

Response to the 'Reviewers' Comments

Tracking #: 776-1756

"A benchmark dataset for the retail multiskilled personnel planning under uncertain demand"

First of all, we would like to thank you for allowing us to revise and resubmit our manuscript. We sincerely appreciate the time and efforts you have devoted to providing us with your valuable comments, which greatly helped us improve our paper. We did our best to answer all comments clearly and explicitly. For clarity, our answers in this response letter are highlighted in blue.

Response to Reviewer #1

We appreciate your recognition of the value of our presentation of real data and that you found the dataset promising. Furthermore, we found your suggestions very helpful and have addressed your comments in the revised version of the paper, as detailed below.

Given that comments C1 and C2 are closely related, we have decided to provide a single response addressing the concerns expressed by you in the two comments.

C1: The authors present a multiskilled personal planning problem dataset from a home department company in Chile. Although the dataset might help to assess further planning algorithms and approaches, the aforementioned literature is not discussed appropriately and is mostly just enumerated (see first and second paragraph). (...) The authors need to elaborate more on certain aspects and the manuscript should therefore be extended (if the general length limit is already reached, I urge the editor to allow for an exception).

C2: Also previous work from the authors, where this dataset was already published, is not mentioned: <https://doi.org/10.1016/j.dib.2020.106066>. Therefore, the authors should mention the difference to their previous article as the dataset seems to be the same and is already available there.

Ans to C1 and C2: We appreciate your guidance. In response to your suggestion, we have extensively revised Section 1 (*Introduction*), and added a new section entitled "*Background on PSPs involving multiskilled staff in a retail setting*," now Section 2 of the manuscript. Both sections complement each other to provide a comprehensive and meaningful discussion of the existing literature on personnel scheduling problems involving multiskilled employees in a retail environment. Consequently, we have gone beyond a mere listing of articles and have been able to identify a significant gap in the literature: *The need for access to datasets to solve Multiskilled Personnel Assignment Problems (MPAPs) in the context of uncertain demand in a retail setting*. Note that we elaborate on the contributions of our data article at the end of Section 1.

In addition, towards the end of Section 2, we discuss existing publicly available datasets. We explicitly outline how the datasets presented in this data article differ from those already accessible in repositories or another published data article (<https://doi.org/10.1016/j.dib.2020.106066>). In particular, we highlight how our datasets contribute to addressing MPAPs under stochastic demand considerations in retail settings. This also allows both academics and practitioners to find robust solutions for similar or identical MPAPs, by performing a benchmark analysis of different approaches to optimization under uncertainty using the datasets provided in our data article.

We hope that these revisions meet your expectations, and we are grateful for the opportunity to improve the quality of our manuscript.

Response to Reviewer #2

We appreciate that you found our datasets, based on real-world data, useful for solving challenging personnel scheduling problems. Your comments and suggestions were very helpful in our attempt to improve the paper, and we have addressed your concerns, as detailed below.

C1: The authors present two datasets for a personnel scheduling problem that involves multiple skills and training. The first dataset comes directly from a software company and presents real-world data. The second dataset is generated by the authors themselves. Having this dataset available may be helpful for other researchers working on similar problems. However, it would have been nice if additional information would have been present in the dataset, eg about employee availability (which is also stochastic) or learning rates.

Ans: We are grateful for your feedback. In response to your suggestion, we recognize the potential value of additional data, such as unscheduled personnel absenteeism and learning and forgetting rates, to enrich the versatility and utility of our data article. Thus, in the new version of the manuscript, we have included real data derived from the experience of Chilean retailers, specifically focusing on unscheduled personnel absences and the learning/forgetting phenomena. These data were previously used in two of our published research articles, Mac-Vicar et al. [24] and Henao et al. [3]. The discussion and justification of these new data are presented in the added paragraphs after Table 2 in Subsection 3.1.

C2: Moreover, the current presentation (specifically the text) is very poor in terms of grammar, which makes it sometimes hard to understand the paper. (...) Mistakes in the text that should be corrected: (...) I stopped correct grammar mistakes here, but there are still many left. I strongly suggest the authors to make use of a proofreading service to further correct the text!

Ans: We would like to thank the reviewer for taking the time to conduct a thorough linguistic and grammatical review of our manuscript. To address your recommendation, we have taken two actions.

First, we made an effort to review all the manuscript looking for sentences and paragraphs where we could be clearer. Consequently, we have rewritten a lot of text in the manuscript, to ensure a comprehensive language check. Second, due to budget limitations, we were unable to hire a proofreading service. However, following the Editor's recommendation, we used a tool such as *DeepL* to improve the grammar and overall flow of the text.

As a result of these actions, we believe that the paper has benefited from the reviewer's recommendation; thus, the new version of the manuscript can now be followed more smoothly.

Response to Reviewer #3

Thank you for your positive feedback and for recommending the acceptance of the article. We sincerely appreciate your recognition of the valuable contribution of our data article, as well as the numerous positive aspects you highlighted. Furthermore, your feedback has been instrumental in enhancing various aspects of the manuscript, as detailed below.

C1: Clarity in Methodology Description: The paper could benefit from a more detailed and explicit explanation of the methodology used for Monte Carlo simulation to generate the simulated dataset. Providing step-by-step details would enhance the transparency of the simulation process for readers.

Ans: We appreciate your insights, and in response to your recommendation, we have taken two specific actions to improve the clarity of the explanation of how our Monte Carlo Simulation (MCS) method generates the simulated datasets. First, we have conducted a thorough review of the entire manuscript, with a particular focus on improving the explanation of Subsections 3.2 and 4.2. Second, in addition to the detailed explanation provided in Subsection 4.2 regarding the set of Excel formulas used in the Excel worksheets to implement the MCS method, we have added a paragraph at the end of that subsection. This new text emphasizes that readers can easily cross-reference and verify the step-by-step details of our MCS method implementation by examining the Excel worksheets and the set of Excel formulas used. All these details are readily available to the reader in the Excel workbook.

C2: Data Validation and Quality Assurance: It would be beneficial to include information on data validation and quality assurance processes applied to both the real and simulated datasets. Ensuring the accuracy and reliability of the data is crucial for the credibility and usability of the database.

Ans: Thank you for your comment. To address your concern, we made improvements to the manuscript.

First, as explained in the manuscript, our real dataset comes directly from the specialized software of the company SHIFT SpA. A company dedicated to the optimization of shift schedules for thousands of employees across Latin America. This software is used to estimate real data based on the experience and historical data of Chilean retailers. As a result, the data provided by SHIFT SpA are already being used by store managers for decision-making in real-world environments, thereby validating and exemplifying the quality of our real data. We have added a brief text highlighting this positive aspect in the first paragraph of Subsection 4.1 in the revised manuscript.

Second, we offer an Excel workbook that allows readers and users to validate the accurate generation of these simulated datasets. Moreover, our revised manuscript now included a detailed and clearer explanation of the MCS method used to generate the simulated datasets.

Finally, in the revised manuscript, we have explicitly stated that the real dataset and the simulated datasets have already been used in various previous research articles for the experimentation and validation stages. These datasets, although generated earlier, were not published before and were used in our previously published research articles, including Henao et al. [2], Henao et al. [21], and Henao et al. [22]. This demonstrates the applicability and validity of the datasets presented in this data article, as they have been used in articles published in high-impact journals.

C4: Comparison with Existing Databases: While the paper mentions the utility of the dataset for benchmarking, a more thorough discussion comparing this database with existing datasets in the field would strengthen the paper. Highlighting the unique features and advantages of this dataset in comparison to others would enhance its appeal to researchers.

Ans: Your comment was aligned with another reviewer's, so we have followed the recommendations of both. In the revised manuscript, we have extensively revised Section 1 (*Introduction*), and added a new section entitled "*Background on PSPs involving multiskilled staff in a retail setting,*" now Section 2 of the manuscript. Both sections complement each other to provide a comprehensive and meaningful discussion of the existing literature on personnel scheduling problems involving multiskilled employees in a retail environment. Consequently, we have gone beyond a mere listing of articles and have been able to identify a significant gap in the literature: *The need for access to datasets to solve Multiskilled Personnel Assignment Problems (MPAPs) in the context of uncertain demand in a retail setting.* Note that we elaborate on the contributions of our data article at the end of Section 1.

In addition, towards the end of Section 2, we discuss existing publicly available datasets. We explicitly outline how the datasets presented in this data article differ from those already accessible in repositories or another published data article. In particular, we highlight how our datasets contribute to addressing MPAPs under stochastic demand considerations in retail settings. This also allows both academics and practitioners to find robust solutions for similar or identical MPAPs, by performing a benchmark analysis of different approaches to optimization under uncertainty using the datasets provided in our data article.

C5: Illustrative Examples: Including illustrative examples or case studies demonstrating the application of the dataset in solving multiskilled personnel assignment problems under uncertain demand would provide practical insights for readers. This could help bridge the gap between the theoretical presentation and real-world application.

Ans: Thank you for your comment, which we believe is closely related to your comments C2 and C4. As already elaborated and discussed in the responses to both comments, we highlight the unique features and advantages of our datasets in the revised version of the manuscript. In addition, we explicitly state that our datasets are highly useful and versatile for addressing MPAPs in the context of uncertain demand in a retail setting. Their applicability is demonstrated by their use in articles published by Henao et al. [2], Henao et al. [21], and Henao et al. [22]. These articles present applied case studies using these datasets and provide managerial insights for retail store managers.

As comments C3 and C6 share a close connection, we have opted to offer a unified response that encompasses the concerns you raised in both comments.

C3: *Discussion of Limitations:* The paper lacks a discussion of potential limitations associated with the dataset. Acknowledging and addressing any constraints or shortcomings in the data, such as assumptions made during simulation or constraints in the real-world data, would provide a more balanced perspective for users.

C6: *Future Directions and Use Cases:* Including a section on potential future directions for research and practical applications based on the dataset would inspire further exploration in the field. This could involve suggesting specific research questions or highlighting industries beyond retail where the dataset could be applied.

Ans C3 and C6: We agree with your valuable comments regarding the need to address the potential limitations associated with our datasets and to outline future research directions. We believe that limitations and future research are related topics because it is necessary to explain the limitations of a study before identifying and providing future research directions. In response to this, in the new Section 6, "*Limitations and future research*," we have linked the potential limitations of our datasets to possible areas for future investigation.

In addition, although not requested by any reviewers, we have decided to summarize the main conclusions of our article in a new Section 5, "*Conclusions*." We believe that reading Sections 5 and 6 will provide users with a clearer overview of the article's contributions and how the potential limitations of the datasets can guide and inspire future research efforts.