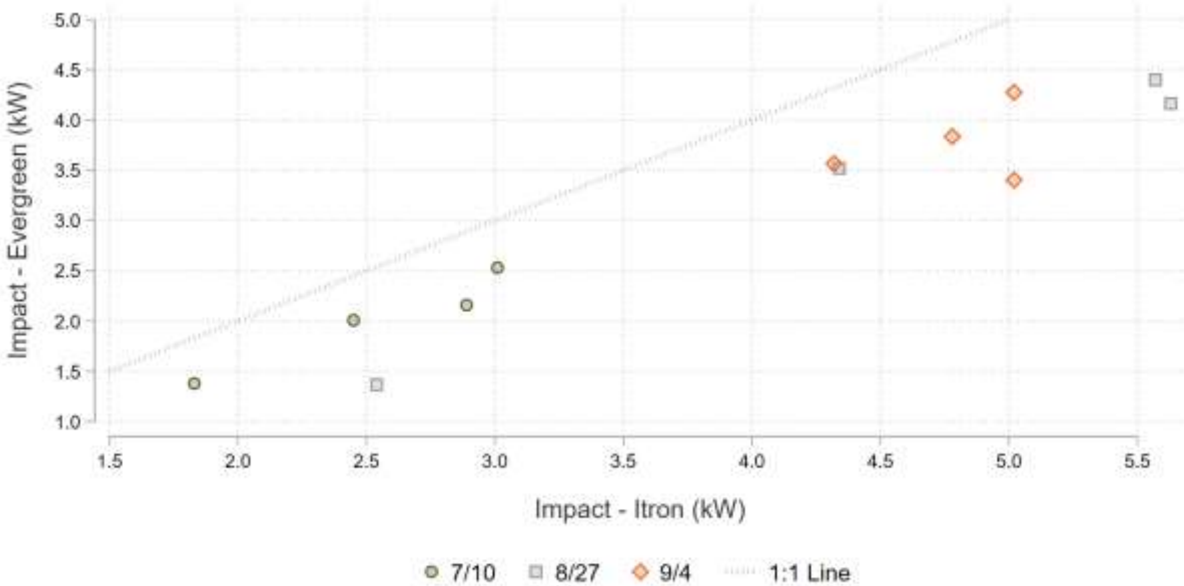
As discussed in the previous section, the Evergreen team thinks the method used to estimate impacts for the Medium Commercial customer class overstates the true average impact. For each event hour during the 2019 DR season, Table 17 shows the estimates produced by the Evergreen team. Our methods differed from Itron's just slightly – in any place where a maximum was called for, we replaced it with the mean.

Table 17: Medium Commercial Impact Results

Date	# of Curtailed Customers	Hour Ending MDT	Temp.	Reference kW	Observed kW	Impact (kW)
7/10/2019	51	16	96	60.23	58.22	2.01
		17*	97	59.60	57.44	2.16
		18	96	57.86	55.33	2.53
		19	96	55.39	54.01	1.38
8/27/2019	51	16	93	70.67	66.27	4.40
		17	95	66.40	62.88	3.52
		18	90	60.85	56.69	4.17
		19	87	54.48	53.11	1.37
9/4/2019	51	16	90	69.33	65.92	3.40
		17	90	66.18	61.91	4.28
		18	88	61.76	57.92	3.84
		19	88	57.38	53.81	3.57

Our reduction estimate is the average of the values in the 'Impact' column during qualified event hours, which is 2.16 kW, compared to 3.05 kW overall. Figure 18 compares Evergreen's ex post hourly impacts with the impacts calculated by Itron. The Evergreen impact is lower in all cases, by about 0.90 kW on average. It is important to note that these impacts are per facility, not per device. Itron notes that there were 2,636 devices installed at 369 facilities at the end of the 2019 DR season, indicating there were approximately 7.14 devices per facility. Thus, Evergreen's per-device estimate during qualified hours is 0.30 kW.

Figure 18: Comparison of Evergreen Ex Post Impacts and Itron Impacts



The 1:1 line shows what the trend would look like if the Evergreen and Itron impacts were identical.

4.3.1 Net Energy Savings

The Evergreen team estimated net energy impacts for the Medium Commercial customer class by summing ex post impacts from the onset of each event through the end of the event day. The calculation of impacts is exactly as described earlier in this section. Table 18 shows the energy savings estimates (per facility) for each event day. On average, net per facility savings were 11.19 kWh per event day. Across the three event days, this means there were 33.57 kWh of energy savings per facility or 4.70 kWh per device. Multiplying this estimate by the number of active devices (2,636 per Itron's report) yields an aggregate savings estimate of 12.4 MWh for the Medium Commercial customer class.

Table 18: Per Facility Energy Savings by Event Day

Date	Event Start (MDT)	Event Savings (kWh)	Snapback (kWh)	Net Savings (kWh)
7/10/2019	3:00 PM	8.08	-0.74	7.34
8/27/2019	3:00 PM	13.45	-5.70	7.75
9/4/2019	3:00 PM	15.08	3.40	18.48
Average		**12.20**	**-1.01**	**11.19**

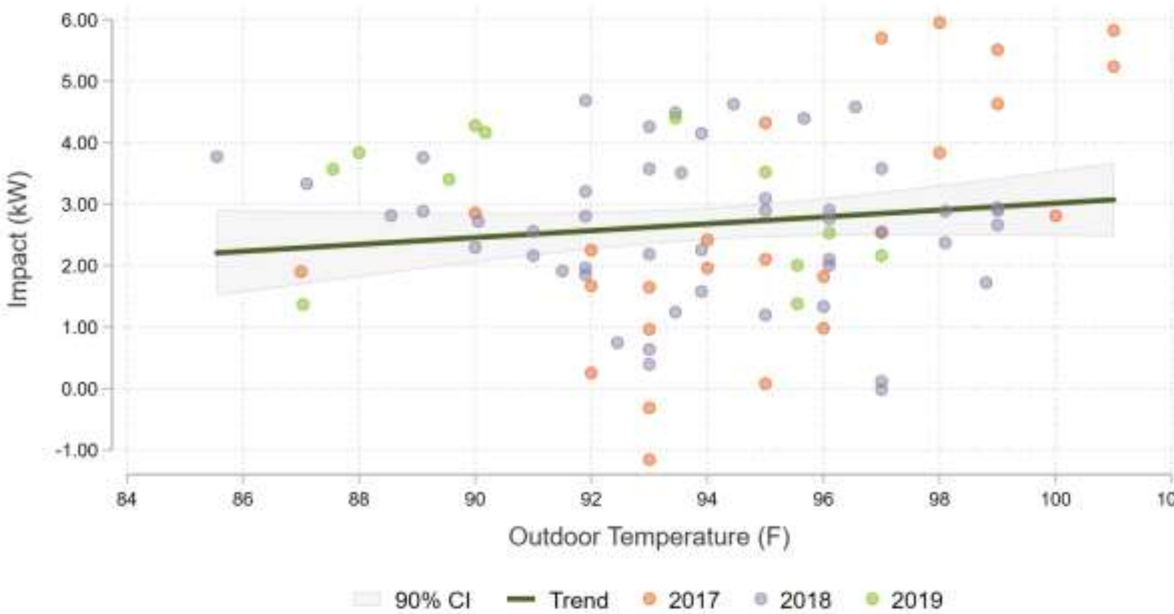
4.4 Evergreen Ex Ante Impacts

The method used by the Evergreen team to calculate ex post impacts for 2019 was the same as what was used in prior years – a baseline method. This allows us to compare impacts across years and use additional data to predict what the program can deliver in terms of load reduction under different planning scenarios.

Figure 19 compares 2017-2019 ex post impact estimates for each event hour with the outdoor air temperature for that hour. (Weather data comes from weather station KABQ in Albuquerque.) The trend in temperature is quite small but positive; impact magnitudes increase as temperature increases. To develop an ex ante impact estimate, the Evergreen team developed a regression model that estimates the ex post impact as a function of temperature and time. The specified model was shown in Section 1.5, and the results from the model are described in more detail below. Using the model, the Evergreen team predicts that the impact of a Medium Commercial DR event at peaking conditions (5:00 PM – 6:00 PM MDT when outdoor temperature is 100 degrees) is 2.82 kW per facility, or 0.39 kW per device (compared to ex post impacts of 2.16 kW per facility and 0.30 kW per device).

It is interesting to note that the 2018 and 2019 load impacts did not actually demonstrate much temperature sensitivity, while 2017 impacts did, in a way that was much more dramatic than what was observed with small commercial customers. With a small sample and large, variable customer loads, any change in sample composition can dramatically affect the overall result, meaning that any trends should be observed with caution.

Figure 19: Hourly Impacts against Outdoor Temperature (F)



The ex ante regression model was run on full event hours (some events in prior summers started mid-hour) and weighted by the number of curtailed devices (each summer had slightly different numbers of dispatched devices). Regression output is shown below. There is no clear relationship between event hour and impact. It should be noted that Hour 20 was extremely rare; only 2 events during the past three years included a full-hour event during this period. Temperature has a positive coefficient, indicating that higher temperatures produce higher impacts. The interaction terms, represented by δ_h, are all negative, indicating that the incremental effect of temperature in a given hour actually decreases the impact. Again, Hour 20 should be interpreted with caution as only two data points were available to fit the model. Note that any coefficient with * next to it is statistically significant. Due to the small sample sizes and year-to-year variability, none of the estimates in this regression are statistically significant.

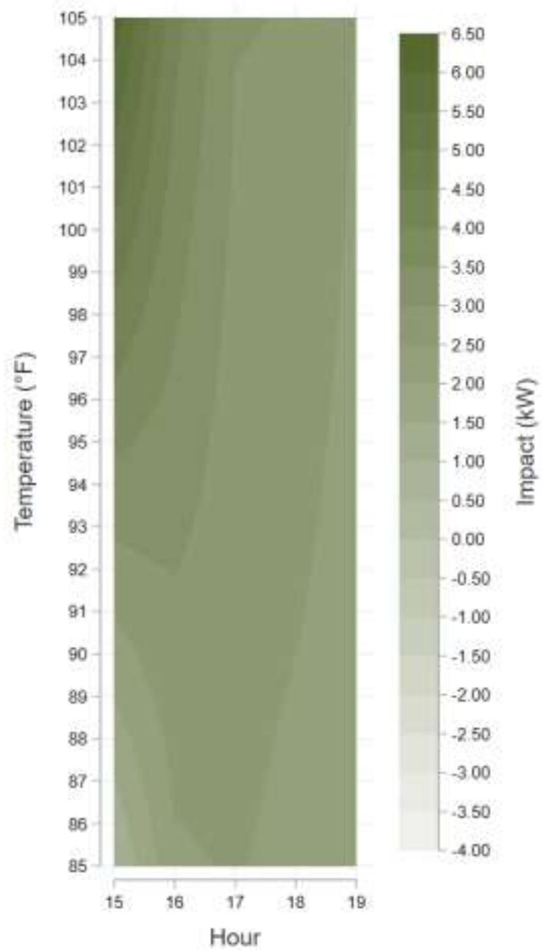
Table 19: Medium Commercial Ex Ante Regression Output

Term	Variable	Coefficient (b)	Standard Error	P-Value	95% CI
β	Temperature	0.26661	0.27088	0.328	(-0.274, 0.807)
γ_h	Hour 15	(base – omitted)			
	Hour 16	16.60100	27.54790	0.549	(-38.342, 71.544)
	Hour 17	22.21209	27.12311	0.416	(-31.883, 76.307)
	Hour 18	21.37336	26.16782	0.417	(-30.817, 73.563)
	Hour 19	22.69612	26.07260	0.387	(-29.304, 74.696)
	Hour 20	58.51108	31.77044	0.070	(-4.853, 121.875)
δ_h	Hour_15_x_Temp	(base – omitted)			
	Hour_16_x_Temp	-0.17813	0.30007	0.555	(-0.777, 0.42)
	Hour_17_x_Temp	-0.24246	0.29540	0.415	(-0.832, 0.347)
	Hour_18_x_Temp	-0.23497	0.28527	0.413	(-0.804, 0.334)
	Hour_19_x_Temp	-0.25214	0.28454	0.379	(-0.82, 0.315)
	Hour_20_x_Temp	-0.65463	0.34897	0.065	(-1.351, 0.041)
α	Constant	-21.71924	24.75628	0.383	(-71.094, 27.656)

Using the regression coefficients shown in Table 19, the Evergreen team created a time-temperature matrix (TTM) that shows expected load reductions (per device) for different outdoor temperatures and at different times of the day. The TTM is shown in Table 20. These results should be interpreted with caution due to their small sample sizes.

Table 20: Medium Commercial Time-Temperature Matrix

Temp	Hour Ending MDT				
	15	16	17	18	19
105	6.27	4.17	3.03	2.98	2.50
104	6.01	4.08	3.00	2.94	2.48
103	5.74	3.99	2.98	2.91	2.47
102	5.47	3.91	2.96	2.88	2.45
101	5.21	3.82	2.93	2.85	2.44
100	4.94	3.73	2.91	2.82	2.42
99	4.68	3.64	2.88	2.79	2.41
98	4.41	3.55	2.86	2.75	2.39
97	4.14	3.46	2.84	2.72	2.38
96	3.88	3.38	2.81	2.69	2.37
95	3.61	3.29	2.79	2.66	2.35
94	3.34	3.20	2.76	2.63	2.34
93	3.08	3.11	2.74	2.60	2.32
92	2.81	3.02	2.71	2.56	2.31
91	2.54	2.93	2.69	2.53	2.29
90	2.28	2.84	2.67	2.50	2.28
89	2.01	2.76	2.64	2.47	2.26
88	1.74	2.67	2.62	2.44	2.25
87	1.48	2.58	2.59	2.41	2.24
86	1.21	2.49	2.57	2.37	2.22
85	0.94	2.40	2.55	2.34	2.21



To get an idea of Medium Commercial resource capability on aggregate, the number of active facilities can be multiplied by the values shown in Table 20. As of the end of summer 2019, there were 369 active medium commercial facilities. Thus, the expected aggregate impact of an event hour ending at 6:00 PM (MDT) when the outdoor temperature is 100 degrees would be 1.0 MW. Adjusted for operability using the 86% adjustment factor, this aggregate impact is 0.89 MW.

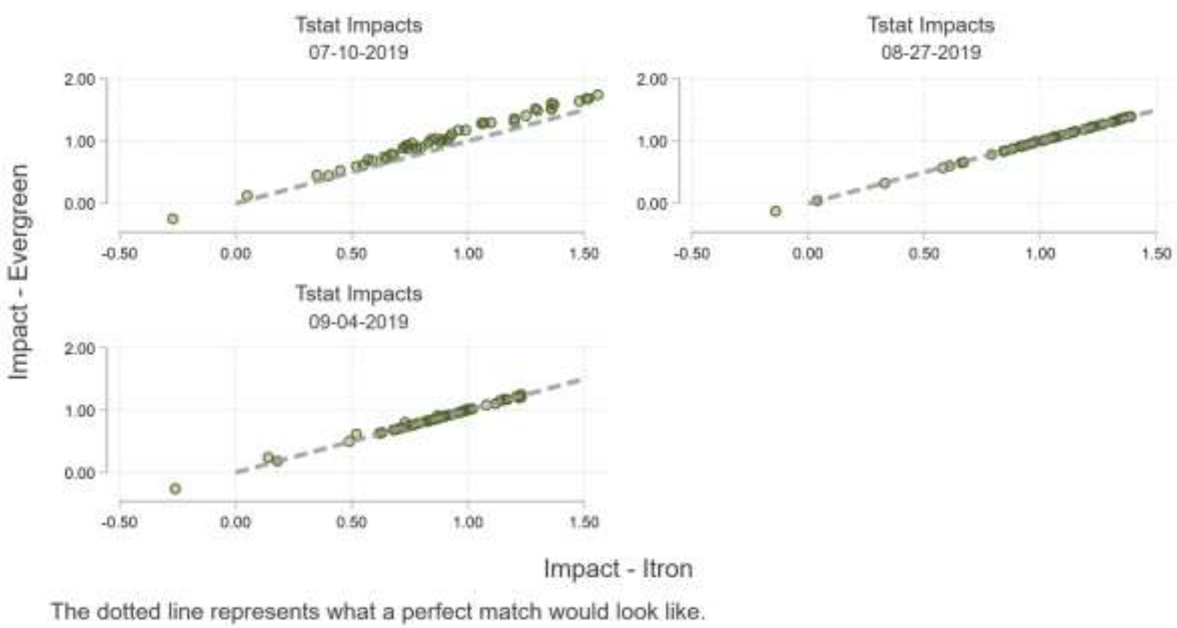
5 Two-Way Smart Thermostat

This section reviews the Two-Way Smart Thermostat impacts calculated by Itron and validated by the Evergreen team. Additionally, the team provides feedback on the evaluation approach used by Itron and provides an alternative impact analysis for summer 2019 events. Finally, ex ante impacts are produced for various temperature scenarios.

5.1 Validation of Calculations

After receiving the participant load data from Itron, the Evergreen team attempted to reproduce the impacts in Itron's Power Saver impact evaluation report. For each event hour other than the hours on 7/10, the Evergreen team was able to replicate Itron's impact estimates for the Two-Way Smart Thermostat segment.[8] Figure 20 compares impacts as calculated by Itron and by Evergreen at the 5-minute level. For reference, Itron's Residential DCU impact estimates are shown in Table 21. Note that an asterisk (*) denotes a qualified event hour. The maximum impact during the single qualified event hour was 0.80 kW for the Two-Way Smart Thermostat segment.

Figure 20: Two-Way Smart Thermostat Impact Verification



[8] To replicate impacts on 8/27, we had to shift all 5-minute records forward in time. To replicate impacts on 9/4, we had to shift all 5-minute records backwards in time.

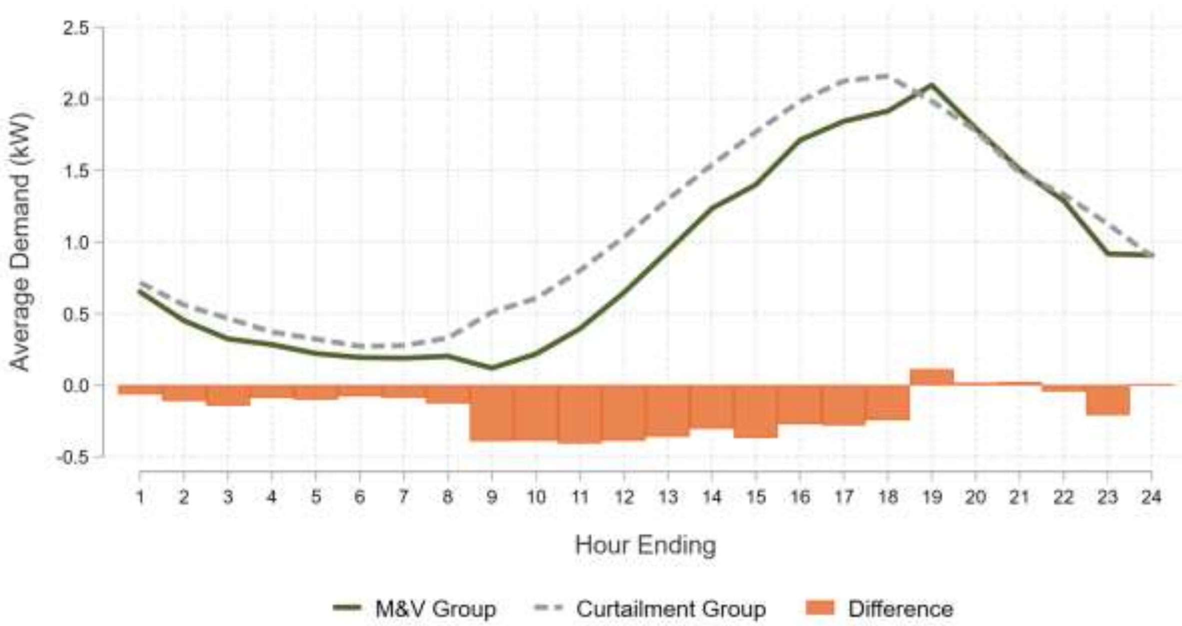
Table 21: Two-Way Smart Thermostat Impact Estimates (kW) by Date and Time

Date	Hour Ending (MDT)			
	4:00 PM	5:00 PM	6:00 PM	7:00 PM
07/10/2019	0.92	0.80*	1.20	1.56
08/27/2019	1.22	1.39	1.26	1.19
09/04/2019	0.90	1.16	1.02	1.23

5.2 Evergreen Ex Post Impacts

For the Two-Way Smart Thermostat segment, Itron's per device kW impact estimate for the 2019 season is the maximum difference between 5-minute rolling average loads for the control and curtailment groups (0.69 kW). (See Section 1.1 for more details.) Though we are generally opposed to using the maximum, the true issue with this segment is the difference in load profiles between the curtailment group and the M&V group (Figure 21). Using the M&V group as a proxy for what the curtailment group load would have looked like absent DR will systematically understate the impacts.

Figure 21: Smart Thermostat Two-Way Load Shapes on Event-Like Days



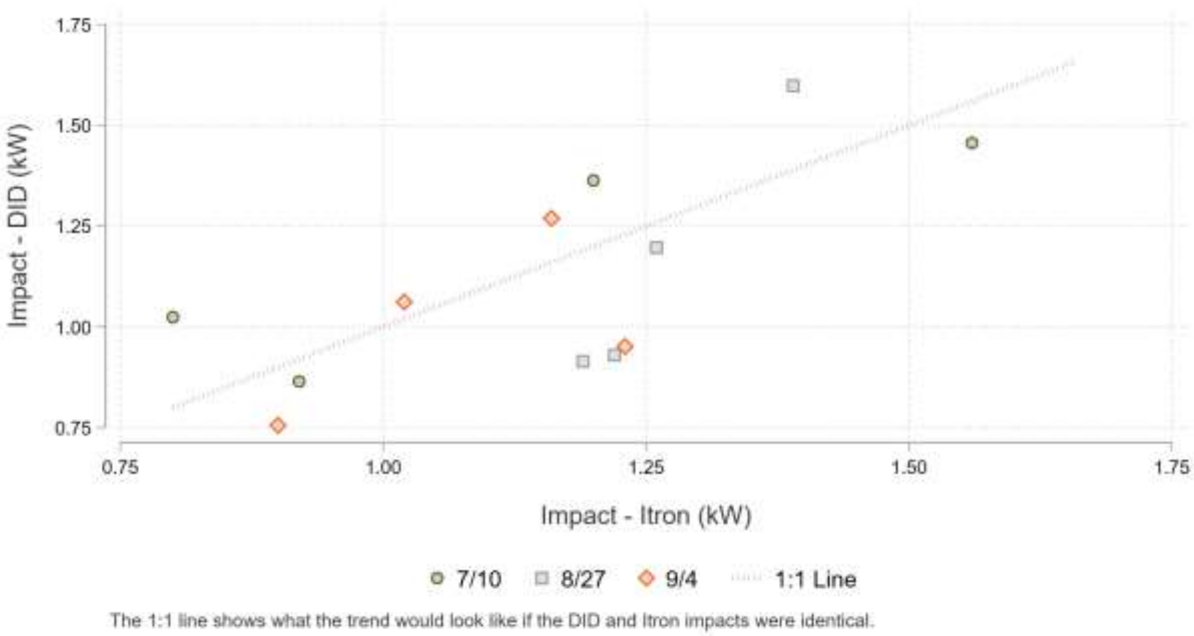
Like the Residential DCU segment, the Evergreen team used a difference-in-differences approach when evaluating the Two-Way Smart Thermostat program. This approach accounts for the large differences between the curtailment group and the M&V on event-like non-event days. An example of this approach was presented in Section 1.2. Results for the 2019 DR season are summarized in Table 22.

Table 22: Impact Calculations for the Two-Way Smart Thermostat Segment

Date	# of Curtailed Devices	Hour Ending MDT	Temp. (F)	Control kW	Curtail kW	Non-Event Diff. (kW)	Impact (kW)
7/10/2019	127	16	96	1.57	0.97	-0.27	0.86
		17*	97	1.41	0.67	-0.28	1.02
		18	96	1.82	0.70	-0.24	1.36
		19	96	2.35	0.78	0.11	1.46
8/27/2019	113	16	93	1.81	1.15	-0.27	0.93
		17	95	2.05	0.73	-0.28	1.60
		18	90	1.73	0.78	-0.24	1.20
		19	87	1.84	0.81	0.11	0.91
9/4/2019	123	16	90	1.47	0.98	-0.27	0.76
		17	90	1.69	0.70	-0.28	1.27
		18	88	1.59	0.77	-0.24	1.06
		19	88	1.86	0.80	0.11	0.95

The average impact during event hours was 1.12 kW. Amongst qualified event hours, the average impact was 1.02 kW. Figure 22 compares Evergreen's ex post hourly impacts with the impacts calculated by Itron. Despite the different methods, the impacts were similar, on average. The biases in Itron's method largely cancelled each other out – using the maximum 5-minute impact overstates the true impact, but the difference in average demand between the curtailment and M&V groups understates the impact.

Figure 22: Comparison of Evergreen Ex Post Impacts and Itron Impacts



5.2.1 Net Energy Savings

The Evergreen team estimated net energy impacts for the Two-Way Smart Thermostat segment by summing ex post impacts from the onset of each event through the end of the event day. The calculation of impacts is exactly as described earlier in this section. Table 23 shows the energy savings estimates (per device) for each event day. On average, net per device savings were 2.18 kWh per event day. Across the three event days, this means there were 6.55 kWh of energy savings per device. Multiplying this estimate by the number of active devices (384 per Itron's report) yields an aggregate savings estimate of 2.5 MWh for the Two-Way Smart Thermostat segment.

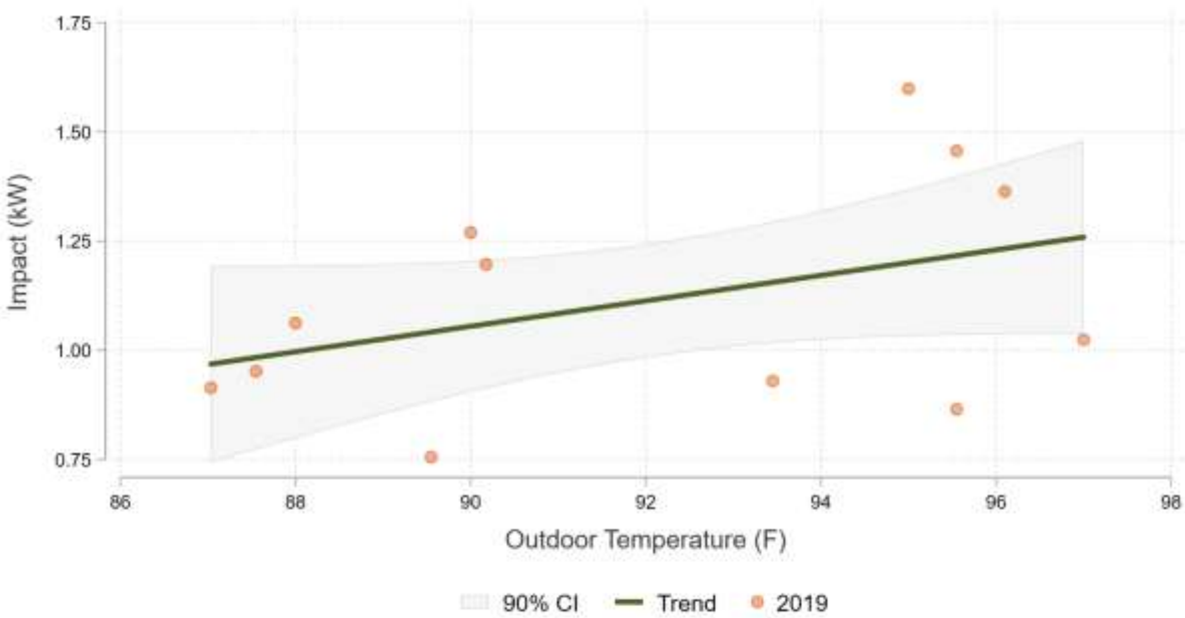
Table 23: Per Device Energy Savings by Event Day

Date	Event Start (MDT)	Event Savings (kWh)	Snapback (kWh)	Net Savings (kWh)
7/10/2019	3:00 PM	4.71	-2.25	2.46
8/27/2019	3:00 PM	4.64	-2.46	2.18
9/4/2019	3:00 PM	4.04	-2.13	1.91
Average		**4.46**	**-2.28**	**2.18**

5.3 Evergreen Ex Ante Impacts

Unlike the other segments, multiple years of ex post results are not available for the Two-Way Smart Thermostat segment. To develop an ex ante impact estimate, the Evergreen team created a regression model that estimates 2019 ex post impacts as a function of temperature and time. Figure 23 shows the relationship between ex post impacts and temperature – though the data set is small, there is a positive trend. The regression model specification was shown in Section 1.5, and the results from the model are described in more detail below. Using the model, the Evergreen team predicts that the impact of a Two-Way Smart Thermostat DR event at peaking conditions (5:00 PM – 6:00 PM MDT when outdoor temperature is 100 degrees) is 1.36 kW per device (compared to an ex post impact of 1.02 kW).

Figure 23: Hourly Impacts against Outdoor Temperature (F)



The ex ante regression model was weighted by the number of curtailed devices in each event hour. Regression output is shown below. Temperature has a positive coefficient, indicating that higher temperatures produce higher impacts. Due to the small sample size, temperature is not considered a statistically significant predictor of the demand reduction. Unlike the other ex ante models, “hour” was not included as an explanatory variable in this model, as there simply are not enough data points to do so. When evaluating 2020 impacts, we will attempt to include the “hour” terms.

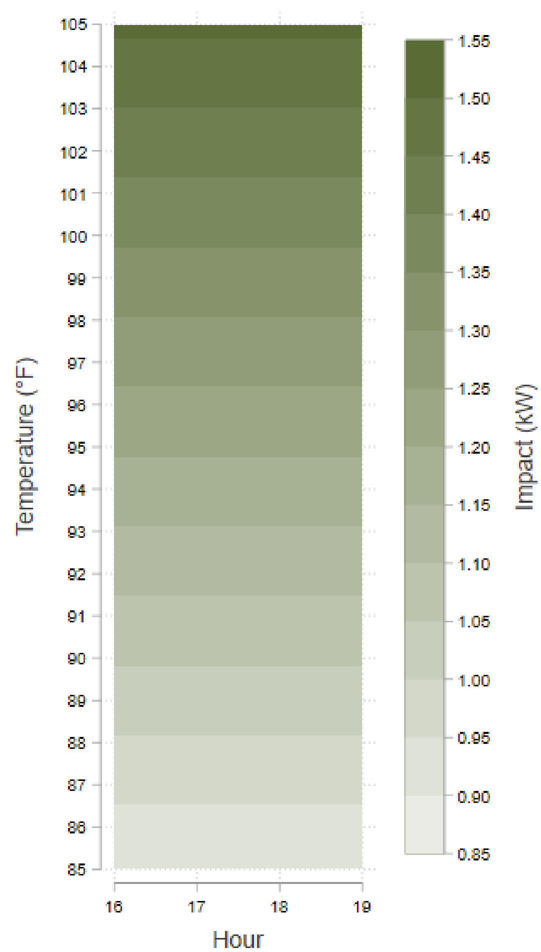
Table 24: Two-Way Smart Thermostat Ex Ante Regression Output

Term	Variable	Coefficient (b)	Standard Error	P-Value	95% CI
β	Temperature	0.03030	0.02021	0.165	(-0.015, 0.075)
α	Constant	-1.67125	1.85788	0.390	(-5.811, 2.468)

Using the regression coefficients shown in Table 24, the Evergreen team created a time-temperature matrix (TTM) that shows expected load reductions (per device) for different outdoor temperatures and at different times of the day. The TTM is shown in Table 25. These results should be interpreted with caution due to their small sample sizes.

Table 25: Two-Way Smart Thermostat Time-Temperature Matrix

Temp	Hour Ending MDT			
	16	17	18	19
105	1.51	1.51	1.51	1.51
104	1.48	1.48	1.48	1.48
103	1.45	1.45	1.45	1.45
102	1.42	1.42	1.42	1.42
101	1.39	1.39	1.39	1.39
100	1.36	1.36	1.36	1.36
99	1.33	1.33	1.33	1.33
98	1.30	1.30	1.30	1.30
97	1.27	1.27	1.27	1.27
96	1.24	1.24	1.24	1.24
95	1.21	1.21	1.21	1.21
94	1.18	1.18	1.18	1.18
93	1.15	1.15	1.15	1.15
92	1.12	1.12	1.12	1.12
91	1.09	1.09	1.09	1.09
90	1.06	1.06	1.06	1.06
89	1.03	1.03	1.03	1.03
88	0.99	0.99	0.99	0.99
87	0.96	0.96	0.96	0.96
86	0.93	0.93	0.93	0.93
85	0.90	0.90	0.90	0.90



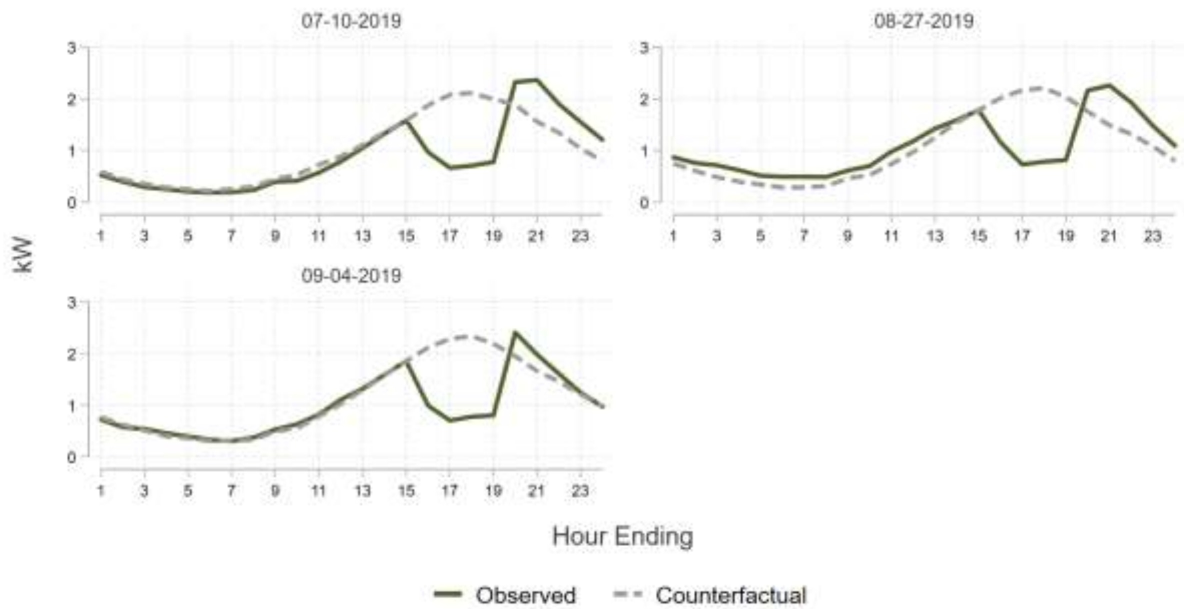
To get an idea of Two-Way Smart Thermostat resource capability on aggregate, the number of active facilities can be multiplied by the values shown in Table 25. As of the end of summer 2019, there were 384 active thermostats (not including the M&V group). Thus, the expected aggregate impact of an event hour ending at 6:00 PM (MDT) when the outdoor temperature is 100 degrees would be 0.5 MW. After applying the 85% online adjustment, this aggregate impact is 0.4 MW.

6 Recommendations

After our review of the 2019 Power Saver program, the Evergreen team offers the following recommendations:

- Ex post impacts provide a helpful look at program performance, but for planning purposes, a consistent, weather-normalized value should be used. The Evergreen team recommends that ex ante program impacts from 5:00 PM to 6:00 PM MDT at 100°F, de-rated for operability, be used for reporting, cost-effectiveness, and planning.
- The Itron contract definition of capacity performance is upwardly biased by capturing favorable noise along with the program impact. If there is a chance to review the terms, we recommend collapsing to the hourly mean rather than the maximum.
- For the Two-Way Smart Thermostat segment, there are significant differences in the load shapes for the M&V group and the curtailment group. We recommend using an alternative impact estimation method for this group. One option would be to use a method similar to what is used for the Small Commercial and Medium Commercial segments (high 3-of-5 baseline calculated in aggregate). Figure 24 shows counterfactual loads for this segment using this approach. The estimated baselines line up well with the actual loads during pre-event hours. Other possible options include using an alternating control group, use regression, or use a difference-in-differences approach.

Figure 24: Two-Way Smart Thermostat 3-of-5 Baseline Approach



Appendix D – Peak Saver Detailed Evaluation Methods and Findings

Public Service New Mexico (PNM) offers the Peak Saver program to non-residential customers with peak load contributions of at least 50 kW. The program compensates participants for reducing electric load upon dispatch during periods of high system load. Peak Saver was implemented by Enbala in 2019, who managed the enrollment, dispatch, and settlement with participating customers. During the summer 2019 demand response season, there were 92 participating facilities and three demand response events. These events are summarized in Table 26.

Table 26: 2019 Peak Saver Event Summary

Date	Weekday	Start Time (MDT)	End Time (MDT)	Daily High at KABQ (F)
07/10/2019	Wednesday	3:00 PM	7:00 PM	97
08/27/2019	Tuesday	3:00 PM	7:00 PM	94
09/04/2019	Wednesday	3:00 PM	7:00 PM	90

After the 2019 demand response (DR) season concluded, Enbala provided the Evergreen team with one-minute interval load data for each site in the Peak Saver population, as well as some workbooks with the performance metrics (10-minute capacity, average participant capacity, participant event capacity, and energy delivered) for each site/event combination. The interval data spanned a period from June 1 to September 30.

The one-minute interval load data also included a field with load impacts calculated using a customer baseline (CBL) method detailed in the contract between PNM and Enbala. A CBL is an estimate of what participant loads would have been absent the DR event dispatch. Load impacts are the difference between the CBL and the metered load during the event. The relevant CBLs were also in the one-minute load data.

With these data sources, the Evergreen team completed our verified savings analysis. The three key steps in the analysis were:

1) Reproduce the performance estimates calculated by Enbala using the contractually-agreed upon CBL method;
2) Assess the accuracy of the contract CBL method by examining its ability to predict loads on non-event weekdays; and
3) Modify the CBL methodology to reduce bias and calculate verified impacts for each event.

The findings from our analysis are described in subsequent sections.

1 Validation of Settlement Calculations

The settlement calculations called for a “high 3-of-5” baseline with an uncapped, asymmetric day-of adjustment. The high 3-of-5 days were determined as follows:

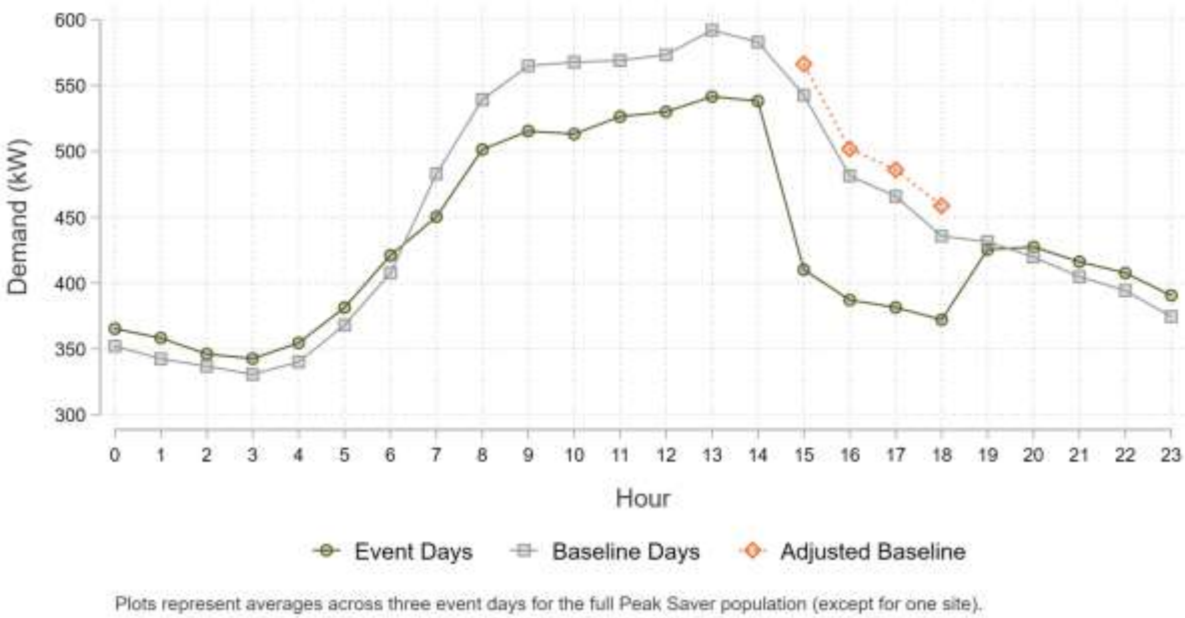
- Collapse the one-minute interval data to fifteen-minute interval data;
- Select the five non-holiday, non-event weekdays that immediately precede the event; and
- Out of those five days, pick the three days with the highest fifteen-minute interval read during the hours in which the event occurred.

Note that the three days with the highest fifteen-minute interval read were not necessarily the days with the highest average event-window load. In the case of a tie, the earlier of the two days was selected as a baseline day. This tie-breaking procedure was not laid out formally; rather, we discovered it when recreating Enbala’s calculations.

For one site, we discovered that baselines were calculated for just one of the three event days. Our team alerted PNM, who then alerted Enbala. Enbala promptly provided an updated data set for this site. Our team had no issues replicating baselines shown in the updated data. Other than this site, the Evergreen team encountered no issues in duplicating Enbala’s baselines.

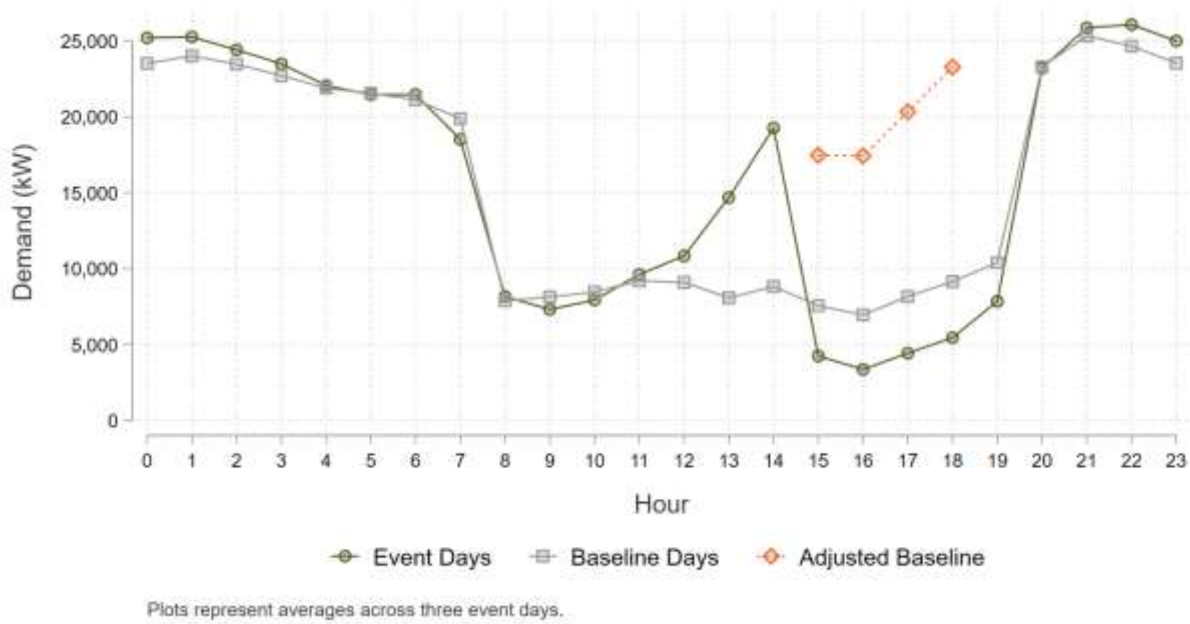
Figure 25 shows average hourly event day loads across the full population, average hourly loads on the baseline days that fed into the baseline, and also average hourly baselines. Note that the largest site (in terms of average kW) was not included in Figure 25, as its size and event day load profile muddied the overall trend. The gap between baseline day loads and event day loads for hours 8 through 14 is driven by a few larger sites.

Figure 25: Peak Saver Loads and Baselines



Averages for the site that was not included in Figure 25 are shown in Figure 26. This site pre-pumps before events (note the increase in event day loads for hours 13 and 14). This operational change makes sense with the DR event looming. However, the pre-pumping feeds into the baseline adjustment mechanism, which scales the baseline based on the ratio of pre-event event day load to pre-event baseline day load. The result is an upward bias (of more than 10 MW) in the adjusted baseline.

Figure 26: Event Day Loads and Baselines for Largest Site



After verifying that the baselines were calculated correctly, our team moved onto the performance metric calculations. The relevant performance metrics are:

- **10-Minute Participant Capacity Performance** – The difference between the CBL and the lowest actual electrical demand measured by a one-minute interval reading between eight and ten minutes after the start of an event.
- **Average Participant Capacity Performance** – The average difference between the CBL and the participant's actual electric demand beginning ten minutes after the initiation of the event.
- **Participant Event Capacity Performance** – Weighted average of 10-Minute Participant Capacity Performance (40% weight) and Average Participant Capacity Performance (60% weight).

During our review in a prior year, Enbala provided some additional insight into the zeroing out of negative performances:

> Regarding the performance calculations, if the (weighted) Event Capacity Performance is less than 0, it is capped to a minimum of 0 and the 10-minute and Average Performance numbers are also displayed as 0. However, if the Event Capacity Performance is greater than 0, the 10-minute and Average Performance numbers are left untouched and allowed to go negative.

Following this approach, our team was able to replicate all performance metrics for each site/event combination.

2 Assessment of CBL Accuracy

Developing an unbiased prediction of what load would have been absent a demand response event is essential to producing a defensible demand response impact estimate. This hypothetical non-event load is the customer baseline (CBL). If the CBL methodology tends to produce unbiased estimates of load (i.e., average error of zero), then demand response impact estimates will also be unbiased. If the CBL tends to overpredict or underpredict load, then demand response impacts will be overstated or understated.

This section details our review of the Enbala contract CBL methodology (described at the beginning of Validation of Settlement Calculations). Specifically, we assess the ability of the CBL methodology to predict load on non-event weekdays, and we explore the distribution of adjustment factors.

2.1 Placebo Event Analysis

Assessing the accuracy of a baseline on an event day is not possible because the counterfactual is unknown. In other words, we do not know what the demand would have been if the event was not called. However, on non-event weekdays there is no demand response, so using the same algorithm to generate a baseline should reasonably predict the metered load. For these days, the true value of demand response is 0 kW so if the baseline yields a non-zero impact estimate, it can be attributed to error. Individual errors are expected as the lookback window is not intended to be a perfect predictor of future load. That said, an unbiased baseline methodology should produce a distribution of errors which are centered around zero, on average.

The Evergreen team used this knowledge of the central tendency of the error to assess the accuracy of the settlement CBL. By creating a set of placebo event days composed of each non-event weekday, we investigated for systematic bias. Each placebo event was assumed to start at 3:00 PM and last for four hours – this mimics the 2019 DR events. Any negative impacts were not zeroed out. For each placebo event, the average CBL during the event window at each site was summed to find the aggregate CBL. The same process was used to find the aggregate metered load. Since no demand response occurred, the impact estimate (difference between CBL and metered load) should be zero and is thus labeled as error. Note that sites with solar power were removed from this analysis.[9] For sites with solar, the baseline adjustment mechanism used in the settlement CBL is truly an adjustment based on cloud coverage rather than an adjustment based on load. That's

[9] The Enbala team provided the evaluation team with a workbook that identifies sites as solar or non-solar.

problematic, of course, but it's a separate issue that we did not want to confound with the results of the exercise described in this section.

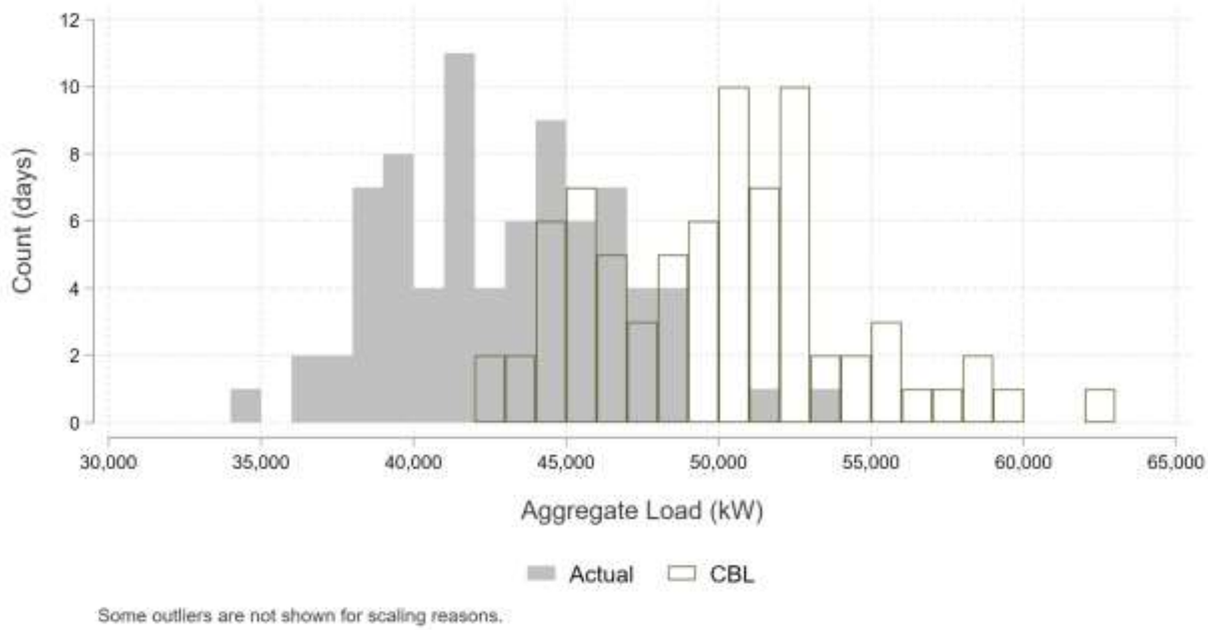
Results for the settlement baseline, aggregated by month, are shown in Table 27. On average, the baseline produced about 7.2 MW of upwards bias (meaning the baseline overstated load by 7.2 MW). The average percent bias across the 77 placebo events was 17.1 percent. Since actual DR reductions are not 100 percent of load, the bias in impact estimates for actual events is necessarily greater than 17.1 percent. Four sites account for 4.1 MW of the bias (approximately 57%).

Table 27: CBL Accuracy Assessment for Placebo Events

Month	Number of Placebo Events	Avg. Daily High Temp at KABQ	Avg. Aggregate Metered Load (kW)	Avg. Aggregate CBL (kW)	Avg. Error (kW)
June	16	88.7	40,423	46,951	6,528
July	21	92.3	41,379	47,859	6,480
August	21	92.2	44,930	52,853	7,923
September	19	84.7	44,965	52,554	7,589
Average	**---**	**89.7**	**43,034**	**50,191**	**7,157**

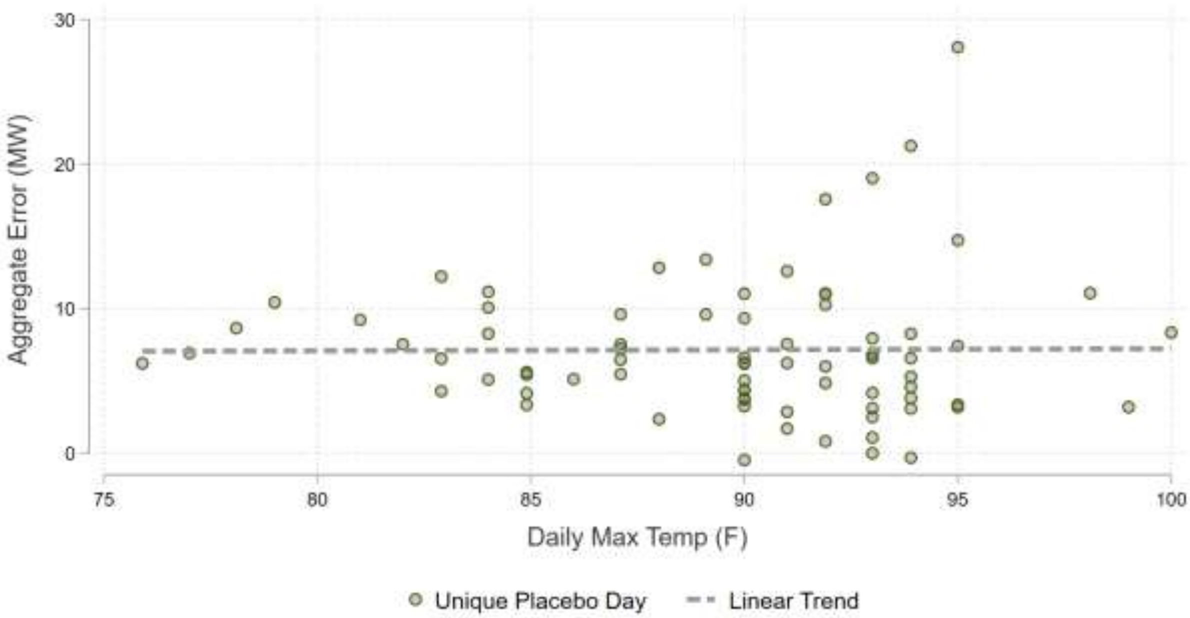
In Figure 27, a histogram compares actual aggregate load from the false event days (gray bars) to aggregate baselines (translucent bars). Ideally, the two distributions would be approximately identical. It is clear from the distribution that the CBL is upward biased.

Figure 27: Histogram of Placebo Event Days – Enbala Method



The placebo days summarized in Table 27 are not perfect representations of actual event days, which tend to be the hottest days of the summer. DR events are called because system operators expect higher than normal loads which will approach the constraints of the system. As a result, the performance of a baseline on hot days is much more important for assessing accuracy than its performance on a mild day. As shown in Figure 28, the performance of the baseline is not correlated with temperature. The average error on a placebo day with a maximum temperature of at least 95 degrees was nearly 10 MW.

Figure 28: Enbala Average Aggregate Baseline Error vs. Temperature



The Evergreen Team believes that the primary reason for such large errors in the settlement CBL is the asymmetric application of the weather-sensitive adjustment. The baseline can only be adjusted up, not down, which naturally biases the error upward. The unadjusted baseline actually produces less aggregate error than the adjusted baseline. While adjusting the baseline using event day loads has been shown to improve accuracy, the adjustment need to be bi-directional. In other demand response markets, including PJM and ISO New England, a symmetric adjustment is employed.

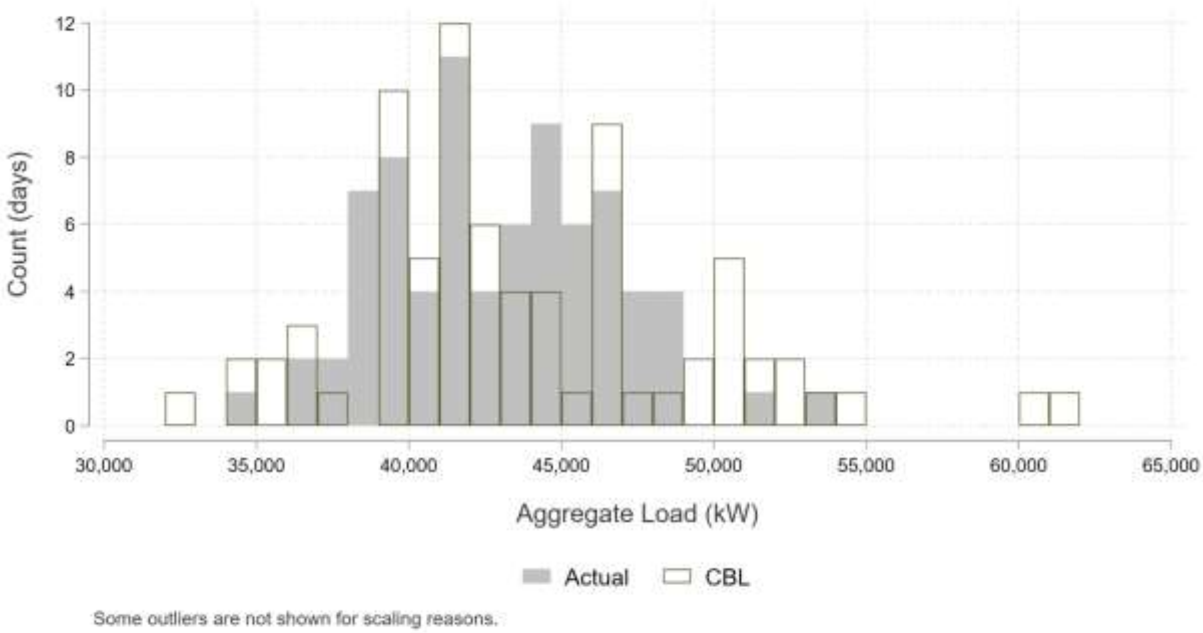
To illustrate the effect of a symmetric adjustment, we altered the CBL methodology to apply the adjustment in either direction depending on its value. Using this new adjusted baseline, we performed the same accuracy test described above. The results are displayed in Table 28. Average error for this method falls under 1 MW and is near 0 MW in two of the four months.

Table 28: Accuracy Assessment with Symmetric Adjustment

Month	Number of Placebo Events	Avg. Daily High Temp at KABQ	Avg. Aggregate Metered Load (kW)	Avg. Aggregate CBL (kW)	Avg. Error (kW)
June	16	88.7	40,423	40,649	226
July	21	92.3	41,379	41,581	220
August	21	92.2	44,930	46,769	1,838
September	19	84.7	44,965	45,540	575
Average	**---**	**89.7**	**43,034**	**43,779**	**745**

Figure 29 shows the histogram as Figure 27 but using the symmetric adjustment rather than the asymmetric adjustment. It is clear that the actual and counterfactual loads are better aligned in this case.

Figure 29: Histogram of Placebo Event Days – Symmetric Adjustment



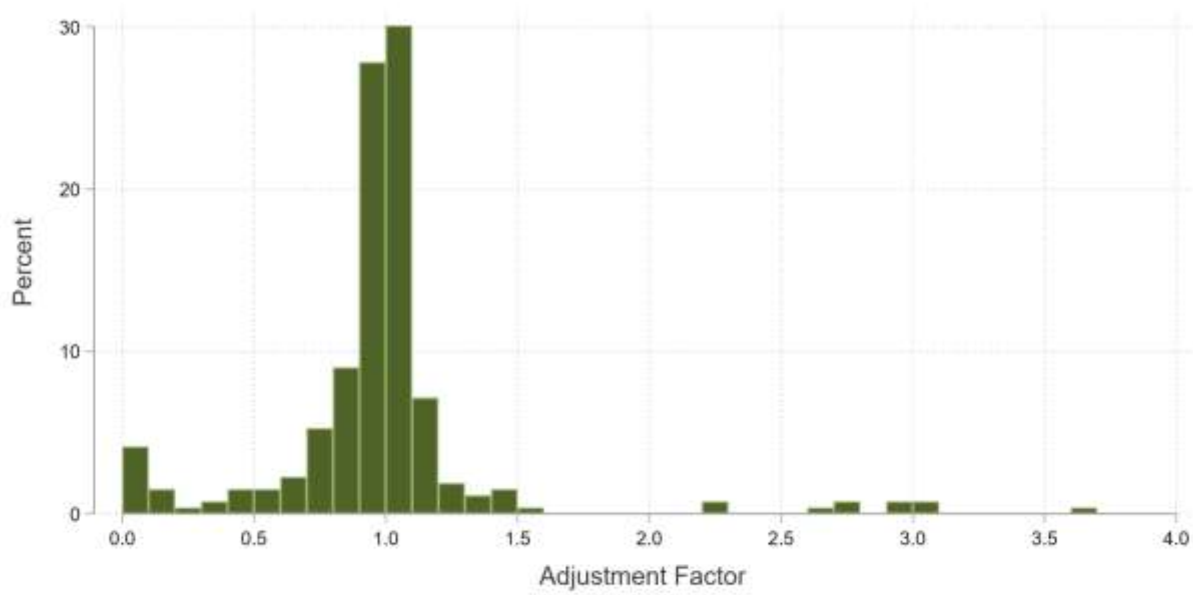
Using an asymmetric adjustment yielded an average error of 7.2 MW and an upwards bias of 17.1 percent. Using a symmetric adjustment yielded an average error of 0.7 MW and an upwards bias of 1.9 percent. While the baseline with a symmetric adjustment still overestimates on average, the distribution of errors falls on both sides of zero and the mean prediction is much closer to true load.

2.2 Adjustment Factors

As demonstrated above, the application of the adjustment factor plays a significant role in the accuracy of the CBL. Because the adjustment in the settlement CBL is applied as a multiplicative adjustment, even values that appear close to 1 (i.e., 1.1) can result in an adjustment of hundreds of kWs for a large customer. The average value of the symmetric adjustment factor across event days and sites was 1.08 and 70 percent of the adjustment factors were within 20 percent of 1 (between 0.83 and 1.2). The median factor, which is unaffected by extreme values, was 0.99.

Figure 30 shows the distribution of adjustment factors (except for one outlier). Recall that the adjustment factors are only applied if they increase the baseline in the contract CBL. In other words, any factor less than one is rounded up to one. In the majority of cases, the adjustments produced baseline values that were reasonable in the context of their distribution of load throughout the summer. Still, there were a handful of adjustment factors larger than two. Even for the most extreme cases of weather sensitivity, adjusting the baseline by a factor of two or more is dubious. Undoubtedly, leaving the asymmetric adjustment factor uncapped leads to an upwards bias in event day baselines, particularly when the adjustment is not symmetric. This again means impacts are, on average, being overstated using the Enbala baseline calculation. This can be addressed by subjecting the offset factor to a cap which prevents the adjustment factor from taking on extreme values.

Figure 30: Distribution of Adjustment Factors



One outlier (adjustment factor = 26.6) is not shown for scaling reasons.

The largest adjustment factor during the 2019 DR season was 26.6. The Evergreen team investigated load at this site to see if we could determine what happened. Figure 31 shows hourly demand for the lookback days and the event day in question. Average demand during the lookback days was about 14 kW and the maximum hourly demand was 53 kW. Right before the event, there was a large spike in demand. This spike, combined with the minimal load on the lookback days, resulted in a large adjustment factor. Though it cannot be seen in Figure 31, such a spike was not atypical at this site. Figure 32 shows a longer time series. The same spike occurs multiple times, typically around the same time of the day – late morning to early afternoon. That said, the spike is never fully coincident with the typical event window (3:00 PM – 7:00 PM), during which average demand is 65 kW. Perhaps the site did curtail load during the event on 7/10, but it seems fair to say that a baseline of 400+ kW is unreasonable for this site during the event window. Also worth noting: (1) adjustment factors for the other two event days were both less than one for this site, and (2) the DR commitment for this site is 500 kW.

Figure 31: Investigating a Large Adjustment Factor

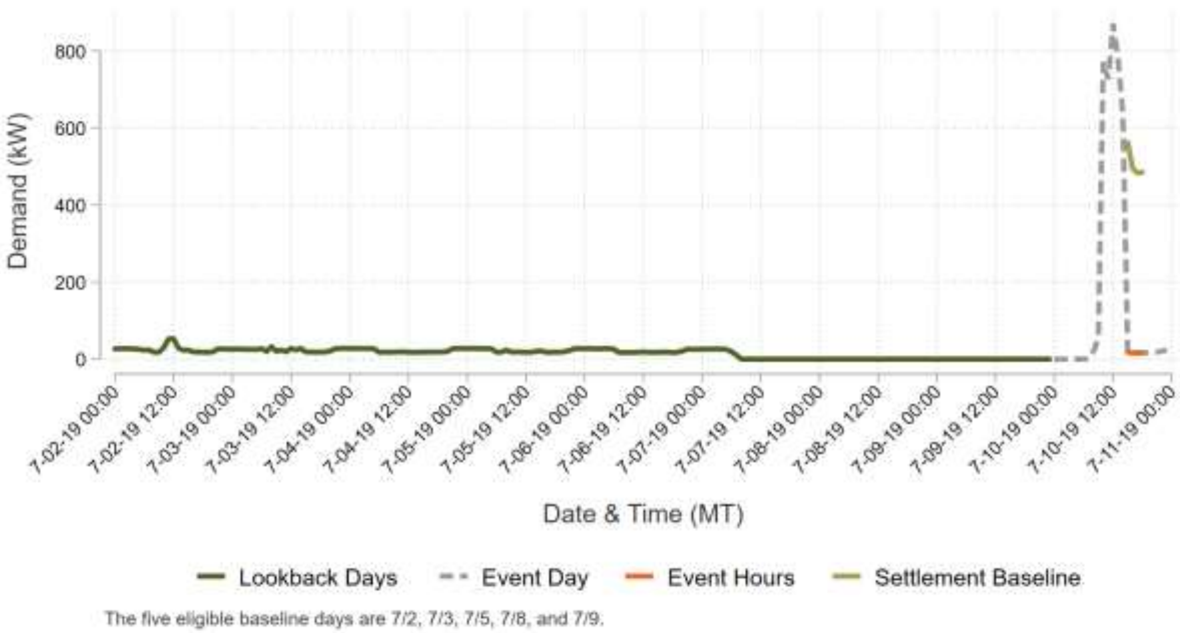
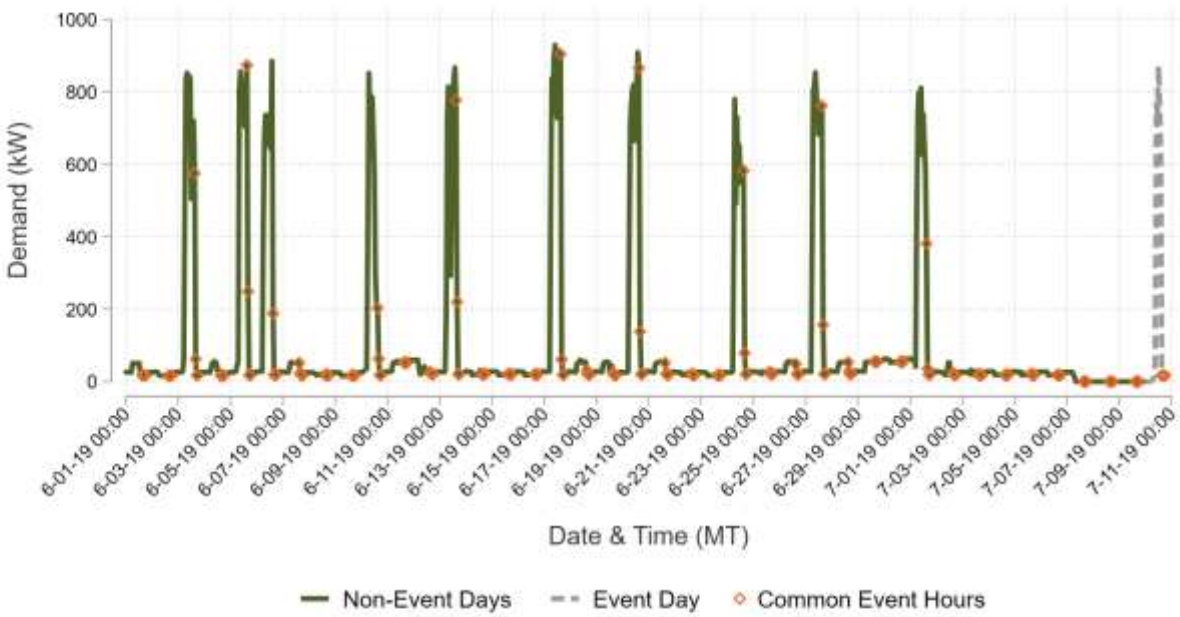


Figure 32: Investigating a Large Adjustment Factor – Longer Time Series



For sites with solar power, the adjustment factor is dependent on a cloud coverage effect that is not accounted for. If cloud cover begins mid-way through the adjustment window on the event day, net utility-supplied load for the hour will increase. If the lookback days were all sunny, then average load during the adjustment window on the lookback days will necessarily be lower than average load during the same window on the event day. This will result in a large adjustment ratio.

A similar effect may occur if sites engage in pre-cooling or pre-pumping in response to the pending demand response event. When this occurs, the adjustment factor will be artificially inflated (see Figure 26).

The adjustment factor is intended to correct for the differences in load between event and baseline days that result from the non-random selection of event-days. Event days are typically the hottest days of the summer and, as such, may be reasonably expected to have higher demand than baseline days. However, a weather adjustment need not be applied to sites which do not have weather sensitive load. It is our view that sites identified as weather sensitive are the only ones which should receive an adjustment to the baseline (excluding those with solar power and those who pre-pump in preparation for the demand response event).

3 Evaluated Impacts

3.1 Approach

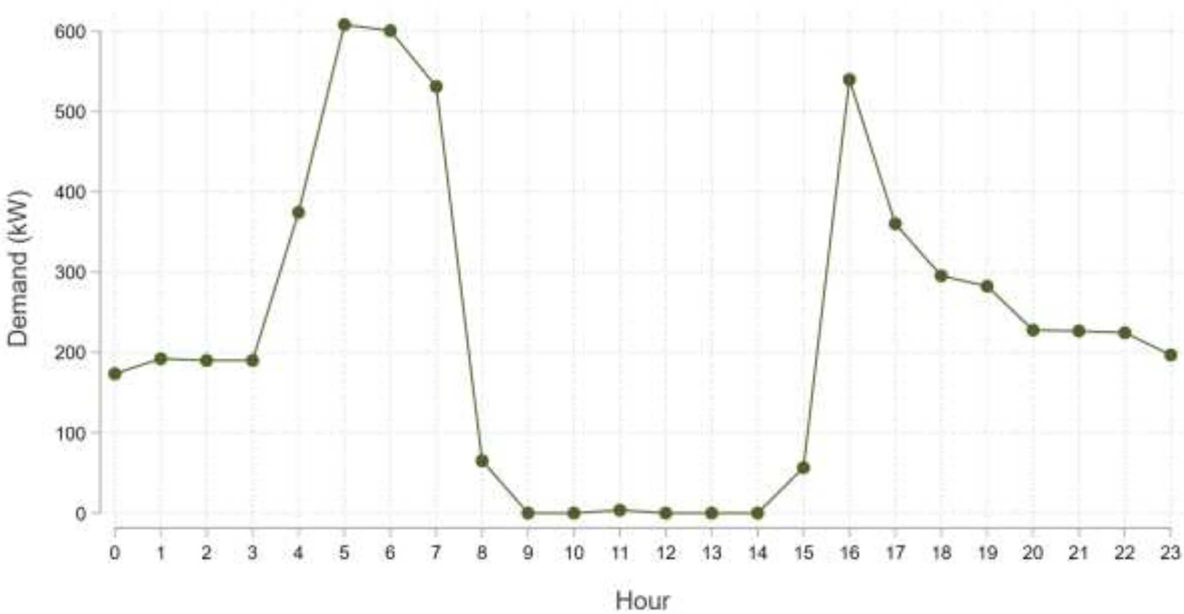
Based on our review of the CBL methodology used to generate Enbala's baselines and impact estimates, the Evergreen team calculated these values (and the performance metrics they feed into) using an adjusted CBL methodology:

- The adjustment factor is symmetric, meaning it can increase or decrease baselines, rather than only serving to increase baselines;
- The adjustment factor is capped at 20 percent rather than uncapped;
- The adjustment factor is only applied to sites that (1) have weather sensitive loads, (2) do not have solar power, and (3) do not pre-pump or pre-cool prior to demand response events;
- For sites that meet (1) and (2) above but not (3), an additive adjustment factor based on weather was applied rather than an adjustment factor based on pre-event load; and
- The 3-of-5 baseline days are selected based on average load during the event window rather than maximum 15-minute kW reading.

Regarding weather sensitive loads, the Evergreen team estimated weather sensitivity at each site by assessing the relationship between load and temperature during common event hours (2:00 PM – 7:00 PM, which includes the adjustment window) on non-event, non-holiday weekdays during the 2019 summer. Sites were considered to be weather sensitive if (1) the correlation between temperature and load was positive and (2) temperature was found to be a statistically significant predictor of load. In total, 52 of the 92 sites met these criteria.

Regarding solar power, the Enbala team provided the Evergreen team with a list of sites with known solar power. Our team also reviewed hourly load profiles for the full population of program participants. Sites that showed the distinct solar profile, as in Figure 33, were treated as solar sites even if they were not identified as such in the Enbala data. In total, 10 of 92 sites were considered sites with solar power.

Figure 33: Example of Solar Load Profile



Regarding pre-pumping or pre-cooling, our team reviewed hourly load profiles on event days and baseline days for the full population of program participants. Figure 26 illustrates this exercise. Sites with a notable incline in pre-event load, relative to load during the same hours on baseline days, were treated as pre-pumpers or pre-coolers. This is a reasonable action for a demand response participant. The issue is that it inflates the baseline adjustment, which is calculated based on pre-event load. In total, only three of 92 sites were considered pre-pumpers.

When these factors are considered in tandem, the load-based adjustment factor was applied to the baselines for 47 of the 92 sites. One other site received a weather-based adjustment. This is an additive adjustment is similar to the weather-based adjustment used by PJM. The adjustment is calculated as:

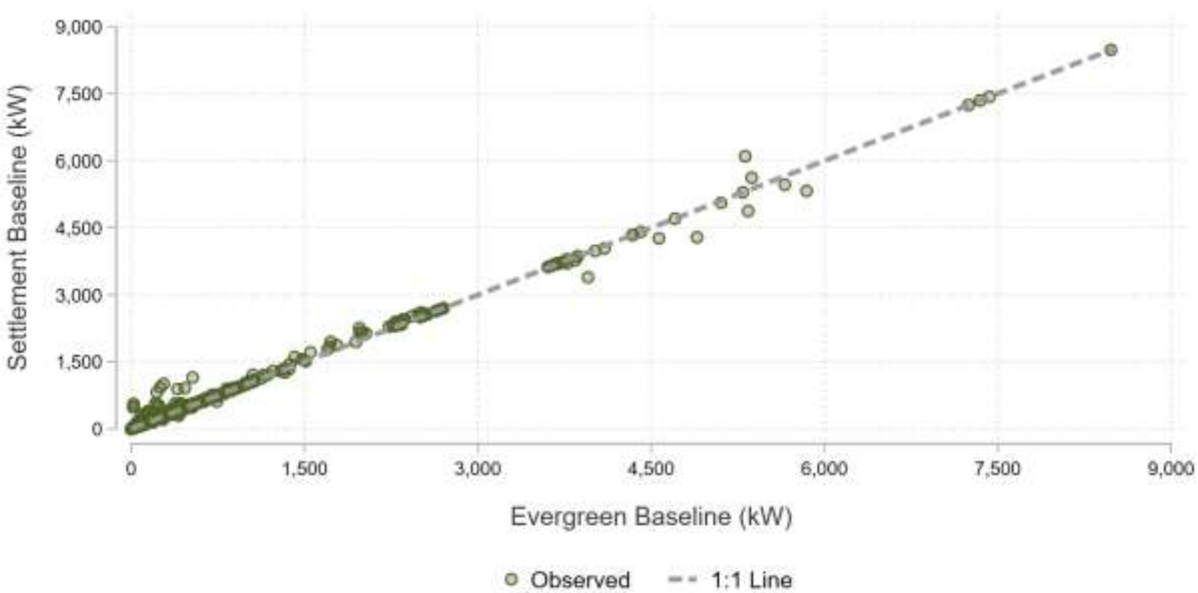
$$Adjustment = Slope * (\Delta_{Temp})$$

In the equation above, "Slope" is a value that quantifies the relationship between outdoor temperature and load for the facility (i.e., for each one unit increase in temperature, how much does load increase on average?). This value is determined via the regression modeling. The second component, Δ_{Temp}, represents the difference between the average outdoor temperature during the event and the average outdoor temperature during the event window on the baseline days.

3.2 CBL Comparison

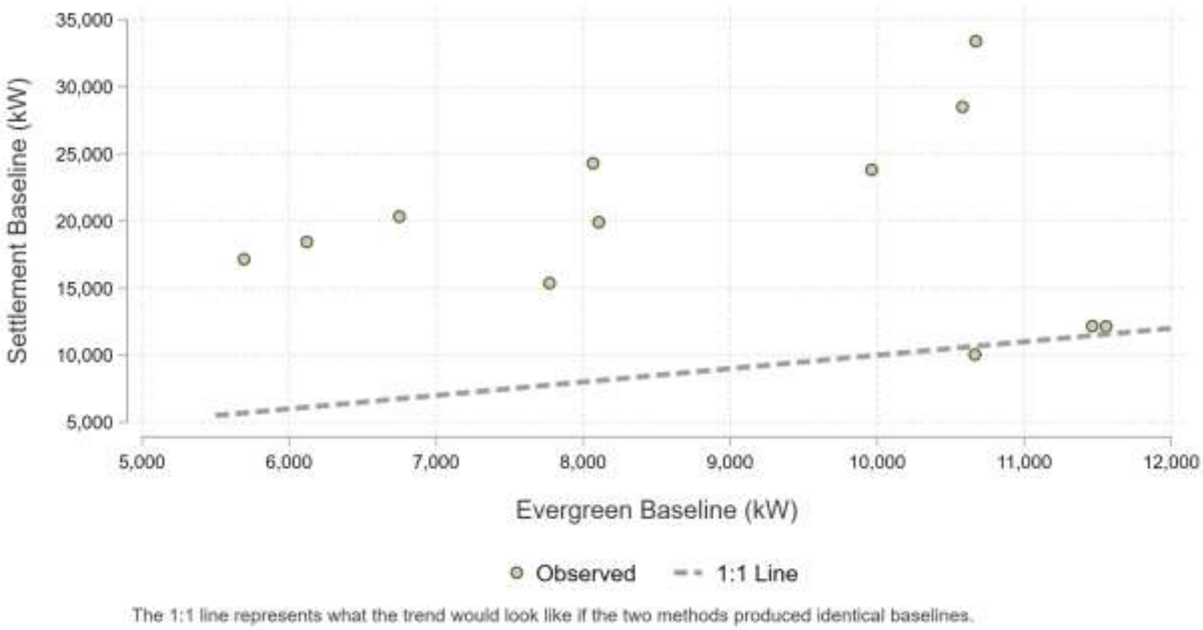
Because the Evergreen team calculated baselines in a manner that was similar to settlement baseline methodology, the baselines themselves were largely similar. This is illustrated in Figure 34, which compares the baselines our team calculated with the settlement baselines. One site, whose settlement baselines are inflated due to pre-pumping, is shown in a separate figure (Figure 35). In the latter figure, note the difference in the scale of the Y-axis and X-axis.

Figure 34: Baseline Comparison – All Sites but One



The 1:1 line represents what the trend would look like if the two methods produced identical baselines.

Figure 35: Baseline Comparison – Separate Site



By date, Table 29 and Table 30 show the average baseline under the settlement method and under the Evergreen method. Table 29 does not include the site that has been singled out elsewhere in this report, as this site distorts the overall averages. Table 29 supports the trend visualized in Figure 34 – the methods produce similar answers when one site is excluded, though the settlement method is naturally going to produce a larger baseline since it uses an asymmetric adjustment mechanism. For the site singled out in Table 30, the settlement baseline is nearly 11 MW higher than the Evergreen baseline, on average. As has been noted, this reflects pre-pumping at the facility that occurs in advance of the event. The adjustment mechanism used by the settlement baseline – a ratio of pre-event load on the event day to load during the same hour on the baseline days – captures this pre-pumping.

Table 29: Baseline Comparison – All Sites but One

Date	Aggregate Settlement Baseline (kW)	Aggregate Evergreen Baseline (kW)	Difference (kW)
07/10/2019	41,434	39,776	1,658
08/27/2019	46,744	44,485	2,259
09/04/2019	48,392	47,873	519
Average	**45,523**	**44,045**	**1,478**

Table 30: Baseline Comparison – Separate Site

Date	Settlement Baseline (kW)	Evergreen Baseline (kW)	Difference (kW)
07/10/2019	20,048	6,658	13,390
08/27/2019	26,401	9,830	16,751
09/04/2019	12,422	10,364	2,058
Average	**19,624**	**8,951**	**10,673**

3.3 Performance Metrics

After calculating adjusted baselines and adjusted impacts, the Evergreen team calculated participant performance metrics in a manner identical to the manner in which Enbala did so with one exception: we did not zero out negative performances as a rule. However, we did zero out negative performances in cases where the program implementer had documentation showing a site informed them that the site would not be participating in the upcoming event.

The results of the Evergreen team's 2019 Peak Saver Demand Response evaluation are shown in Table 31. For comparison, the savings produced by the program implementer are shown in Table 32. Our findings indicate the Peak Saver program is approximately a 17 MW capacity resource. On average, the verified capacity performance estimates using the Evergreen methodology are 56 percent of the values calculated by Enbala using the settlement CBL. The majority of the difference can be attributed to one site. Without that site, the verified capacity performance estimates using the Evergreen methodology are 85 percent of the settlement values. Section 2 described some of the other drivers leading to lower estimates for the Evergreen method.

Table 31: Evaluated Performance Summary by Event

Event Date	10-Minute Capacity Performance (kW)	Average Capacity Performance (kW)	Verified Capacity Performance (kW)	Energy Performance During Event Hours (kWh)
07/10/2019	14,416	12,953	13,539	51,180
08/27/2019	15,186	14,091	14,529	54,999
09/04/2019	22,777	18,981	20,499	74,570
Average	**17,460**	**15,342**	**16,189**	**60,250**

Table 32: Performance Summary – Program Implementer

Event Date	10-Minute Capacity Performance (kW)	Average Capacity Performance (kW)	Verified Capacity Performance (kW)	Energy Performance During Event Hours (kWh)
07/10/2019	29,864	28,257	28,900	113,057
08/27/2019	29,532	33,343	31,819	131,677
09/04/2019	31,862	21,334	25,545	85,139
Average	**30,419**	**27,645**	**28,754**	**109,958**

Table 33 presents daily energy savings. This is the aggregate difference between energy use on an event day and the baseline for all hours following the beginning of the event (including the event hours). Comparing the capacity performance, energy savings during the event, and the daily energy savings helps illustrate the extent to which event load was shifted to other hours. On average, aggregate energy used decreased by 63.8 MWh on event days. One would expect daily energy savings to be less than event energy savings due to snapback. This was not the case for Peak Saver in 2019, as several of the large customers saved energy in the post-event hours (i.e., their actual load was less than their baseline). The table also shows how these numbers change if the two largest sites (in terms of demand) are removed. These two sites swing the results by about 21 MWh for each event.

Table 33: Daily Energy Savings – Event Hours and Post-Event Hours

Event Date	Daily Energy Impact (kWh)	Daily Energy Impact (kWh) – Without Two Largest Sites
07/10/2019	59,793	37,781
08/27/2019	49,455	31,507
09/04/2019	82,014	58,189
Average	**63,754**	**42,493**

3.4 Nominations

The following sections detail comparisons the Evergreen team made between site-level DR kW commitments ("nominations"), average demand, and DR impacts. The latter section is a comparison between nominations and demand. As is often the case, this investigation spurred another: how do nominations compare with load on non-event days? Findings from this section are presented in 3.4.1. Throughout these two sections, note that results are presented at the participant level rather than the site level. That is, if one participant has three sites in the program, those three sites will be aggregated.

It is important to note that nominations will change throughout the summer for some participants. For most participants, this is not the case. The comparisons made in this section use nominations from August.

3.4.1 Comparing DR Nominations and Average Demand

In comparing DR nominations to load, our team only investigated common event hours (3:00 PM – 7:00 PM) on non-event, non-holiday weekdays. Additionally, any hours where the temperature was below 80 were removed. Under these conditions, we calculated average hourly demand for each participant, then compared these averages to the nominations. For the comparison, two metrics were calculated: raw differences and ratios. Raw differences are simply the difference between average demand and the nomination. Ratios were calculated as the nomination divided by average load (and multiplied by 100%).

Figure 36 shows the distribution of differences. A difference greater than zero implies average demand exceeds the nomination – this is what we would expect to see for all sites (though this may get muddied for sites with solar power). Indeed, most sites fall to the right of zero, but not all do. For one site in particular, the nomination was more than 3 MW above average demand. For this site, we will note that demand did sometimes exceed the nomination (in 28 of the 274 hours that fed into this investigation).

Figure 36: Comparing Nominations and Non-Event Demand

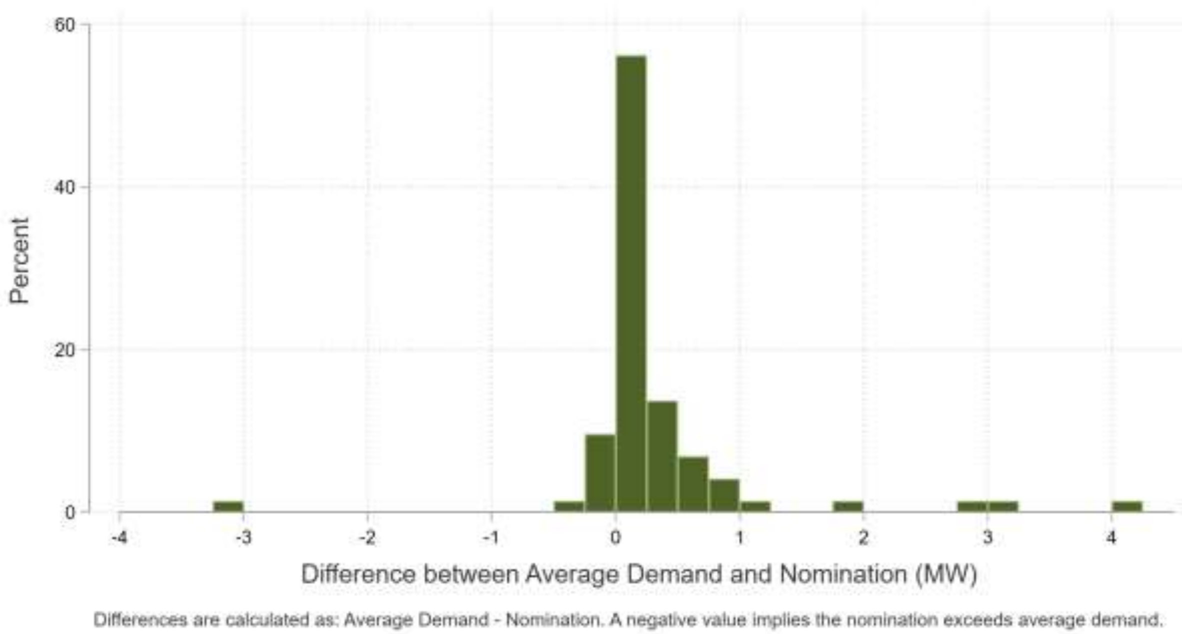
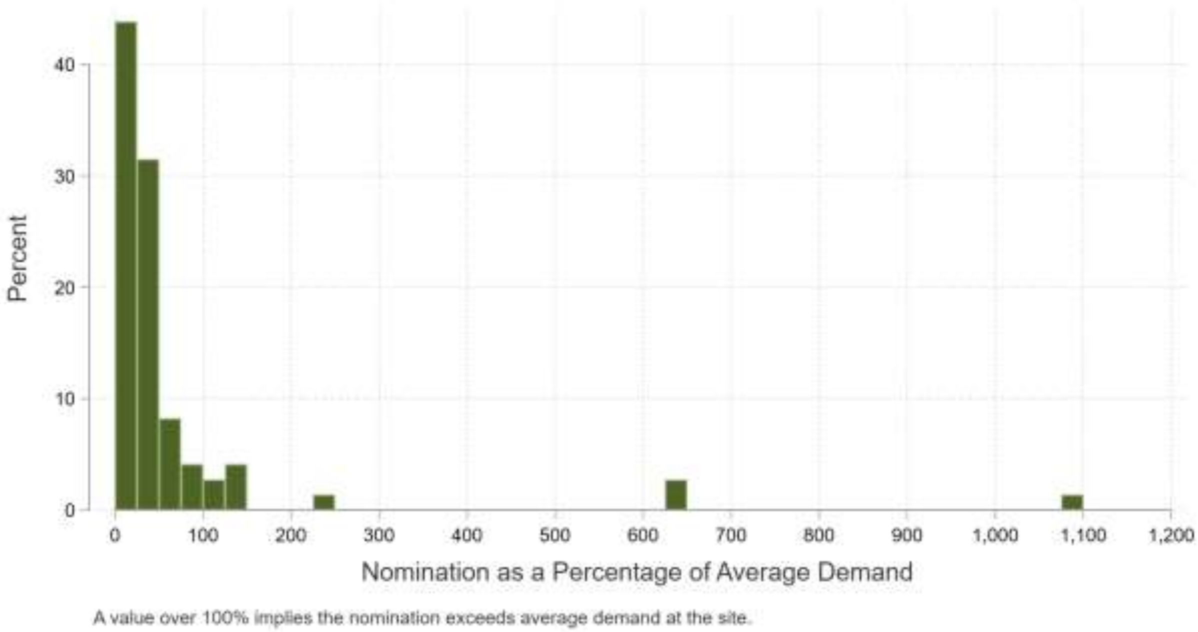


Figure 37 shows the distribution of ratios (ratio = nomination / average demand * 100%). A value greater than 100 percent implies the nomination exceeds average demand. For a handful of sites, the ratio was considerably greater than 100 percent. The outlier on the far right is a site that is known to have solar power. That said, even if the maximum demand value had been used rather than the average, the ratio for this site would still exceed 900 percent. Perhaps this site has a combination of solar and storage.

Figure 37: Nominations as a Percentage of Demand



For most participants, DR nominations make sense relative to their average hourly demand on non-event summer afternoons. For a handful of others, we would recommend reviewing the loads and nominations with Enbala (and possibly the customer).

3.4.2 Comparing DR Nominations and DR Performance

This section compares DR nominations with verified performance metrics (as calculated by the Evergreen team). The metric our team reviewed was the percent of the nomination achieved, calculated as follows:

$$Percent\ of\ Nomination\ Achieved = 100\% * \frac{Verified\ Reduction}{Nominated\ Reduction}$$

Figure 38 shows the distribution of these percentages. For each participant, unique percentages were calculated for each event. Instances where actual reductions do not exceed nominated reductions result in percentages that are less than 100 percent, and viceversa. The majority of the distribution falls below 100 percent, implying that most sites did not achieve their nominated load reduction on most event days.

Figure 38: Distribution of Percent Differences

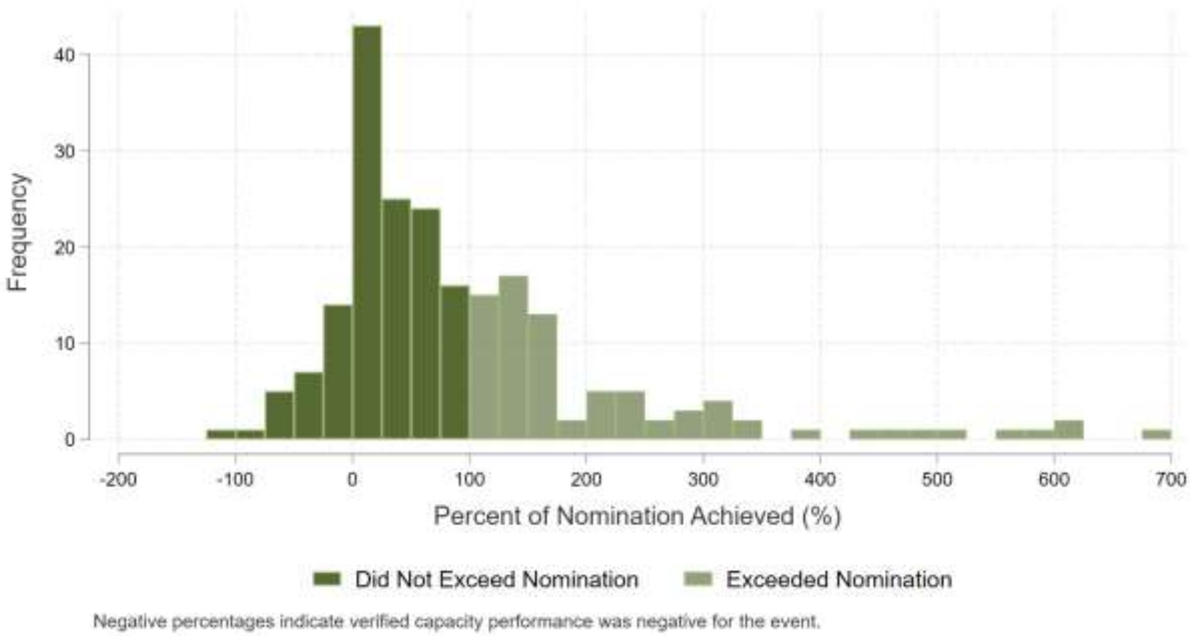


Table 34 groups participants based on how their verified reductions compared to their nominated reductions. Of the 73 participants, 27 exceeded their nomination on average.[10] Another 40 participants – accounting for 75 percent of the total nominations –did not exceed their nomination but did provide demand reductions. Figure 39 shows, on average, what percentage of their nomination each site achieved. The five participants with negative verified reductions are not included in the figure. None of these five sites have solar PV and three of them are schools.

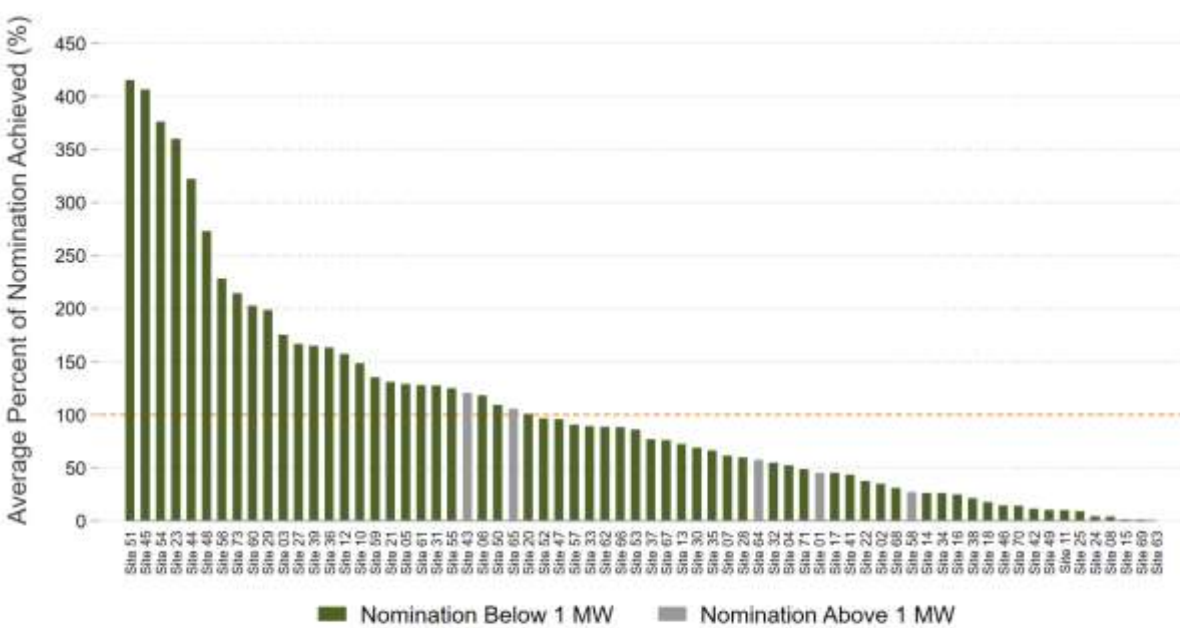
[10] Recall that sites are aggregated to the participant level. Some participants had multiple sites.

Table 34: Comparing Performance and Nominations

Result	Frequency	Aggregate Nomination (kW)[1]
Did Not Exceed Nomination	40	18,675
Exceeded Nomination	27	5,955
Negative Performance	5	370
Nomination of 0 kW	1	0
Total	**73**	**25,000**

[1] Nominations from August 2019 are shown.

Figure 39: Average Performance by Site



4 Recommendations

After our review of the 2019 Peak Saver program, the Evergreen team offers the following recommendations:

- Make the multiplicative adjustment symmetric rather than asymmetric. As per the assessment of CBL accuracy presented in Section 2.1, using an asymmetric adjustment results in an upwards bias in the baseline. Biasing the baseline inherently biases the performance metrics. The bias is greatly reduced when using a symmetric adjustment.

- Add a cap to the multiplicative adjustment factor. Otherwise, baselines are apt to approach unrealistic levels.
- Examine load data for solar patterns or pre-pumping/pre-cooling on event days. Pre-pumping/pre-cooling on event days is fine, but sites that do so should not receive the adjustment factor (or the adjustment factor should be based on weather rather than load). For sites with solar, consider using a smaller adjustment factor cap, using an additive adjustment, or removing the adjustment factor altogether.
- Compare DR nominations to the average demand on typical summer afternoons. If any nominations seem too high, update them. (We'll note that nominations for some sites do change throughout the summer.)
- PNM should also consider collecting all meter channels for sites with solar PV. This would allow the CBL to fully capture the load shape of sites that are net exporters during key times of day. It's possible that these sites reduced load and thus became larger exporters than they would have been on a non-event day, but the available data doesn't allow for a measurement. Also, an additive adjustment may work better than a multiplicative one for sites whose load can cross zero during the event period or adjustment window.
- Customer loads are volatile and baselines are not perfect. When metered load is higher than the baseline, performance estimates should be recorded as negative values and not zeroed out.

Appendix E – Commercial Comprehensive Desk Review Results Summary