**Responses to reviewers' comments:**

We would like to thank the reviewers and the editor for the time dedicated to reviewing this manuscript and for their constructive comments, which is useful for improving the quality of this manuscript. Based on the editors and reviewers’ comments, we have made changes to the content of the manuscript accordingly. Below, you will find our responses (in blue) to the reviewers’. Revised portions of the manuscript from each comment are also provided below (track change and clean version) where applicable.

***Reviewer #1:***

1. This research develops a useful package to extend the application of EnergyPlus into the R environment. This package would be helpful for parametric energy analysis in the field of building energy analysis since the R has a lot of packages on advanced statistical functions. The examples added in the appendix are also a good start point to be familiar with this new package. This is a well-structured paper to clearly present useful information on this new R package.  
   **Response**: Thank you for taking time to review our paper. We are grateful for your positive comments.

***Reviewer #2:***

1. The motivation for integrating data-driven analytics with BES is not well described in section 1. For instance, what are data-driven approaches and why/how they can be integrated with BES. A very concise summary will be necessary for the readers as general background.  
   **Response**: Thank you for the comments and suggestions. A summary of current research gaps has been added, including why data-driven analytics should be used when large parametric BES is involved. We have also merged the State-of-the-art section into the Introduction section for better readability and flow. (Pages 1 and 2 of the revised manuscript with and without changes tracked.)
2. An overall summary of state-of-art will be necessary to form up the research gaps.  
   **Response**: Thank you for the suggestion. An overall summary of the state-of-the-art has been included in the revised manuscript. (Pages 1 and 2 in the revised manuscript with and without changes tracked.)
3. Why choose R package particularly? What is the situation of other language. Does Python has similar package?  
   **Response**: Over the last decade, the R programming language has become a vital tool for implementing data analysis algorithms in many fields. It has extensive libraries of statistical modeling, data and database manipulation, wrangling, and visualization. New advanced statistical methods are usually first enabled through R libraries. This makes it an ideal choice for data-driven analytics. Another reason for choosing R is because of its good support for reproducible research, including literate programming, pipeline toolkits, project workflows, code/data formatting tools etc. Those facilities can be easily adapted to BES domain with the proposed R-based framework. Table 1 in the introduction gives a summary of the situation of other programming languages, including Ladybug & Honeybee and eppy packages (Python based), Modelkit (Ruby based), EpXL (Excel and VBA based), MLE+ (Matlab based). However, to the authors' knowledge, the proposed framework is the first one that focuses on seamless integration between BES and data-driven analytics, and BES reproducibility.
4. Can the developed framework be transferred into other language for larger application?  
   **Response**: Yes, it can. The developed framework is based on R which has a rich ecosystem of cutting-edge interface packages to communicate between open-source languages. Also, the proposed computational environment for BES is based on Docker, which is easy to be deployed for larger applications, including cloud computing.
5. Will the tool be possible for optimisation of building and energy systems? If so, how does it work?  
   **Response**: Thank you for the question. The proposed framework is leaned on the R language and its ecosystem. R offers rich facilities for solving various kinds of optimization problems, including optimization infrastructure packages, general-purpose solvers, mathematical programming solvers, and multi-objective optimization. A good example is the epluspar package mentioned in the Applications section, which integrates the proposed framework with multi-objective optimization solvers existing in R. With the flexible and extensible API, it is also possible to optimize building and energy systems if the problem can be abstracted and coded into R.

***Reviewer #3:***

1. 1. Abstract: the functionality of the R package should be provided.  
   **Response**: Thank you for the suggestion. The abstract has been revised to more clearly describe the functionalities of the eplusr package. (Page 1 of the revised manuscript with and without changes tracked.)
2. State of the art: it is better to merge the State-of-the-art section into the Introduction section to point out the knowledge gaps of the proposed study.  
   **Response**: Thank you for the suggestion. We have merged the State-of-the-art section into the Introduction section. We also added a dedicated subsection in the revised manuscript to describe the aim and objectives of this paper. (Page 5 in the revised manuscript with and without changes tracked).
3. Methodology: it is better to show the functionality of the R package rather than the detailed software structure and coding. The authors may also need to point out the novelty or academic contribution of each feature.  
   **Response**: Thank you for the suggestion. The main functionality of the proposed framework has been described by elaborating on the three main components and their respective features. Details on the novelty of each feature has also been included in the revised manuscript. (Page 5 of the revised manuscript with and without changes tracked).
4. Figures 8, 11, 16 are not easy to understand.  
   **Response**: Thanks for the comment. Figures 8, 11, and 16 were included to demonstrate the workflow of each application using the proposed framework. For improved clarity and to aid understanding, a short description has been included below each method/function.

(Figure 8 on page 14; Figure 11 on page 15; Figure 13 on page 17; and Figure 16 on page 19 of the revised manuscript without changes tracked.)

(Figures 8 on page 14, Figure 11 on page 16, Figure 13 on page 18, and Figure 16 on page 20 of the revised manuscript with changes tracked.)

1. Figure 9: this figure can be easily drawn based on the data in the eplus.html result file. The authors need to change the example to demonstrate the necessity of using the R package. For example, the users can use this R package to easily create some result figures, while it will require far longer without using the R package. Same comments for sections 4.2, 4.3 and 4.4.   
   **Response**: Thanks for the comments. We agree that Figure 9 may not be a good example for showcasing the necessity of using the R package. To better illustrate the capabilities of using the R package to streamline data exploration exploratory analysis using structured output that meets the Tidy data principle, the following changes were made to the revised manuscript: (1) Figure 9 was replaced with a stacked area plot of monthly breakdown of electricity consumption; and (2) Figure 10 was replaced with an energy signature diagram of outdoor temperature against electricity computation. Descriptions have also been added to showcase the advantages of structured tidy time-series output when performing data aggregation and visualization. We would also like to note that existing tools lack an easy way to extract data from the EnergyPlus output file (eplusout.html.), which is significantly simplified with the proposed R package. Additionally, although Figures 12 (Section 3.2), 14 (Section 3.3), and 18 (Section 3.4) are not complex visualization, the data extraction and preprocessing of large number of parametric simulation can be time-consuming and labor-intensive. With the eplusr package, such data extraction and preprocessing efforts becomes intuitive and effortless.

(Figure 9 on page 14 and Figure 10 on page 15 of the revised manuscript without changes tracked.)

(Figures 9 and 10 on page 15 of the revised manuscript with changes tracked.)

1. "Lines 59 - 124 in Listing 1", "line 42 in Listing 1", "only 15 lines of codes (line 60 - 82 in Listing 1)": It may not be necessary to show such detailed coding information. Fewer lines of code do not mean that they are easier to write.  
   **Response**: Thanks for the comments. We agree that fewer lines of code do not mean that they are easier to write. Nevertheless, one of the contributions of the proposed framework is that it brings reproducible research to the building energy simulation. Regardless, we agree that such code should not be included in the main manuscript, and thereby place it in the Appendix to avoid distractions to the readers.
2. The authors may need to show more about the "data-driven analytics" as indicated in the title.  
   **Response**: Thanks for the suggestion. To better illustrate “data-driven analytics”, energy signature analysis that is based on the time-series output of outdoor temperature and electricity computation has been included in the revised manuscript. We also further include the demonstration of extracting structured output using the proposed framework, thereby easing efforts to perform data-driven analytics.

(Page 14 of the revised manuscript without changes tracked.)

(Page 14 of in the revised manuscript with changes tracked.)

***Reviewer #4:***

1. This paper presents a framework for integrating building energy simulation and data-driven analytics (eplusr). In summary there is some good underlying work here, but the paper does not follow a typical peer reviewed paper format. It is presented as a final report. For example, there is no stand alone section to talk about the exact research methodology, results, and discussions; specially about figures/tables.  
   **Response**: Thank you for the comment and suggestion. The paper format has been changed in the revised manuscript to reflect a typical peer reviewed paper format. Specifically,
   1. The State-of-the-art section has been merged into the Introduction section. The research gaps have been move clearly discussed there.
   2. The novelties of the proposed framework have been added into the Methodology section.
   3. A Discussion section has been added to compare the proposed framework with existing solutions, including the advantages and current limitations.

(Pages 1-5 and 21 of the revised manuscript without changes tracked, and pages 1-5 and 22 of the revised manuscript with changes tracked.)

1. In my opinion, the article does not present any innovative contribution as it is mainly based on already existing research, and the significance of the results is quite limited. The results are not critically discussed in the context of the state of the art and uncertainties, strengths and weaknesses of the work are not discussed.  
   **Response**: Thank you for the comment. The main contribution of this work lies in the software development of a framework that enables seamless integration between BES and data-driven analytics. The proposed framework also encourages reproducible BES research. To the authors’ knowledge, this is the first research that bridges this gap. To bring clarity to the innovative contribution of this work, a comparison of the proposed framework to existing solution including discussion of its strengths and weaknesses has been included in the revised manuscript.

(Page 21 of the revised manuscript without changes tracked)

(Page 22 in the revised manuscript with changes tracked.)

1. Quality assurance on EnergyPlus simulations needs to be addressed carefully in this paper. Working with simulations facilitates mistakes, which are seldom found without equally thorough quality assurance procedures. Please describe which sort of tests and measures were adopted to assure the quality of the presented results.  
   **Response**: Thank you for the comment. However, although important, quality assurance of the simulation models does not fall within the scope of the current paper since the emphasis is on the integration between the energy simulation engine and data driven analytics. However, we would like bring to attention that the eplusr package contains the following that would help with the quality assurance of the simulation model. (1) a rule-based model data validator to avoid any possible input errors before modifying EnergyPlus models. It includes 13 different checks and is triggered whenever any methods are called to modify an IDF. This is also true for the parametric model generation. Issues with detailed reasons and possible solutions are directly reported to the users; and (2) an EnergyPlus simulation error (ERR) file parser to show simulation debug information and summarize error messages. The simulation manager has an *errors* method to collect error messages from single simulation and parametric simulations. Also, the Tidy data interface will check if the simulation completes successfully and will issue a warning message otherwise to remind users that the collected data is not reliable.

Nonetheless, we agree that quality assurance of the R package that has been developed is an important aspect that should be included. Therefore, a Quality assurance and quality control subsection has been added to the Methodology section in the revised manuscript to further discuss those efforts which includes the Test-Driven Development process that the proposed R package follows. Currently, there are more than 3600 unit tests which covers around 90% of the total codebase. The released version of eplusr is distributed via CRAN which runs all the tests automatically on Windows, macOS, Linux and Solaris at each eplusr submission and also each new release of the R language itself. The development version is held in a GitHub public repository with Continuous Integration (CI) using GitHub Actions which runs all the tests whenever any code changes occur. (See Page 12 of the revised manuscript without changes tracked and pages 12-13 of the revised manuscript with changes tracked.).

1. The topic is interesting and worthy of investigation, but it is not clear how this study contributes to advance the state of art about methodology, data processing, and reliability of results.  
   **Response**: Thank you for the comment. As mentioned in our response to comment 2 above, to the best of the authors’ knowledge, this is the first research focusing on providing seamless integration between BES and data-driven analytics, and solutions for reproducible BES research. Further discussion on the novelty and advantages has been added in the discussion section of the revised manuscript (Page 21 of the revised manuscript without changes tracked and pages 21-22 of the revised manuscript with changes tracked.). The Applications section, together with the code in the Appendix section further demonstrates the benefits of the tidy-formatted BES results which can be easily fed to various data-centric analytics using existing tools in R. The efforts of quality assurance and result reliability of the proposed framework has been added to the Methodology section of the revised manuscript (page 12 of the revised manuscript without changes tracked and pages 12-13 of the revised manuscript with changes tracked.).