Data Article

Dataset for solving a hybrid flexibility strategy on personnel scheduling problem in the retail industry

Andrés Felipe Porto\textsuperscript{a,b}, César Augusto Henao\textsuperscript{b,*}, Héctor López-Ospina\textsuperscript{b}, Esneyder Rafael González\textsuperscript{b}, Virginia I. González\textsuperscript{b}

\textsuperscript{a} Department of Industrial Engineering, Corporación Universitaria Americana, Barranquilla, Colombia
\textsuperscript{b} Universidad del Norte, Barranquilla, Colombia

\textbf{A R T I C L E  I N F O}

Article history:
Received 1 July 2020
Accepted 17 July 2020
Available online 24 July 2020

Keywords:
Personnel scheduling
Workforce flexibility
Staffing
Tour scheduling
Multiskilling
Flexible contracts
Retail services

\textbf{A B S T R A C T}

This data article describes datasets from a home improvement retail store located in Santiago, Chile. The datasets have been developed to simultaneously solve a staffing and tour scheduling problem that incorporates flexible contracts and multiskilled staff. This Data in Brief article is related to the published article “Hybrid flexibility strategy on personnel scheduling: Retail case study” \cite{1}. The datasets contain real, processed, and simulated data. Regarding the real and processed datasets, they are presented for three different store sizes (4, 5 or 6 departments). Real datasets include information about the employment-contract characteristics, cost parameters, and a forecast of the number of employees required in each department for each day of the week and each time period into which the operating day is divided. As regards the data processed for the case study, they include the set of skill sets considering that the employees can be trained in a maximum of two store departments. Regarding the simulated datasets, they include information about the random parameter of staff demand in each store department. The simulated data are presented in 90 text files classified by: (i) Store size (4, 5 or 6 departments). (ii) Coefficient

DOI of original article: 10.1016/j.cie.2019.04.049
* Corresponding author.
E-mail address: cahenao@uninorte.edu.co (C.A. Henao).
of variation (10, 20, 30%). (iii) Instance identification number (10 instances per scenario that resulted from combining the store sizes and coefficients of variation). Researchers can use the datasets for benchmarking the performance of different approaches with the one presented by Porto et al. [1], and in consequence, they can find solutions to the same (or similar) type of personnel scheduling problem. The dataset includes an Excel workbook that can be used to randomly generate staff demand instances according to a chosen coefficient of variation.

© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license. (http://creativecommons.org/licenses/by/4.0/)

### Specifications table

| Subject | Management Science and Operations Research |
| --- | --- |
| Specific subject area | Personnel scheduling |
| Type of data | Tables, Figures, Text files, Excel workbook |
| How data were acquired | Real and processed data from a retail store and simulated data generated using a Monte Carlo simulation in an Excel workbook |
| Data format | Mixed (raw and analyzed) |
| Parameters for data collection | Real and processed datasets are presented for three different store sizes. Real data including the employment-contract characteristics, cost parameters, and a forecast of the number of employees required in each department by weekday and time period of the day. Processed data including the set of skill sets. Simulated data containing randomly generated staff demand in each store department |
| Description of data collection | Real and processed data were collected from a home improvement retail store. Simulated data were generated using the Excel formulas of inverse probability distribution and random values. The outputs are presented in 90 text files classified by: (i) Store size (4, 5 or 6 departments), (ii) Coefficient of variation (10, 20, 30%), (iii) Instance identification number (10 per scenario) |
| Data source location | Retail store in Santiago, Chile |
| Data accessibility | Data can be downloaded from: Mendeley repository name: Dataset for solving a hybrid flexibility strategy on personnel scheduling problem in the retail industry Data identification number: doi:10.17632/rzf6whbm6g6.1 Direct URL to data: https://data.mendeley.com/datasets/rzf6whbm6g6/draft?a=c2d30c00-5d6a-40bc-8bc5-04bcfc8329d8 |
| Related research article | Porto, A. F., Henao, C. A., López-Ospina, H., & González, E. R. (2019). Hybrid flexibility strategy on personnel scheduling: Retail case study. Computers & Industrial Engineering, 133, 220–230. |

### Value of the data

- The datasets contain real, processed, and simulated information that can be used to simultaneously solve a staffing and tour scheduling problem that incorporates flexible contracts and multiskilled staff, considering a retail store with uncertain demand.
- Researchers can use the datasets for benchmarking the performance of different approaches with the one presented by Porto et al. [1], and in consequence, they can find solutions to the same (or similar) type of personnel scheduling problem.
- The data from this research can be useful to determine which are the staffing levels by contract type, the cost-effective multiskilling levels, and the weekly shift programming that minimize the costs associated with training and over/understaffing in a retail store.
• Researchers can use these data to solve other types of personnel scheduling problems, such as a shift scheduling problem, a day-off scheduling problem, or a personnel assignment problem.
• The datasets include an Excel workbook that can be used to randomly generate staff demand instances according to a chosen coefficient of variation.

1. Data description

The human resources management in the retail industry faces predictable phenomena such as demand seasonality, as well as unpredictable phenomena such as demand uncertainty and unscheduled staff absenteeism [1]. Such phenomena produce periods of over and understaffing that can increase labor costs and deteriorate customer service levels (CSL) [1]. To minimize this mismatch between employee supply and demand, companies have used different labor flexibility strategies to solve personnel scheduling problems [2]. Porto et al. [1] expressed that there are four type of strategies typically implemented: flexible contracts, multiskilled staff, collaborative teams and temporary employees.

The database presented in this article was used by Porto et al. [1] to simultaneously solve a staffing and tour scheduling problem that combines the following two labor flexibility strategies: (i) flexible contracts, which allow to relax the duration of the shifts and the number of weekly hours employees must work; and (ii) multiskilled staff, employees trained to work on multiple departments, which allows store managers to transfer available multiskilled employees from overstaffed departments to understaffed departments.

Staffing is a type of personnel scheduling problem that determines how many employees are required in each type of contract, and how many of them will be multiskilled employees and in which task types (or store departments). In addition, tour scheduling is another type of personnel scheduling problem where days-off and shifts are scheduled simultaneously over a given planning horizon (typically one week). The solution of a staffing and tour scheduling problem must minimize labor costs while maintaining or improving the CSL.

The data presented in the next sections are derived from a home improvement retail store in Santiago, Chile. In this store the employees are assigned to different departments, each of which constitutes a store business unit and at that level is where the store’s training and employee scheduling decisions are made. This database contains real, processed, and simulated data that will be described below.

1.1. Real data

The real data include information about the sets and parameters that can be used to solve a hybrid flexibility strategy on personnel scheduling problem in the retail industry. Table 1 shows the notation, description, and values of these data.

Regarding the sets, it presents the days of the week that the store is operating (\(D\)), the time periods into which the operating day is divided (\(P\)), the contract types (\(C\)), shifts (\(T\)), workday types (\(J\)), store departments (\(L\)), workdays for each contract type (\(J_c\)) and the shifts for each contract type and workday type (\(T_{ij}\)). As for the parameters, it presents the costs of staff shortage (\(U\)) and staff surplus (\(O\)), and the base wage for each type of contract (\(G_c\)). Finally, it provides some additional parameters which are the start and end times of shifts (\(s_t, e_t\)), the start and end times of period (\(u_p, b_p\)), the length of each shift (\(h_t\)), and the number of weekly working hours according to the contract type (\(e_c\)).

Particularly, the parameter associated to the forecast of the number of employees required in each department for each day of the week and period (\(r_{ldp}\)) is presented in Table 2. This staff demand is expressed in terms of 28 time periods between 8:00 and 22:00, each period has a length of 30 min.
Table 1
Real data description from a retail store with six departments.

| Notation | Description | Value |
|----------|-------------|-------|
| **Sets** |             |       |
| $D$      | Days of the week, indexed by d. | $D = \{1, 2, 3, 4, 5, 6, 7\}$ |
| $P$      | Time periods into which the store's operating day is divided, indexed by p. | $P = \{1, 2, 3, 4, \ldots, 28\}$ |
| $C$      | Contract types, indexed by c. | $C = \{1, 2\}$ such that $c = 1 \rightarrow FT45$; $c = 2 \rightarrow PT30$ |
| $T$      | Shifts per day, indexed by t. Each shift is defined by a start and end time. | $T = \{1, 2, 3, 4, \ldots, 56\}$ |
| $J$      | Workday types, indexed by j. Defines the length of a shift. | $J = \{1, 2, 3, 4\}$ such that $j = 1 \rightarrow 5$ hrs per day; $j = 2 \rightarrow 6$ hrs per day; $j = 3 \rightarrow 9$ hrs per day; $j = 4 \rightarrow 10$ hrs per day |
| $L$      | Store departments, indexed by l. | $L = \{1, 2, 3, 4, 5, 6\}$ |

**Additional model sets derived from the sets defined above**

| Notation | Description | Value |
|----------|-------------|-------|
| $J_c$    | Workdays for employees with contract type $c, \forall c \in C, j \leq J$. Indicating that each contract type may have different feasible shift durations. | $J_1 = \{3\}$; $J_2 = \{1, 2, 4\}$ |
| $T_{cj}$ | Shifts for employees with contract type $c$ and workday type $j, \forall c \in C, j \in J_c, T_{cj} \subseteq T$. | $T_{c1} = \{3, 7, 11, 15, 19, 23, 27, 31, 35, 39, 42\}$; $T_{c2} = \{1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 40, 43, 45, 47, 49, 51, 53, 55, 56\}$; $T_{c2} = \{2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 41, 44, 46, 48, 50, 52, 54\}$; $T_{c4} = \{4, 8, 12, 16, 20, 24, 28, 32, 36\}$ |

**Parameters**

| Notation | Description | Value |
|----------|-------------|-------|
| $U$      | Cost per understaffed employee-period. | 61 US$; /period |
| $O$      | Cost per overstaffed employee-period. | 15 US$; /period |
| $G_c$    | Base wage of an employee according to the contract type $c, \forall c \in C$. | $G_1 = 100$ US$; /week; $G_2 = 90$ US$; /week |
| $r_{dp}$ | Number of employees required in department $l$, on day $d$, in period $p, \forall l \in L, d \in D, p \in P$. Start time of shift $t, \forall t \in T$. | See Table 2 |
| $s_{it}$ | End time of shift $t, \forall t \in T$. | $s_7 = 8 \, \text{Vr} = 1, 2, 3, 4; s_3 = 8.5 \, \text{Vr} = 5, 6, 7, 8; s_4 = 9 \, \text{Vr} = 9, 10, 11, 12; s_5 = 9.5 \, \text{Vr} = 13, 14, 15, 16; s_6 = 10 \, \text{Vr} = 17, 18, 19, 20; s_7 = 10.5 \, \text{Vr} = 21, 22, 23, 24; s_8 = 11 \, \text{Vr} = 25, 26, 27, 28; s_9 = 11.5 \, \text{Vr} = 29, 30, 31, 32; s_{10} = 12 \, \text{Vr} = 33, 34, 35, 36; s_{11} = 12.5 \, \text{Vr} = 37, 38, 39; s_{12} = 13 \, \text{Vr} = 40, 41, 42; s_{13} = 13.5 \, \text{Vr} = 43, 44; s_{14} = 14 \, \text{Vr} = 45, 46; s_{15} = 14.5 \, \text{Vr} = 47, 48; s_{16} = 15 \, \text{Vr} = 49, 50; s_{17} = 15.5 \, \text{Vr} = 51, 52; s_{18} = 16 \, \text{Vr} = 53, 54; s_{19} = 16.5 \, \text{Vr} = 55; s_{20} = 17 \, \text{Vr} = 56 |
| $e_{it}$ | End time of shift $t, \forall t \in T$. | $e_1 = 13 \, \text{Vr} = 1; e_2 = 13.5 \, \text{Vr} = 5; e_3 = 14 \, \text{Vr} = 2, 9; e_4 = 14.5 \, \text{Vr} = 6, 13; e_5 = 15 \, \text{Vr} = 10, 17; e_6 = 15.5 \, \text{Vr} = 14, 21; e_7 = 16 \, \text{Vr} = 18, 25; e_8 = 16.5 \, \text{Vr} = 22, 29; e_9 = 17 \, \text{Vr} = 3, 26, 33; e_{10} = 17.5 \, \text{Vr} = 7, 30, 37; e_{11} = 18 \, \text{Vr} = 4, 11, 34, 40; e_{12} = 18.5 \, \text{Vr} = 8, 15, 38, 43; e_{13} = 19 \, \text{Vr} = 12, 19, 41, 45; e_{14} = 19.5 \, \text{Vr} = 16, 23, 44, 47; e_{15} = 20 \, \text{Vr} = 20, 27, 46, 49; e_{16} = 20.5 \, \text{Vr} = 24, 31, 48, 51; e_{17} = 21 \, \text{Vr} = 28, 35, 50, 53; e_{18} = 21.5 \, \text{Vr} = 32, 39, 52, 55; e_{19} = 22 \, \text{Vr} = 36, 42, 54, 56 |

(continued on next page)
In order to maintain the optimization problem computationally tractable, Porto et al. [1] considered three store size (SS). Each SS represents the number of departments that the store has. The first SS has four departments, the second SS included five departments, and the third SS included six departments. Tables 1 and 2 show a SS of six departments, but to represent the other SS it would only be necessary to adjust L to the number of departments (i.e., $L = \{1, 2, 3, 4\}$ for the SS of four departments and $L = \{1, 2, 3, 4, 5\}$ for the SS of 5 departments). Also, for the parameter $r_{ldp}$, the demand for the corresponding SS can be selected in the order presented in Table 2.

The retailer data associated to the case study were provided by SHIFT SpA [3], a firm that optimizes the shift schedules of thousands of employees across Latin America. Two types of contracts were defined for the addressed personnel scheduling problem, based on established practices in the Chilean retail sector. The first contract type is FT45, for full-time employees working 45 weekly hours, while the second contract type is PT30, for part-time employees working 30 weekly hours. As mentioned before, the set $C$ represents the contract types, indexed by $c$, and the values $c = 1$ indicates a FT45 contract and $c = 2$ a PT30 contract.

As regards the workday types, there are three possible workdays in the PT30 contract (i.e., 5, 6 and 10 h) and there is one possible workday in the FT45 contract (i.e., 9 h). As noted above, the set $J$ represents the workday types, indexed by $j$, and the values $j = 1$ indicate a workday of 5 h, $j = 2$ a workday of 6 h, $j = 3$ a workday of 9 h and $j = 4$ a workday of 10 h.

As a result, Table 3 shows the set of shifts per day ($T$, indexed by $t$) that was obtained for each combination of contract type and workday type (i.e. workday groups). Column 1 shows the workday groups, which are structures formed by the type of contract, the working days per week, and the working hours per day. For example, the workday group FT45 ($5 \times 9$) represents an employee under a full-time contract that works 45 h per week, spread over 5 days per week and 9 h per day.

Columns 2 and 3 in Table 3 present detailed information for each shift belonging to a workday group. Column 2 indicates the start ($s_t$) and end ($e_t$) time associated to each shift $t$. Note that, the shifts can only start every 30 min and are limited by the store operating hours (i.e., 8:00–22:00). Column 3 shows for each shift presented in the Column 2 what is their respective value of $t \in T_{cj}$. Note that each workday group is equivalent to one of the sets $T_{cj}$, which represents the shifts for employees with contract type $c$ and workday type $j$, $\forall c \in C, j \in J, T_{cj} \subseteq T_c$. The workday group FT45 ($5 \times 9$) is equivalent to the set $T_{13}$, the workday group PT30 ($6 \times 5$) is equivalent to the set $T_{21}$, the workday group PT30 ($5 \times 6$) is equivalent to the set $T_{22}$, and the workday group PT30 ($3 \times 10$) is equivalent to the set $T_{24}$. The last column presents the number of shifts that can be considered for each workday group, whose total sum is equivalent to $|T| = 56$.

Finally, a detailed description of how we estimate the personnel demand and the staff costs will be addressed in the experimental design, materials, and methods section.

### 1.2. Data processed for the case study

In this section, we present the data processed for the case study solved in Porto et al. [1], which refers to the set of skill sets (i.e., $W$, $L_w$, $W_i$, $W_{i^\text{in}}, W_{i^\text{out}}$). Table 4 shows the set of skill sets

| $u_p$ | Start time of period $p$, $\forall p \in P$. |
| $b_p$ | End time of period $p$, $\forall p \in P$. |
| $h_t$ | Length of shift $t$, $\forall t \in T$. |
| $e_c$ | Number of weekly working hours according to the contract type $c$, $\forall c \in C$. |

---

**Table 1 (continued)**

| $u_p$ | Start time of period $p$, $\forall p \in P$. |
| $b_p$ | End time of period $p$, $\forall p \in P$. |
| $h_t$ | Length of shift $t$, $\forall t \in T$. |
| $e_c$ | Number of weekly working hours according to the contract type $c$, $\forall c \in C$. |

In order to maintain the optimization problem computationally tractable, Porto et al. [1] considered three store size (SS). Each SS represents the number of departments that the store has. The first SS has four departments, the second SS included five departments, and the third SS included six departments. Tables 1 and 2 show a SS of six departments, but to represent the other SS it would only be necessary to adjust $L$ to the number of departments (i.e., $L = \{1, 2, 3, 4\}$ for the SS of four departments and $L = \{1, 2, 3, 4, 5\}$ for the SS of 5 departments). Also, for the parameter $r_{ldp}$, the demand for the corresponding SS can be selected in the order presented in Table 2.

The retailer data associated to the case study were provided by SHIFT SpA [3], a firm that optimizes the shift schedules of thousands of employees across Latin America. Two types of contracts were defined for the addressed personnel scheduling problem, based on established practices in the Chilean retail sector. The first contract type is FT45, for full-time employees working 45 weekly hours, while the second contract type is PT30, for part-time employees working 30 weekly hours. As mentioned before, the set $C$ represents the contract types, indexed by $c$, and the values $c = 1$ indicates a FT45 contract and $c = 2$ a PT30 contract.

As regards the workday types, there are three possible workdays in the PT30 contract (i.e., 5, 6 and 10 h) and there is one possible workday in the FT45 contract (i.e., 9 h). As noted above, the set $J$ represents the workday types, indexed by $j$, and the values $j = 1$ indicate a workday of 5 h, $j = 2$ a workday of 6 h, $j = 3$ a workday of 9 h and $j = 4$ a workday of 10 h.

As a result, Table 3 shows the set of shifts per day ($T$, indexed by $t$) that was obtained for each combination of contract type and workday type (i.e. workday groups). Column 1 shows the workday groups, which are structures formed by the type of contract, the working days per week, and the working hours per day. For example, the workday group FT45 ($5 \times 9$) represents an employee under a full-time contract that works 45 h per week, spread over 5 days per week and 9 h per day.

Columns 2 and 3 in Table 3 present detailed information for each shift belonging to a workday group. Column 2 indicates the start ($s_t$) and end ($e_t$) time associated to each shift $t$. Note that, the shifts can only start every 30 min and are limited by the store operating hours (i.e., 8:00–22:00). Column 3 shows for each shift presented in the Column 2 what is their respective value of $t \in T_{cj}$. Note that each workday group is equivalent to one of the sets $T_{cj}$, which represents the shifts for employees with contract type $c$ and workday type $j$, $\forall c \in C, j \in J, T_{cj} \subseteq T_c$. The workday group FT45 ($5 \times 9$) is equivalent to the set $T_{13}$, the workday group PT30 ($6 \times 5$) is equivalent to the set $T_{21}$, the workday group PT30 ($5 \times 6$) is equivalent to the set $T_{22}$, and the workday group PT30 ($3 \times 10$) is equivalent to the set $T_{24}$. The last column presents the number of shifts that can be considered for each workday group, whose total sum is equivalent to $|T| = 56$.

Finally, a detailed description of how we estimate the personnel demand and the staff costs will be addressed in the experimental design, materials, and methods section.

### 1.2. Data processed for the case study

In this section, we present the data processed for the case study solved in Porto et al. [1], which refers to the set of skill sets (i.e., $W$, $L_w$, $W_i$, $W_{i^\text{in}}, W_{i^\text{out}}$). Table 4 shows the set of skill sets
Table 2  
Forecast of the number of employees required in each department for each day of the week (D) and each time period into which the retail store's operating day is divided (P).

| D | p |
|---|---|
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 |
| Department 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Department 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 |
| Department 3 | 1 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 1 | 1 |
| Department 4 | 1 | 1 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 2 | 2 |
| Department 5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Department 6 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 |

Department 1

```plaintext
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
4 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
5 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
6 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
7 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```
for the SS of six departments, but the building process is the same for the SS of four departments and the SS of five departments. In fact, the detailed building process of $W$ for any store size will be addressed in the experimental design, materials, and methods section. The other sets are derived from $W$.

The first set is $W$ which represents the skill sets, indexed by $w$. Given that the employees can be trained in a maximum of two store departments, the cardinality of $W$ is determined by the square of the SS. For example, if the SS is six departments then $|W| = 36$, such that this set includes 6 subsets with only one skill (i.e., single-skill) and 30 subsets with maximum two skills (i.e., multi-skill). Note that, each skill set $w$ is the result of the combination of each store department with itself and with the other departments. On one side, when a department is combined with itself the result is a single-skill subset. On the other hand, when a department is combined with another department the result is a multi-skill subset.

The second set is $L_w$ which represents the subset of store departments that are included in the skill set $w$, $\forall w \in W, L_w \subseteq L$. According to the value of $w$, $L_w$ can have one or maximum two departments. On one side, if $L_w$ has a single department, it is associated with a single-skill subset. On the other hand, if $L_w$ has two departments, it is associated with a multi-skill subset. In this last case, a distinction is made between which is the primary department and which is the secondary one.

In Table 4 for example, $L_1 = \{1\}$ is associated with the single-skill subset $w = 1$, where employees are trained to work only in department 1. $L_2 = \{1, 2\}$ is associated with the multi-skill subset $w = 2$, where employees are trained to work in department 1 as primary, and in department 2 as secondary. It should be noted that even though $L_2 = \{1, 2\}$ and $L_7 = \{2, 1\}$ are associated with multi-skill subsets where employees are trained to work in department 1 and department 2, they differ because their primary department is 1 and 2 respectively. This property also applies to the rest of the multi-skill subsets.

The third set is $W_l$ which represents the skill sets that allow an employee to work in the department $l$, $\forall l \in L, W_l \subseteq W$. To differentiate the primary department from the secondary department, we defined $W^\text{out}_l$ and $W^\text{in}_l$. First, $W^\text{out}_l$ represents the multi-skill sets where department $l$
Table 4
Description of the data processed for the case study of a retail store with six departments.

| Notation | Description | Value |
|----------|-------------|-------|
| \(W\)   | Set of skill sets, indexed by \(w\). Includes subsets with only one skill (i.e., single-skill) and subsets with maximum two skills (i.e., multi-skill). | \(W = \{1, 2, 3, 4, \ldots, 36\}\) |

Additional model sets derived from the sets defined above

- \(L_w\) Subset of store departments that are included in the skill set \(w, \forall w \in W, L_w \subseteq L\). Each subset has maximum two departments.
- \(W_l\) Skill sets that allow an employee to work in the department \(l, \forall l \in L, W_l \subseteq W\). Each skill set has \(|2|L| - 1\) skills.
- \(W_{l}^{\text{out}}\) Multi-skill sets where department \(l\) is considered the primary department of the employee, but he/she also has an additional skill to work in a secondary department, \(\forall l \in L, W_l \subseteq W\). Each multi-skill set has \(|L| - 1\) skills.
- \(W_{l}^{\text{in}}\) Multi-skill sets where an employee has a primary department different to \(l\), but he/she also has an additional skill to work in the department \(l, \forall l \in L, W_l \subseteq W\). Each multi-skill set has \(|L| - 1\) skills.

Parameters

- \(M_w\) Cost of training an employee according to the skill set \(w, \forall w \in W\). \(M_w = 1\ USD/week/multi-skilled\ employee\ →\ M_w = 0\ ∀w = 1, 8, 15, 22, 29, 36;\ M_w = 1\ ∀w = 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 16, 17, 18, 19, 20, 21, 23, 24, 25, 26, 27, 28, 30, 31, 32, 33, 34, 35\)

is considered the primary department of the employee, but he/she also has an additional skill to work in a secondary department, \(\forall l \in L, W_l \subseteq W\). Second, \(W_{l}^{\text{in}}\) represents the multi-skill sets where an employee has a primary department different to \(l\), but he/she also has an additional skill to work in the department \(l, \forall l \in L, W_l \subseteq W\).

In addition, we assume a minimal training cost per each hired multisilled employee, such that \(M_w = 1\ USD/week/multi-skilled\ employee\). Note that, the training cost is zero for the skill sets \(w\) with only one skill (i.e., single-skill).

Finally, we provided three files with the sets and parameters of the real and processed data written in AMPL, for each SS. The files named 'SS4-real-processed-data.dat', 'SS5-real-processed-data.dat' and 'SS6-real-processed-data.dat' contain the values that correspond to a SS of four departments, five departments and six departments, respectively. All files can be downloaded from the Mendeley repository that was provided in Data accessibility Section (see Specifications Table).
Table 5
Simulated datasets for the random realizations of the number of employees required in each department, day, and time period.

| CV  | 4 Departments | 5 Departments | 6 Departments |
|-----|---------------|---------------|---------------|
| 10% | SS4-CV10–01.dat, SS4-CV10–02.dat, SS4-CV10–03.dat, SS4-CV10–05.dat, SS4-CV10–07.dat, SS4-CV10–09.dat, SS4-CV10–10.dat | SS5-CV10–01.dat, SS5-CV10–02.dat, SS5-CV10–03.dat, SS5-CV10–05.dat, SS5-CV10–07.dat, SS5-CV10–09.dat, SS5-CV10–10.dat | SS6-CV10–01.dat, SS6-CV10–02.dat, SS6-CV10–03.dat, SS6-CV10–05.dat, SS6-CV10–07.dat, SS6-CV10–09.dat, SS6-CV10–10.dat |
| 20% | SS4-CV20–01.dat, SS4-CV20–02.dat, SS4-CV20–03.dat, SS4-CV20–05.dat, SS4-CV20–07.dat, SS4-CV20–09.dat, SS4-CV20–10.dat | SS5-CV20–01.dat, SS5-CV20–02.dat, SS5-CV20–03.dat, SS5-CV20–05.dat, SS5-CV20–07.dat, SS5-CV20–09.dat, SS5-CV20–10.dat | SS6-CV20–01.dat, SS6-CV20–02.dat, SS6-CV20–03.dat, SS6-CV20–05.dat, SS6-CV20–07.dat, SS6-CV20–09.dat, SS6-CV20–10.dat |
| 30% | SS4-CV30–01.dat, SS4-CV30–02.dat, SS4-CV30–03.dat, SS4-CV30–05.dat, SS4-CV30–07.dat, SS4-CV30–09.dat, SS4-CV30–10.dat | SS5-CV30–01.dat, SS5-CV30–02.dat, SS5-CV30–03.dat, SS5-CV30–05.dat, SS5-CV30–07.dat, SS5-CV30–09.dat, SS5-CV30–10.dat | SS6-CV30–01.dat, SS6-CV30–02.dat, SS6-CV30–03.dat, SS6-CV30–05.dat, SS6-CV30–07.dat, SS6-CV30–09.dat, SS6-CV30–10.dat |

1.3. Simulated data

The forecast values of the parameter $r_{ldp}$, which represents the number of employees required in department $l$, on day $d$, in period $p$, $\forall l \in L, d \in D, p \in P$, were presented in Table 2. But also, in order to evaluate the potential benefits of the joint use of flexible contracts and multiskilled employees in the face of different changes in staffing demand, Porto et al. [1] created simulated demands considering three different variability levels in each department: $CV = 10\%$, $20\%$ and $30\%$. Such that, CV is the coefficient of variation of demand from the forecast values presented in Table 2.

Combining the SS ($4$, $5$ and $6$ departments) and the CV ($10\%$, $20\%$, $30\%$) the resulting number of scenarios is $9$. A Monte Carlo simulation was used to randomly generate instances for demand in each store department following a zero-truncated normal probability distribution (this prevents negative demand values). In our case, $10$ demand instances were generated for each of the scenarios, resulting in $90$ instances. These $90$ simulated demands are provided in the text files listed in Table 5 and can be downloaded from the Mendeley repository that was provided in Data accessibility Section (see Specifications Table).

Each file contains the number of employees required in each store department, such that each row represents one of the $7$ days of the week, and each column represents one of the $28$ time periods into which the retail store’s operating day is divided. The file names are identified by a three-character code $i$-$j$-$k$, where $i =$ SS4, SS5, SS6 indicates the store size ($4$, $5$ or $6$ departments); $j =$ CV10, CV20, CV30 indicates the coefficient of variation ($CV = 10$, $20$ or $30\%$); and $k = 01$, $02$, $03$, $04$, $05$, $06$, $07$, $08$, $09$, $10$ represents the instance identification number ($10$ instances per scenario).

As an example, Figs. 1 and 2 were created to visualize the simulated demands. Fig. 1 shows the first instance of the simulated demand, for the scenario with six store departments and a coefficient of variation equal to $10\%$ (i.e., the data from the ‘SS6-CV10–01.dat’ file). And Fig. 2 shows the first instance of the simulated demand, for the scenario with six store departments and a coefficient of variation equal to $30\%$ (i.e., the data from the ‘SS6-CV30–01.dat’ file). In both instances, departments $3$ and $4$ are the ones with the highest demands, and department $1$ has the lowest demand, just like the forecast values of $r_{ldp}$ presented in Table 2. For $CV=10\%$, the number of required employees ranges from $1$ to $7$, considering all departments, days, and time periods. Meanwhile for $CV=30\%$ the number of required employees ranges from $0$ to $8$, considering all departments, days, and time periods.
Fig. 1. Data visualization for the first instance of the simulated demand, for the scenario with SS=6 and CV=10%.
**Fig. 2.** Data visualization for the first instance of the simulated demand, for the scenario with SS=6 and CV=30%.
Finally, a detailed description of the tool we used to perform the Monte Carlo simulations will be addressed in the experimental design, materials, and methods section.

2. Experimental design, materials, and methods

In this section, we present a complete description of the experimental design and methods used to acquire the data regarding the skill sets, the staff costs, and the forecast and simulated demand. First, we describe how the skill sets were designed. Second, we explain the reasoning behind the estimation of the base wage per type of contract and the costs of training, staff shortage and staff surplus. Third, we indicate the methods used to obtain the forecast staff demand in each store department. Finally, we provide a description of the program used for the Monte Carlo simulation, which generates random parameters of staff demand.

2.1. Skill sets

The sets $W$ were designed using square matrices such that the number of rows and columns is equal to the store size (SS). Fig. 3 shows the matrix notation, matrix representation, and linear vector for each set $W$ associated with each SS (4, 5 and 6 departments). In the matrix representation of sets $W$ the main diagonal represents the single-skill subsets (indicated in bold). Furthermore, the entries above and below the main diagonal represent the multi-skill subsets. Note that, in the model, the set $W$ was finally written as a linear vector, thus the matrix representation is for illustration purposes only.

For example, when SS=6, the matrix notation $a_{1,1}$ is associated with the single-skill subset $w=1$, where employees are trained to work only in department 1. In addition, the matrix notation $a_{1,2}$ is associated with the multi-skill subset $w=2$, where employees are trained to work in department 1 as primary, and in department 2 as secondary. As well as that, the matrix notation $a_{2,1}$ is associated with the multi-skill subset $w=7$, where employees are trained to work in department 2 as primary, and in department 1 as secondary. Finally, as already explained in Section 1.2, the other skill sets (i.e., $L_w$, $W_l^f$, $W_l^{out}$,$W_l^{in}$) are derived from $W$.
2.2. Staff costs

Regarding the cost structure, we set the base wage per type of contract according to the Chilean Labor Law effective in 2018. The weekly base wage for a full-time employee (i.e., FT45 contract) is equal to $G_1 = 100 \text{ US\$}/\text{week}$, whereas for a part-time employee (i.e., PT30 contract) is equal to $G_2 = 90 \text{ US\$}/\text{week}$. Note that, the weekly cost of a full-time employee is greater than the weekly cost of a part-time employee, since the former works 15-h more than the latter. However, one hour of work by a part-time employee is typically more costly than one hour of work by a full-time employee.

In regards to the cost of training associated with multi-skilled employees, we assume a minimal training cost of $M_w = 1 \text{ US\$}-\text{week/multi-skilled employee}$. Henao et al. [4], Henao et al. [5], Henao et al. [6], and Porto et al. [1] explain that this assumption allows for interpreting the results as an upper bound on the potential contribution of multiskilling to store performance.

As for the shortage and surplus costs, we assumed to be the same for all departments, time periods and days. The shortage cost is equivalent to the cost of the expected lost sales, and according to historical data of the retail store is 61 US$/period on average. The surplus cost represents an opportunity cost, incurred by paying for idle staff who could be assigned to other productive tasks in the store, and the average cost was determined to be 15 US$/period using historical data.

2.3. Demand forecast from a retail store

SHIFT SpA is a specialized firm that provides advisory services in workforce management for different companies in Latin America [3]. As mentioned before, SHIFT SpA provided the real data for our case study, which are derived of a home improvement retail store located in Santiago, Chile. In particular, the staff demand by department was obtained through a specialized software that runs in two steps: (1) forecast the sales and transactions and (2) generate the staff requirements.

In the first step, the software uses a multiple linear regression to forecast the amount of sales and number of transactions of the store by department, day, and time period. This procedure is based on historical data. To improve the estimation of the regression, the store must have at least 2–6 years of data. In the second step, the software transforms the predicted sales and transactions in a staff demand expressed in man-hours, considering the typical customer service times. Then, given a pre-established level of service, the number of employees required in each store department per day and time period is determined.

2.4. Monte Carlo simulation

In the Mendeley repository that was provided in Data accessibility Section (see Specifications Table), we supply an Excel workbook named ‘Monte Carlo Simulator.xlsx’ which contains the Monte Carlo simulation used to generate the instances for demand in each store department following a zero-truncated normal probability distribution (simulated data). This workbook contains one worksheet that can be used to generate demand instances one by one, for a given store department.

Since the staff demand follows a normal distribution, two parameters are required to generate the instances: (i) the forecast of the number of employees required in the department for each day of the week and each time period, which is given in Table 2; and (ii) the coefficient of variation, which can be selected by the user to represent the uncertainty in the staff demand that best fits its case study. In the Excel worksheet, these parameters or inputs are placed in a yellow fill cell, this color indicates that the values can be edited.
The rest of the input data is calculated using Excel formulas, and are presented in gray text to indicate that these cells should not be edited. In the worksheet, the standard deviation is calculated as the product between the forecast demand and the CV. Then some statistics are calculated to truncate the outputs. Since the distribution is zero-truncated, the standard score (z) and its respective quantiles were calculated.

Finally, the demand instance is calculated using the Excel formulas of inverse probability distribution and random values. The random values range between the quantile (associated with the standard value of demand equal to zero) and 1, and a step size of 0.0001. This ensures that the generated demand values are not negative, that is, they are greater than or equal to zero. The random parameters of staff demand outputs are organized in 7 rows representing the days of the week, and 28 columns representing the time periods into which the retail store’s operating day is divided. Such values are presented in blue text to indicate that these cells are the outputs and should not be edited.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

Acknowledgments

This research was supported by “Fundación para la Promoción de la Investigación y la Tecnología (FPIT)” under Grants 4.254 and 4.523 and Universidad del Norte under Grant 2017–13. Finally, the authors would like to thank their colleagues at SHIFT SpA for providing the real data from our case study.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2020.106066.

References

[1] A.F. Porto, C.A. Henao, H. López-Ospina, E.R. González, Hybrid flexibility strategy on personnel scheduling: retail case study, Comput. Ind. Eng. 133 (2019) 220–230, doi:10.1016/j.cie.2019.04.049.
[2] M. Mac-Vicar, J.C. Ferrer, J.C. Muñoz, C.A. Henao, Real-time recovering strategies on personnel scheduling in the retail industry, Comput. Ind. Eng. 113 (2017) 589–601, doi:10.1016/j.cie.2017.09.045.
[3] SHIFT SpA, Available: http://www.shiftlabor.com/, accessed 20 June 2020.
[4] C.A. Henao, J.C. Muñoz, J.C. Ferrer, The impact of multi-skilling on personnel scheduling in the service sector: a retail industry case, J. Oper. Res. Soc. 12 (66) (2015) 1949–1959, doi:10.1057/jors.2015.9.
[5] C.A. Henao, J.C. Ferrer, J.C. Muñoz, J. Vera, Multiskilling with closed chains in a service industry: a robust optimization approach, Int. J. Prod. Econ. 179 (2016) 166–178, doi:10.1016/j.ijpe.2016.06.013.
[6] C.A. Henao, J.C. Muñoz, J.C. Ferrer, Multiskilled workforce management by utilizing closed chains under uncertain demand: a retail industry case, Comput. Ind. Eng. 127 (2019) 74–88, doi:10.1016/j.cie.2018.11.061.