An Optimization Model for a New Scheduling Problem: Application in a Molecular Biology Lab

GUIERDY CONCHA [ID]1, FELIPE GAINZA [ID]2, AND ALFREDO CANDIA-VEJAR [ID]3

[1]Programa de Magíster en Gestión de Operaciones, Department of Industrial Engineering, Faculty of Engineering, Universidad de Talca, Curicó 3340000, Chile
[2]Centro de Investigación e Innovación, Viña Concha y Toro, Pencahue 3550000, Chile
[3]Department of Industrial Engineering, Faculty of Engineering, Universidad de Talca, Curicó 3340000, Chile

Corresponding author: Alfredo Candia-Vejar (acandia@utalca.cl)

ABSTRACT A new batch process scheduling problem is studied in this paper. The problem considers several machines where the jobs are processed and a team of specialists who analyze the jobs' results. Two operations that add complexity to the problem are the potential repetition of one or more processes and the probabilistic decision about the reprocessing of the jobs. A known State-Task Network partially represents the problem, so it is extended to include the two operations mentioned before and also the participation of a technical team. Based on this representation, an integer programming model is formulated for the integrated scheduling problem so that all the resources, material and human, are used in the best possible way. Actual data from a research lab located in the Region del Maule, Chile, illustrates the model's performance. The results showed that the scheduling obtained significantly contributed to planning the resources at the research lab. Changes in the technical team and instruments are possible so that the model could also be executed, only changing the corresponding parameters. Furthermore, additional experiments to the case study were conducted to study the performance of the model by increasing the size of the parameters.

INDEX TERMS Laboratories, operations research, scheduling, state-task network.

I. INTRODUCTION

Scheduling optimization comprises several classes of problems that include the well-known Flow Shop, Job Shop, and Open Shop. The essential components in a scheduling problem are the finite number of jobs and the finite number of machines. Usually, the case is separated when only one machine exists (Single Machine Models) and, on the other hand, several machines exist (Parallel Machine Models). In the classical book of [1], a detailed classification of scheduling problems is found. In our research, we use the Parallel Machine Model, and the machines are, in general, different from each other. A fundamental characteristic of the our scheduling problem is the machines can process several jobs simultaneously. In the literature, this situation is called batching machine. It is assumed that the processing times of the jobs in a batch are the same. Another characteristic of our problem is the recirculation of a batch. This operation means that a batch may visit a machine more than once.

The associate editor coordinating the review of this manuscript and approving it for publication was Wen-Sheng Zhao [ID].

Our problem also assumed that a team analyzes the results of the jobs processed by the machines. The team is multi-skilled, so two different people could operate various machines. The problem of multi-skilling in scheduling problems is well-known, and [2] presented a recent review paper. This problem has in project scheduling problems one of the more important applications. Our problem integrates both scheduling machines and scheduling a multi-skilling team. Therefore, a formal definition of the integrated problem and an optimization formulation are presented in this paper in Section 3. To our best knowledge, scarce research has been conducted in the study of the integrated problem. Some of them are commented on below and in the Literature Review.

The primary motivation of our research comes from a scheduling problem from one of the molecular research labs of a winery in Chile. The lab includes a set of machines, at least one per process (there may be parallel machines only for certain process), where there are machines that can be used only in a particular process, while there are machines that can be used in more than one process (only one process at a time). Machines run processes in batch, and each machine

can have different capabilities depending on the process it is running. Each machine that executes a process has to be managed by a human resource for the entire process time. Human resources have different skills, depending on their training, therefore they are qualified only to operate the machines and execute processes for which they are trained.

The main challenge in the lab involves managing many samples to be processed. The samples do not need to go through all the analysis processes continuously; they can stop at specific points and continue at another time. There are trained personnel to perform all types of analyses and personnel non-trained only to do specific tasks. Equipment and instruments for analysis are limited as well as human resources. In the lab, the planning and scheduling of resources and tasks are critical. These tasks are neither static nor constant during the year but rather depend on the analytical demand associated with the season. The case study presented refers to the grapevine plant production season. Thus, the need for efficient scheduling and allocation of resources must be for weekly/monthly terms. The models have to be flexible enough to reorder based on the demands that arise in the season.

A potential application of scheduling problems, including the characteristics described above, is semiconductor manufacturing. A complex job shop contains parallel machines, batch processes, and re-entrant process flows in semiconductor manufacturing. Wafer fabs are considered the first main stage of semiconductor manufacturing. At this stage, a Burn-in oven is a batching machine where several chips can be tested simultaneously. Complex scheduling problems also appear in other stages of semiconductor manufacturing, like in the Assembly and Test stages. A more detailed explanation of the application of scheduling problems in this area is found in [3].

Another application is the industry of chemical processes, such as the manufacture of lubricants, as explained by [4], where it is pretty common for intermediates to be shared among two or more products, batches of the material may have to be split or merged. The same material can be produced by more than one task, often due to recycling unreacted feedstocks. Batches of more than one product may be made simultaneously within the same plant. Similar problem characteristics can be found in polymer batch plant and steel-making casting plant as explained by [5].

Although the semiconductor and chemical process industries share similarities with the problem studied in this paper, one difference is the integration of machines and workers, which has not been explored in both industries. Also, in our situation, jobs depend on a known probability of reentrant, which has not been studied to the best of our knowledge.

Therefore, despite the literature's related works on batch processes, the research lab's problem includes other particularities and thus needing innovation in modeling the problem. So, new insights are necessary to formulate the scheduling problem as a mathematical programming model.

Our paper contributes to the literature on scheduling in several aspects.

- An optimization model for a new batch process scheduling problem is formulated. The model includes some typical constraints in traditional scheduling problems and introduces new types of complex constraints that integrate the scheduling of the technical staff and the machines. The existing State-Task-Network model has the particularity that operations (tasks) and intermediate and final products (states) can be mathematically optimized. Our optimization model extends the State-Task-Network, introducing new constraints and a technical team's participation in the whole system operation. At the same time, the optimization model presents a new integration between machines and a team to the shown by [6] since our model includes different characteristics. Thus, a challenging computational problem must be solved.

- A case study located in a research laboratory in Chile is presented. The scheduling problem existing in the research lab was solved using data from the 2019 season since 2020 was an abnormal year. The results delivered by the optimization model showed that the scheduling of people and equipment found could plan the work of the analysts efficiently. Additionally, in the presence of unexpected problems, the model could be used for re-planning.

- The laboratory analysts validated the results and considered it a potential tool for the management of the laboratory. Furthermore, the model could also be adapted to solve scheduling problems that include both machines and technical teams in other contexts like manufacturing.

The paper continues in Section II with the Literature review. In Section III, the problem and several of its characteristics are presented, while in Section IV the optimization model proposed is detailed. A case study illustrating the application of the model to a problem of analysis of samples in a research lab in Chile is presented in Section V. Additional experiments with the model are discussed in Section VI. Finally, conclusions are presented in Section VII.

II. LITERATURE REVIEW

We start the review by discussing scheduling problems that share characteristics with ours. Then, we present formulations and solution approaches that help us to the treatment of our problem.

A. RELATED MODELS FOR SCHEDULING PROBLEMS

According to Pinedo's notation [1], the problem presented in this paper is a Scheduling problem with Identical machines in parallel, machine eligibility restrictions, with recirculation and batch processing. However, not every characteristic of the problem is defined in Pinedo's notation, so this scheduling problem could not be classified as a classical one, being a

Flexible Job Shop (FJS) with recirculation that most resembles our problem.

There is a wide variety of types and bibliography of FJS models. Still, we highlight the papers that bear a relationship to this work, such as [6] who proposed what they called "a double flexible job-shop scheduling problem". Here, workers and machines are flexible simultaneously, considering their multi-objective function green production, human factor indicators, and minimization of the makespan. However, in this model, one of the characteristics is the following: "Each operation can be performed only once on one machine, and its sequence is respected for every job." This fact is different in our scheduling problem, where a batch could be reprocessed. On the same line, reference [7] studied the same problem of [6] but only aiming to minimize the makespan.

Reference [8] presented a solution for the flow shop scheduling with a multi-skill workforce and multiple types of machines using a framework based on machine learning. They take as inspiration a multi-stage, multi-machine, and multi-product manufacturing system operated by a multi-skilled workforce. Only a small example is described in the paper. Also, the machine scheduling is different from ours.

Reference [9] reviewed scheduling problems with batching. They presented research on models that integrate scheduling with batching decisions. However, their models do not include key characteristics of our problem as jobs recirculation and duplication.

In some papers have been studied problems faced in laboratories and modeled as scheduling problems. A work that shares common characteristics with our problem was developed by [10], which solved a scheduling problem in a quality control laboratory in the pharmaceutical industry. They aimed to minimize the total flow time with the least jobs not meeting the due date. Still, no mathematical programming formulation is presented as a basis for the problem solution. Reference [11] worked with a chemical testing laboratory from a water and waste-water company that had to do different tests and analyses on water and soil samples. Their goal was to develop a decision support system to schedule the lab operations to consider capacity allocation, batching, and sequencing issues. Reference [12] developed an extension of the Resource-Constraint Project Scheduling Problem (RCPSP), aiming to minimize the makespan with a genuine medical research project that dealt with the relationship between polyamine synthesis and cancer. It was a case study with resource availabilities and resource requirements varying with time. Reference [13] studied the problem to achieve optimal weekly programming of activities in a nuclear research laboratory and presented the preemptive Multi-Skill Resource-Constraint Project Scheduling Problem with a penalty for preemption. Reference [14] studied the problem of scheduling laboratory personnel in a clinical lab, assuring every workstation is filled, and the workers' skills are exercised by regularly rotating through all work areas.

Reference [2] reviewed multi-skilling and flexible resources in scheduling problems. The author presented 160 papers published between 2000 and the middle of 2020 and gave each paper according to the number and type of objective functions, the structure of parameters, mathematical formulations, solving methods, and case studies implemented by the authors. However, characteristics related to scheduling problems in the context of our problem are not discussed.

B. APPROACHES FOR PROBLEM SOLUTION

A valuable reference for our research is the review paper of [5], where they presented a general classification for scheduling problems of batch processes and the corresponding optimization models. In particular, they focused on the modeling of an optimization approach that utilizes discrete time. The modeling uses a network representation of the problem, and they commented that it had been a successful tool for solving practical problems in chemical engineering. Initially, this modeling was proposed in [4] and [15] where they detailed the MILP formulation for a scheduling problem for multipurpose batch chemical plants. The model has the particularity that operations (which they call "tasks") and the intermediate and final products (which they call "states") are explicitly included as nodes in the network representation of the problem, which they call State-Task-Network (STN). STN fits well for representing the problem. It has the flexibility to allow a process to feed more than one state to feed state early in the network, meaning a recirculation task.

In [16] are presented three MILP models for scheduling multi-tasking multipurpose batch processes in a scientific service facility. They used STN for representation, including two different objective functions: maximizing productivity and minimizing the makespan. They solved instances with commercial software to illustrate the formulations' capability and compared them with mathematical models in the literature.

About RCPSP, [17] proposed four discrete-time model formulations for the resource-constrained project scheduling problem with flexible resource profiles where the resource usage of an activity can be adjusted from period to period. They compared the results of each model and a priority-rule heuristic, using instances from Project Scheduling Problem Library (PSPLIB).

Reference [18] developed a model to improve the scheduling for phlebotomists aiming to reduce the excess work for personnel, hoping to balance the workload between the shifts. They used a two-stage stochastic integer linear programming with a stochastic component for the work demand and solved it with a heuristic algorithm they proposed. Reference [19] aimed to solve the problem of reducing the delivery of results after the deadline in a histopathology laboratory and spread the workload to reduce peaks of physical works. They modeled the process as batch processing machines and solved it with a two-phase decomposition approach.

In summary, scheduling problems offer various characteristics, optimization modeling, and solution approaches.

Related to our problem, the integration between machines and workers has recently started to be studied in the context of scheduling problems in recent years. Only three articles were found in this research line. To our best knowledge, no model exists in the literature that can fit all the particularities present in the problem addressed in this paper. Still, Networks STN with the structure of tasks and states will help formulate the optimization model in Section 4. The following section presents in detail the characteristics of the problem to be studied.

III. THE PROBLEM: CHARACTERISTICS AND ASSUMPTIONS
The scheduling problem to be studied has the following components.
- There is a set of jobs to be processed and analyzed. The jobs are grouped in batches to be processed by machines.
- There is a set of different machines. Each machine could process a batch in one operation (called batching machines in the literature).
- There is a set of processes to be executed by machines. One or more machines could complete a single process.
- There is a set of precedence relations between the processes that defining the route of a batch.
- There is a set of human resources that can operate (or not) the different machines or instruments.
- There is a period for the horizon planning of the scheduling, typically taking between one and four weeks.

Additional specifications about the relations among the different components are detailed below. Besides the precedence relations of processes, the flow not always follows only one path, as explained in Fig. 1, where each box represents a process that is a step in analyzing the job. Each arrow represents the flow of that job through processes. The flow is not always linear; it depends on the result of each process. For example, if the result of process $k+1$ is not satisfactory, it has to return to the beginning of process k and repeat from there (following dashed arrow). Still, if the result is fair, it continues through a straight arrow. Also, it is essential to note that there are specific points where an operation is finished the job is "duplicated." It can continue in two or more different paths simultaneously (as at the end of process k in Fig. 1).

Each process has a set of machines that can perform it, as represented in Fig. 2 by M. Each of these machines can be operated only by a qualified worker W. Some workers are flexible enough since they can use any machine (like W2). Workers with more limited qualifications can operate only specific machines (W1 and W3).

One of the critical characteristics of the problem is related to the use of time. It is assumed that the jobs must be processed every day, and it is necessary that every day, the specialists are working in the analysis done by the machines. Thus, the system continuously operates every day, but the processes and the analysis can take several minutes or hours. Therefore, the time is discretized in hours every day, and the entire period to be planned could take several days or
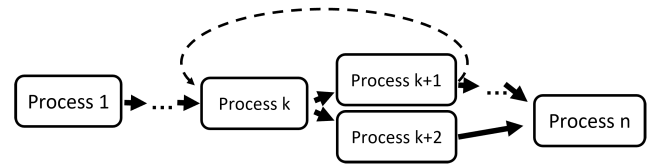

FIGURE 1. Representation of flow of jobs in a laboratory process. Each process has precedence relations that have to be fulfilled, represented by arrows. This precedence relations are not only horizontal, it is possible to repeat a processes as a reentrant task (dashed arrow). Also, there are "duplication" of tasks, as it is shown at the end of process k.
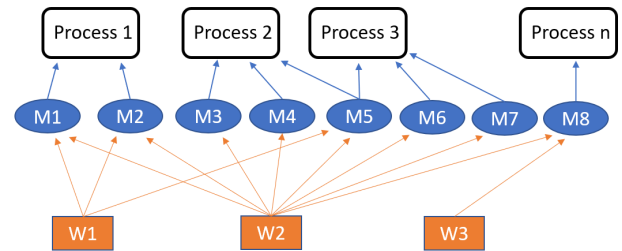

FIGURE 2. Example where workers and machines are flexible. M represents machines, W workers. As it can be noticed, a worker can operate more than one machine and therefore different processes.

weeks. In practice, several facts could interrupt the analysis of samples, either by changes in the equipment composition or the specialists. Under these conditions, convenient time in the scheduling is relatively short, for example, one or two weeks.

Therefore, the model to be formulated must consider this reality to optimize the resources during short periods while also allowing the tasks to remain. According to [5], this strategy for managing the schedule horizon has the advantage that scheduling constraints must be satisfied only at specific and known time points.

In the next section, the problem is formulated as an integer programming model.

IV. AN INTEGER PROGRAMMING FORMULATION FOR THE PROBLEM
A. THE OPTIMIZATION MODEL
In the structural model, precedence relations can be represented as a STN network. A STN has two types of nodes: state and task nodes, where states show the input, intermediate and final products, and task nodes represent the operations that take one or more input states to one or more output states [4]. This model has flexibility in using resources, as the machines work processing not a single but a group of jobs. This allows for the same machine to work in different processes and has various capacities for each process. The formulation is based on a discrete representation of time. The planning horizon is divided into an equal duration number of intervals, so all the scheduling happens between the interval boundaries. This representation has an advantage, as it provides a reference grid where all operations compete for the resources, facilitating the formulation [4]. We propose the following model based on this representation, including new features compared to the original model. They are human resources

and a probabilistic decision about the recirculation of jobs. Additionally, all processes that begin to run on a working day must end within the same working day, allowing workers not to stay overtime.

Definition of sets

I: Set of processes, $I = \{1, \ldots, |I|\}$

S: Set of states, $S = \{0, \ldots, |S|\}$

T: Set of time periods to schedule, $T = \{0, \ldots, |T|\}$

J: Set of machines and instruments, $J = \{1, \ldots, |J|\}$

H: Set of human resources, $H = \{1, \ldots, |H|\}$

H_j: Set of human resources that can operate a machine or instrument j, $j \in J$

J_h: Set of machines that can be operate by human resource h, $h \in H$

I_j: Set of processes that can be executed on a machine or instrument j, $j \in J$

J_i: Set of machines or instruments that can execute a process i, $i \in I$

T_s: Set of processes receiving jobs from state s, $s \in S$

$\overline{T_s}$: Set of processes that send their outputs to state s, $s \in S$

Definition of parameters

D_t: Last period t for a day, $t \in T$

V_{ij}^{min}: Minimum capacity of machine j when used to performing process i; $j \in J$, $i \in I$

V_{ij}^{max}: Maximum capacity of machine j when used to performing process i; $j \in J$, $i \in I$

C_s: Maximum storage capacity of jobs in state s, $s \in S$

Q_s: Number of jobs in state s at the beginning of schedule, $s \in S$

ρ_{is}: Proportion of input of state s that needs process i; $i \in I$, $s \in S$

$\overline{\rho_{is}}$: Proportion of output to state s that produce process i if there is no recirculation; $i \in I$, $s \in S$

$\overline{\rho 2_{is}}$: Proportion of output to state s that produce process i if there is recirculation; $i \in I$, $s \in S$

p_i: Execution time of process i, $i \in I$

$g_{it} \in \{0, 1\}$: Allows to decide if the process i at time t has recirculation or not, $i \in I$, $t \in T$. If $g_{it} = 1$, it means in process i at time t there is not recirculation, so it has to take the set $\overline{\rho_{is}}$ for parameters, if $g_{it} = 0$, in process i at time t there is recirculation, so it has to take set $\overline{\rho 2_{is}}$ for parameters. Its value is obtained from M_i which is the probability for a process i to not have recirculation with $i \in I$, and it can be calculated as follows: For all $i \in I$, $t \in T$, a random value between 0 and 1 is obtained that is compared with the current M_i value; if the random value is less than M_i, then $g_{it} = 1$, otherwise $g_{it} = 0$.

Decision Variables

$W_{ijht} \in \{0, 1\}$: $W_{ijht} = 1$ if human resource h starts to operate machine j to execute the process i at the beginning of period t. $W_{ijht} = 0$ otherwise, $i \in I$, $j \in J$, $h \in H$, $t \in T$

B_{ijht}: Amount of jobs that start to execute in process i in machine j by human resource h, at the beginning of period t; $i \in I$, $j \in J$, $h \in H$, $t \in T$

S_{st}: Number of jobs stored in state s at the beginning of period t; $s \in S$, $t \in T$

The Integer Programming (IP) formulation of the problem is as follows (1)–(15), as shown at the bottom of the next page.

The objective function is looking to maximize the number of tasks that start within the time to schedule. Constraints (1) and (2) assure that at any period, an idle machine j can start at most one process i, and if the machine begins to execute, that machine cannot begin any other process until it is finished. Constraints (3) and (4) assure that at any period, an idle human resource h can start to operate at most one machine j. If the human resource starts to use that machine, she (he) cannot begin any other process until it is finished. Constraints (5) and (6) indicate that a machine j cannot execute a process if it does not belong to I_j, and a human resource h cannot operate a machine j if he/she does not belong to H_j, respectively. Constraint (7) makes sure that the number of jobs executed by human resource h in machine j for process i at time t is bound by its maximum and minimum capacity. Constraint (8) indicates that at any time, t, the number of jobs stored in the state s can be at most its maximum capacity of storage. Constraint (9) indicates that the amount of jobs stored in-state s at time t is the amount that the state has at the previous period before plus the amount that came from the process that produced the state as output in time t, minus the amount that is used to feed processes. Constraint (10) makes sure that the initial number of jobs stored in state s is known. Constraint (11) indicates that scheduling starts at time 1. Constraint (12) makes sure that it must finish within the workday hours if a process begins to execute. Constraints (13), (14), and (15) specify the domains of variables. It is important to note that the IP formulated for the problem contains three sets of integer variables. The number of variables is dominated by the variables B and W. There are $|I| \times |J| \times |H| \times |T|$ integer variables due to B and the same number of binary variables due to W.

B. ILLUSTRATIVE EXAMPLE

Let us suppose that in a research lab, there are three machines (machine1, machine2, and machine3), three workers (worker1, worker2, and worker3), two process (process1 and process2), and three states (state1, state2, and state3), where state1 store all the jobs that don't have passed by any process, and state3 store all jobs that have passed all process, as is shown in Fig. 3. As the image suggests, state1 feed process1. Process1 takes a 1-time block to complete, and its output always goes 20% back to state1 and 80% to state2. State2 feeds process2, which takes a two-time block to complete. Process2 has two possible paths to send its output: all jobs go back to state two, or all jobs go forward to state3. There is a decision to choose which paths jobs must follow, which happens in the rhombus in the image. There is a probability to follow each path, and this probability is calculated by using empirical values. These probabilities are 4% for the dashed path (goes back to state 2) and 96% continuous path (goes forward to state 3). Workers work eight hours continuously per day. After those hours, any machine and any human resource cannot work. So if a process is scheduled to execute a

specific day, it must begin and finish within the eight working hours of the same day. Machine1 can execute process1 with ten jobs, machine2 can execute process1 with ten jobs and process2 with five jobs, and machine3 can execute process2 with five jobs. Worker1 can operate machine1, worker2 can manage machine2 and machine3, and worker3 can operate machine3. The schedule is for one week, and this is 40 hours. One block of time represents 1 hour. There is no limit to storage in any of the states.

According to this example, sets can be defined as follow: There are two process, so $I = \{1, 2\}$, three states so $S = \{1, 2, 3\}$, forty time periods to schedule $T = \{1, 2, 3, 4, 5, \ldots, 40\}$, three machines $J = \{1, 2, 3\}$, three human resources $H = \{1, 2, 3\}$, $H_1 = \{1\}$, $H_2 = \{2, 3\}$, $H_3 = \{3\}$. $I_1 = \{1\}$, $I_2 = \{1, 2\}$, $I_3 = \{3\}$. $J_1 = \{1, 2\}$, $J_2 = \{3\}$. $T_1 = \{1\}$, $T_2 = \{2\}$, $T_3 = \{\}$. $\overline{T_1} = \{1\}$, $\overline{T_2} = \{1, 2\}$, $\overline{T_3} = \{2\}$.

Parameters can be defined as shown in Table 1 for D_t, Table 2 for V_{ij}^{min} and V_{ij}^{max} and Table 3 for ρ_{is}, $\overline{\rho_{is}}$ and $\overline{\rho 2_{is}}$.

As there are limits in the storage capacity of any state, the maximum capacity is defined as a number big enough,

$C_1 = 1000$, $C_2 = 1000$, $C_3 = 1000$. As is the beginning of the processing, there are only jobs in the first state and no jobs in the intermediate and last state: $Q_1 = 500$, $Q_2 = 0$, $Q_3 = 0$.

Execution time is defined as $p_1 = 3$, $p_2 = 2$. Parameter $g_{i,t}$, as it is a random number between 0 and 1 and its value can be different in each run, it is calculated in a pre-process. The values of this parameter for this specific run for g_{it} with $i = 1$ were $g_{it} = 1 \ \forall \ t \in T$. With $i = 2$ the values were $g_{it} = 0 \ \forall \ t \in \{2, 5, 12, 20\}$, and $g_{it} = 1 \ \forall \ t \in T \mid t \notin \{2, 5, 12, 20\}$. As it can be noticed, parameter $g_{1,t}$ never is 0, which means that never is conditional reentrant, as it was defined in the example, while parameter $g_{2,t}$ vary, as it was defined that process 2 can be conditional reentrant.

Table 4 shows the results of the model application to the illustrative example with data from Table 1, Table 2, and Table 3. After running twice the model, the results of the two resulting scheduling are shown in Tables 4a and 4b and commented below.

As it can be noticed, all processes begin and end during the 8 hours of a workday. No processes overlap from one

$$\text{Maximize} \sum_{i \in I} \sum_{j \in J} \sum_{h \in H} \sum_{t \in T} W_{ijht}$$

$$\text{Subject to :} \sum_{h \in H} \sum_{i \in I} \sum_{t'=t}^{t-p_i+1} W_{ijht'} \leq 1 \quad \forall \ j \in J, t \in T - \{t : t > max_{(i \in I)} \ p_i\} \tag{1}$$

$$\sum_{h \in H} \sum_{i \in I} \sum_{t'=t}^{t+p_i-1} W_{ijht'} \leq 1 \quad \forall \ j \in J, t \in \{1, \ldots, max_{(i \in I)} \ p_i\} \tag{2}$$

$$\sum_{j \in J} \sum_{i \in I} \sum_{t'=t}^{t-p_i+1} W_{ijht'} \leq 1 \quad \forall \ h \in H, t \in T - \{t : t > max_{(i \in I)} \ p_i\} \tag{3}$$

$$\sum_{j \in J} \sum_{i \in I} \sum_{t'=t}^{t+p_i-1} W_{ijht'} \leq 1 \quad \forall \ h \in H, t \in \{1, \ldots, max_{(i \in I)} \ p_i\} \tag{4}$$

$$\sum_{t \in T} \sum_{h \in H} W_{ijht} = 0 \quad \forall \ i \notin I_j, j \notin J_i \tag{5}$$

$$\sum_{t \in T} \sum_{i \in I} W_{ijht} = 0 \quad \forall \ h \notin H_j, j \notin J_h \tag{6}$$

$$W_{ijht} V_{ij}^{min} \leq B_{ijht} \leq W_{ijht} V_{ij}^{max} \quad \forall \ i \in I, t \in T, j \in J_i, h \in H_j \tag{7}$$

$$S_{st} \leq C_s \quad \forall \ s \in S, t \in T \tag{8}$$

$$S_{st} = S_{st-1} + \sum_{i \in \overline{T_s}} \sum_{h \in H} \sum_{j \in J_i} (\overline{\rho_{is}} g_{it} + \overline{\rho 2_{is}} (1 - g_{it})) B_{ijht-p_i} - \sum_{i \in T_s} \sum_{h \in H} \sum_{j \in J_i} \rho_{is} B_{ijht} \ \forall s \in S, t \in T \tag{9}$$

$$S_{s0} = Q_s \quad \forall \ s \in S \tag{10}$$

$$W_{ijh0} = 0 \quad \forall \ i \in I, j \in J, h \in H \tag{11}$$

$$W_{ijht}(t + p_i - 1) \leq D_t \quad \forall \ i \in I, j \in J, h \in H, t \in T \tag{12}$$

$$S_{st} \in \mathbb{N} \quad \forall \ s \in S, t \in T \tag{13}$$

$$B_{ijht} \in \mathbb{N} \quad \forall \ i \in I, j \in J, h \in H, t \in T \tag{14}$$

$$W_{ijht} \in \{0, 1\} \quad \forall \ i \in I, j \in J, h \in H, t \in T \tag{15}$$

day to another in the five workdays, so workers can go home after their working day as no machines or human resources are scheduled to work after those 8 hours. According to the probabilistic component in the model, it can be noticed that the two schedules shown in Tables 4a and 4b are different even though the same parameters were used. The difference is more noticeable in the last two workdays scheduled (from hour 25 to hour 40). One important issue in the example is the flexibility of the model to manage a diversity of solutions provided by the probabilistic component. Thus, the lab manager could select the solution that best represents the conditions existent in the lab.

The following section presents the case study and the results obtained with the application of the optimization model.

V. THE CASE STUDY

A. A LAB FROM WINE INDUSTRY IN CHILE

Viña Concha y Toro (https://conchaytoro.com/) is the biggest winery company in Latin America (the second worldwide in vineyard surface) and is located in Chile. It has a Center of Research and Innovation (https://www.cii.conchaytoro.com/) that promotes applied research, technological development, and knowledge transfer to enhance the competitiveness and multi-origin excellence of the company and the Chilean wine industry in a dynamic international market with increasingly demanding consumers. One facility of this center is the Molecular Biology Laboratory.

The laboratory features cutting-edge equipment, enabling early detection of the primary diseases that affect vineyards (grapevine trunk diseases, viruses, and bacteria) using molecular tools. The equipment includes instruments for the isolation and identification of microorganisms (Microbiology area), nucleic acid purification and amplification for the identification of pathogens with automated sample processing capacity (Polymerase Chain Reaction (PCR) and quantitative Polymerase Chain Reaction (qPCR) area), and in vitro culture for the sanitation and multiplication of selected plant materials. This laboratory is responsible for processing and analyzing samples from all of the company's vineyards around the country, focusing on the mother blocks and the nursery processes.

The problem is related to grapevine viruses and trunk diseases, focusing on mother blocks. In this context, the most
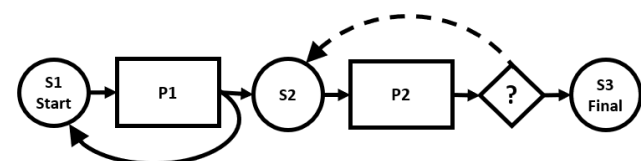
TABLE 1. Values of each parameter D_t.

Parameter	Value
$D_1, ..., D_8$	8
$D_9, ..., D_{16}$	16
$D_{17}, ..., D_{24}$	24
$D_{25}, ..., D_{32}$	32
$D_{33}, ..., D_{40}$	40

TABLE 2. Values of each parameter V_{ij}^{min} and V_{ij}^{max}.

Parameter	Value	Parameter	Value	Parameter	Value
V_{11}^{min}	24	V_{12}^{min}	24	V_{13}^{min}	0
V_{21}^{min}	0	V_{22}^{min}	48	V_{23}^{min}	48
V_{11}^{max}	24	V_{12}^{max}	24	V_{13}^{max}	0
V_{21}^{max}	0	V_{22}^{max}	48	V_{23}^{max}	48

TABLE 3. Values of each parameter ρ_{is}, $\overline{\rho_{is}}$ and $\overline{\rho2_{is}}$.

Parameter	Value	Parameter	Value	Parameter	Value
ρ_{11}	1	ρ_{12}	0	ρ_{13}	0
ρ_{21}	0	ρ_{21}	1	ρ_{23}	0
$\overline{\rho_{11}}$	0.2	$\overline{\rho_{12}}$	0.8	$\overline{\rho_{13}}$	0
$\overline{\rho_{21}}$	0	$\overline{\rho_{22}}$	0	$\overline{\rho_{23}}$	1
$\overline{\rho2_{11}}$	0	$\overline{\rho2_{12}}$	0	$\overline{\rho2_{13}}$	0
$\overline{\rho2_{21}}$	0	$\overline{\rho2_{22}}$	1	$\overline{\rho2_{23}}$	0

crucial step in establishing a profitable vineyard is planting clean plants [20]. The principal cause of all these pathogens are disseminated across long distances is by vegetative propagation (scions and rootstocks) [21], [22]. Eventually, this causes substantial crop losses, reduces plant vigor, and shortens the longevity of new vines [23], also having an impact on the main characteristics of wines, such as acidity, sugar contents, and pigments [24], [25]. The analysis of mother block plants leads to a control based on preventive measures such as sanitary selection for healthy plants to establish new vineyards [26].

The optimization model was applied to schedule the processes for identifying pathogens in wood and leaves in vineyards plants, shown graphically in Fig. 4. Around 1,500 samples of leaves have to pass through 17 processes with eight machines (represented in Fig. 5). About 900 samples of wood have to pass through 9 processes (depicted in Fig. 6) with seven machines or instruments. Also, four human resources for each were used as the base case, which means the actual scenario of the company's laboratory.

Because the purpose of the analysis is to find healthy plants, at the end of all processes, each sample is labeled as "accepted" or "rejected," which means the plant sample tested negative for diseases or positive, respectively. First, all the samples from leaves are processed. Suppose a leaf type sample is "rejected" once its analysis is finished. In that case, there is no necessity to analyze the wood of that specific plant, and only when leaf analysis finds healthy plants, its wood

FIGURE 3. STN representation of a small example of the problem. Circles represent the states, and rectangles represent the processes. Continuous arrows represent paths that occur almost always. In contrast, dashed arrows represent paths that occur only according to the given probability, calculated in the diamonds. It is decided which directions the task must follow: dashed path or continuous path.

TABLE 4. Two different scheduling (two runs with the same parameters) for a week (40 hours) are shown in Tables 4a and 4b. Each row represents each worker's scheduling at the given hour, where empty cells represent idle periods, P1 and P2 process1 and process2 respectively, and M1 to M3 machines 1 to 3.

a

Hour	1	2	3	4	5	6	7	8
Worker1	P1 M1	P1 M1					P1 M1	P1 M1
Worker2	P1 M2	P1 M2			P1 M2	P1 M2	P2 M2	P2 M2
Worker3					P2 M3	P2 M3	P2 M3	P2 M3
Hour	9	10	11	12	13	14	15	16
Worker1			P1 M1	P1 M1			P1 M1	P1 M1
Worker2	P1 M2	P1 M2	P2 M2	P2 M2			P1 M2	P1 M2
Worker3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3
Hour	17	18	19	20	21	22	23	24
Worker1			P1 M1	P1 M1			P1 M1	P1 M1
Worker2	P1 M2	P1 M2	P2 M2	P2 M2			P1 M2	P1 M2
Worker3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3
Hour	25	26	27	28	29	30	31	32
Worker1			P1 M1	P1 M1			P1 M1	P1 M1
Worker2	P2 M2	P2 M2	P1 M2	P1 M2			P1 M2	P1 M2
Worker3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3
Hour	33	34	35	36	37	38	39	40
Worker1	P1 M1	P1 M1			P1 M1	P1 M1		
Worker2	P2 M2	P2 M2	P2 M2	P2 M2	P2 M2	P2 M2	P2 M2	P2 M2
Worker3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3

b

Hour	1	2	3	4	5	6	7	8
Worker1	P1 M1	P1 M1					P1 M1	P1 M1
Worker2	P1 M2	P1 M2			P1 M2	P1 M2	P2 M2	P2 M2
Worker3				P2 M3	P2 M3	P2 M3	P2 M3	
Hour	9	10	11	12	13	14	15	16
Worker1			P1 M1	P1 M1			P1 M1	P1 M1
Worker2	P2 M2	P2 M2	P1 M2	P1 M2			P1 M2	P1 M2
Worker3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3
Hour	17	18	19	20	21	22	23	24
Worker1			P1 M1	P1 M1			P1 M1	P1 M1
Worker2	P2 M2	P2 M2	P1 M2	P1 M2			P1 M2	P1 M2
Worker3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3
Hour	25	26	27	28	29	30	31	32
Worker1	P1 M1	P1 M1					P1 M1	P1 M1
Worker2	P2 M2	P2 M2	P2 M2	P2 M2	P2 M2	P2 M2	P2 M2	P2 M2
Worker3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3	P2 M3
Hour	33	34	35	36	37	38	39	40
Worker1	P1 M1	P1 M1			P1 M1	P1 M1		
Worker2	P2 M2	P2 M2	P2 M2	P2 M2	P2 M2	P2 M2	P2 M2	P2 M2
Worker3	P2 M3	P2 M3	P2 M3	P2 M3			P2 M3	P2 M3

is analyzed. So a healthy plant means both leaf and wood samples were accepted.

The planning was made for two weeks and divided into 90 time periods representing an hour. Given that the optimization model has the number of samples stored in each state as a parameter, the scheduling for the first two weeks can be used as input for scheduling the next two weeks, planning for a longer time.

In Fig. 5 the representation of STN network for leaves processes is shown, and in Fig. 6 is the representation of STN network for wood processes. Each state (circles) acts as a queue through which samples must pass before entering the respective process (rectangles). There are processes where the sample is "duplicated," following two paths at the same time (for example, at the end of process P6 and so on in Fig. 5) and even "triplicated" at the end of process P5 in Fig. 5. This represents that there are stages in which the sample from the same plant may be being processed in more than one process at the same time. Also, at the end of specific processes, it is necessary to make a decision: the batch of samples continues to the following processes (continuous arrows in Fig.5 and Fig.6), or should it go back as reentrant (dashed arrows in Fig.5 and Fig.6). This is calculated according to a given probability. This random component simulates an error in the process, such as a human error, sample contamination, inadequate pipetting, and a spoiled sample. This error can be noticed only at the end of the process, and it means that the results obtained are not reliable, so they must be repeated.
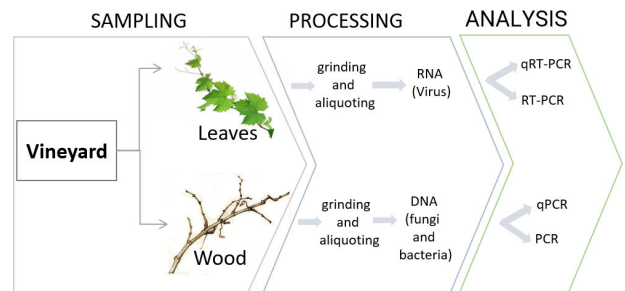
FIGURE 4. Flowchart in molecular biology laboratory in case study.

The model was run separately for each type of tissue sample, leaves and wood, and then the results were analyzed.

B. COMPUTATIONAL EXPERIMENTS AND ANALYSIS

The computational implementation of the optimization model was done using the solver CPLEX 12.8. The experiments run on a computer with Intel Core i7 8550U CPU 2.0 GHz and 8 GB RAM. Due to the nature of the solution method proposed, which has random components, the model runs several times to obtain statistics and dispersion analysis of results for the objective function and the schedule.

Model Application: Results and Findings.

The experiments aim to exploit the model's potential to provide (near) optimal scheduling of the machines and workers and some indicators about the use of the machines and time occupied by the workers. First, Fig. 5 and
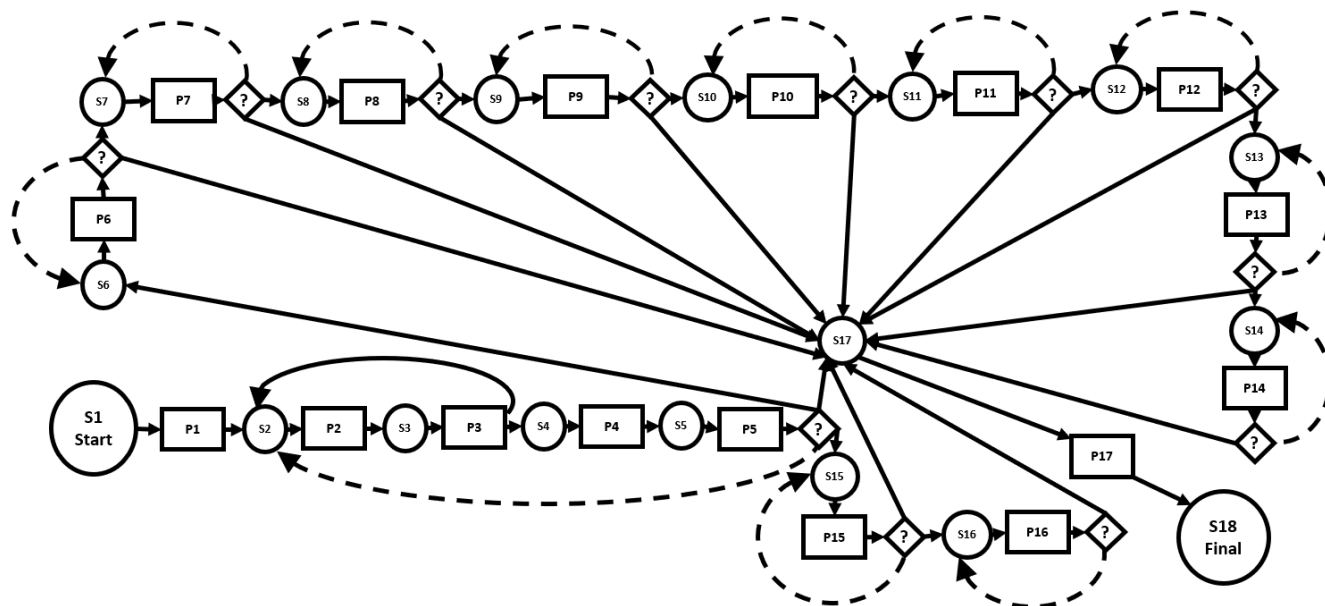
FIGURE 5. STN representation of leaves analysis processes. Circles represent the states, and rectangles represent the processes. Continuous arrows represent paths that occur almost always. In contrast, dashed arrows represent paths that occur only according to the given probability, calculated in the diamonds. It is decided which directions the task must follow: dashed path or continuous path.

TABLE 5. Relations between machines, processes, and workers for leaves analysis.

| Machines	1	2	3	4	5	6	7	8
Workers	1 and 2	1 and 2	1 and 2	1 and 2	1 and 2	3 and 4	3 and 4	3 and 4
Process	1	1	2	2	3	4, 15 and 16	5,6,7,8,9,10,11,12, 13 and 14	17

TABLE 6. Process time (in hours) for each process and the probability of the process not having conditional recirculation (probability to follow black arrows in Fig. 5) for leaves analysis.

Process	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Time	2	3	3	4	3	3	3	3	3	3	3	3	3	3	4	4	1
Probability	1	1	1	1	0.95	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	1

Tables 5 and 6 illustrate the input data for applying the model on leaves. Table 7 presents the model results by detailing basic information like the value of the objective function and the running time. Also, statistics about the busy time of workers and machines is presented. Table 8 illustrates the detailed scheduling of workers and the specification of the process assigned to a machine. The same scheme of input data and results are then presented for the case of woods.

1) LEAVES SAMPLES: INSTANCE AND RESULTS
a: INPUT DATA SPECIFICATION
The instance used for leaves analysis consists of 90 hours, four workers, eight machines, 17 processes, and 18 states. The relations between workers, machines, and processes are shown in Table 5. Workers 1 and 2 represent technical human resources, while Workers 3 and 4 represent analyst human resources. Machines 1,2,3,4 and 5 can process batches

with 24 samples, machine 6 with 24 samples for process 4, and 96 samples for processes 15 and 16, machine 7 with 288 samples for processes 5,6,7,8,9,10 and 11, and 96 samples for processes 12, 13, 14, machine eight can process 288 samples at the same time. At the end of process 3 (see Fig. 5), 87.5 % of the samples processed go to state 4, and 12.5% return to state 2. The model size is 100,646 variables, of which 49,504 are binary and 159,889 constraints.

In Table 6 the processing time (expressed in hours) is shown for each process and the probability that the output of a process can have recirculation or not (follow dashed arrows or straight arrows in Fig. 5). Each state had unlimited capacity, and the start amount of samples in each state (representing the beginning of the season) is 0, except State 1 that has all the samples. State 1 is the initial number of samples that have not gone through any process, and State 18 is the number of samples that have gone through all the processes, while the rest are intermediate states.
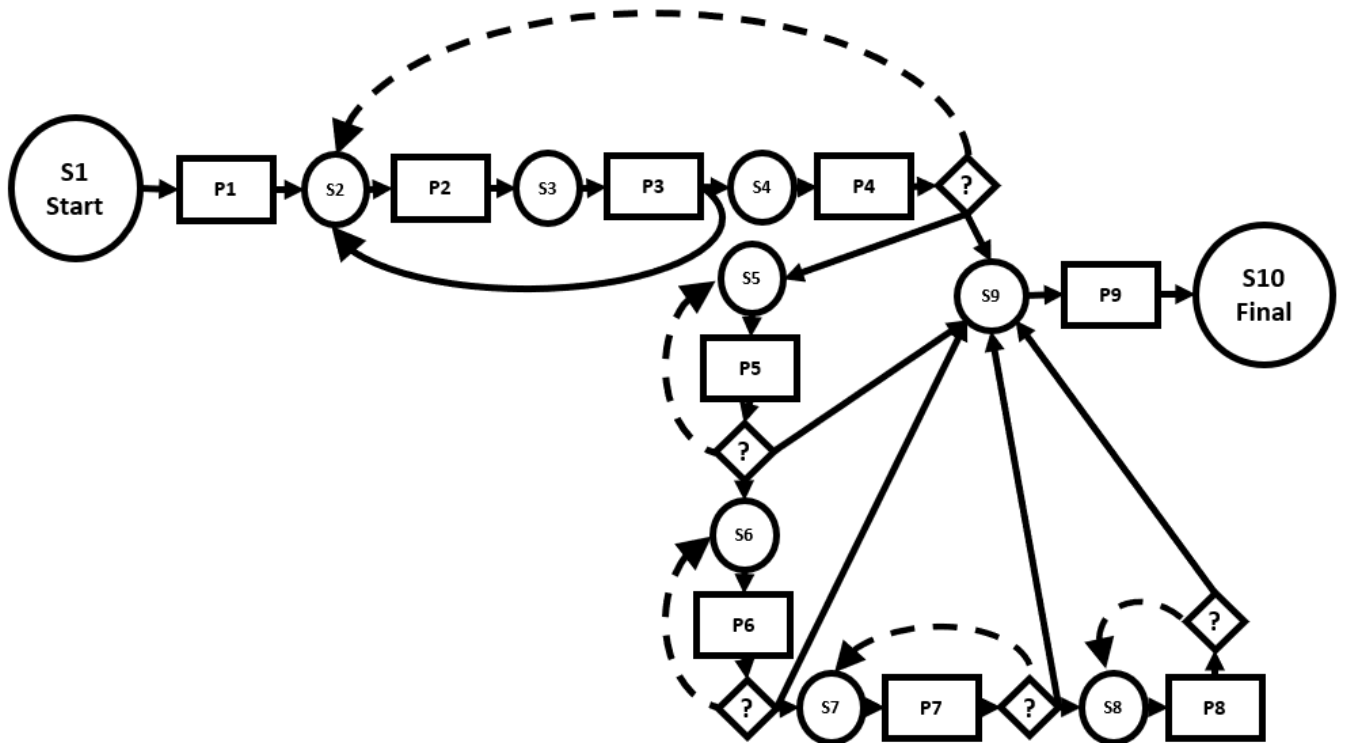
FIGURE 6. STN representation of wood analysis processes. Circles represent the states, and rectangles represent the processes. Continuous arrows represent paths that occur almost always. In contrast, dashed arrows represent paths that occur only according to the given probability, calculated in the diamonds. It is decided which directions the task must follow: dashed path or straight path.

TABLE 7. Average, Standard deviation, maximum and minimum for objective function value, execution time, and busy hours for each worker and machine for leaves analysis.

	Average	%	Max. Value	Min. Value	Standard Deviation
Execution Time (s)	10.66	-	48.13	4.82	9.33
Objective Function Value	109.2	-	125	105	5,59
Busy Time periods for Worker 1	89.60	99.56	90	88	0.68
Busy Time periods for Worker 2	89.60	99.56	90	88	0.75
Busy Time periods for Worker 3	59.95	66.61	79	42	9.28
Busy Time periods for Worker 4	62.70	69.67	88	47	10.77
Busy Time periods for Machine 1	24.00	26.67	32	16	3.84
Busy Time periods for Machine 2	22.30	24.78	32	10	4.74
Busy Time periods for Machine 3	40.95	45.50	57	18	10.54
Busy Time periods for Machine 4	40.80	45.33	51	24	6.77
Busy Time periods for Machine 5	51.15	56.83	69	39	10.64
Busy Time periods for Machine 6	74.80	83.11	80	72	2.28
Busy Time periods for Machine 7	37.20	41.33	75	27	13.11
Busy Time periods for Machine 8	10.65	11.83	14	9	1.27

Given that during the first two weeks of the season, there is only work for the first steps of the workflow, technicians are mostly busy, and analysts are mostly idle. So, to illustrate the scheduling of technicians and analysts, we run the model for the first two weeks (results not shown) and then another two weeks (results shown). So, we take the results obtained by the first two weeks (put values of S_{s90} in Q_s, $\forall s \in S$) as the input for the two successive weeks.

b: MODEL RESULTS FOR LEAVES SAMPLES

After running the model 20 times for the instance described above, the results are shown below. For leaves analysis, in Table 7 is shown the average, standard deviation, the maximum and minimum value of the objective function, and the total of busy hours for human resources and machines. It shows an average of 109.2 for the objective function, which means the sum of all processes that start and finish within

TABLE 8. Representative scheduling for the second 90 hours for leaves analysis. Each row represents each worker's scheduling at the given hour, where empty cells represent idle periods, P1 to P17 process 1 to process 17, M1 to M8 machines 1 to 8.

Hour	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Worker1	P3 M5				P3 M5			P3 M5		P1 M1	P3 M5				P1 M2		P1 M1	
Worker2	P2 M3				P2 M4			P2 M3		P2 M3		P2 M3			P3 M5			
Worker3						P4 M6					P4 M6				P4 M6			
Worker4	P15 M6						P11 M7			P11 M7		P12 M7			P12 M7			

Hour	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Worker1	P3 M5				P2 M3			P2 M4			P3 M5			P2 M4		P3 M5		
Worker2	P2 M4				P3 M5			P2 M3			P2 M4			P3 M5		P2 M4		
Worker3		P13 M7				P4 M6				P13 M7				P14 M7		P13 M7		
Worker4	P4 M6							P13 M7	P17 M8	P4 M6				P4 M6				P17 M8

Hour	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
Worker1	P1 M1			P2 M4		P1 M2			P1 M2	P3 M5				P2 M3		P3 M5		
Worker2	P3 M5			P3 M5						P2 M4				P3 M5		P2 M3		
Worker3	P14 M7			P12 M7			P12 M7				P13 M7			P14 M7				P17 M8
Worker4	P4 M6					P4 M6				P4 M6				P17 M8		P4 M6		

Hour	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
Worker1	P2 M4				P2 M3		P2 M3			P2 M3				P2 M3		P2 M3		
Worker2	P1 M2		P1 M1		P2 M4		P1 M1			P2 M4			P2 M4			P2 M4		
Worker3	P16 M6				P17 M8		P16 M6			P8 M7				P9 M7		P10 M7		
Worker4	P5 M7				P6 M7		P7 M7			P15 M6				P17 M8		P15 M6		

Hour	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
Worker1	P1 M1		P3 M5			P1 M1		P1 M1		P1 M1		P1 M2		P2 M3		P1 M2		
Worker2	P1 M2		P1 M2		P1 M2			P2 M4		P1 M2		P2 M4		P1 M2		P1 M1		
Worker3	P11 M7			P12 M7		P12 M7				P13 M7				P13 M7		P17 M8		P17 M8
Worker4	P17 M8	P17 M8	P17 M8	P17 M8	P17 M8	P15 M6				P16 M6				P16 M6				P17 M8

these 90 hours is around 109 on average, with a standard deviation of 5.59. Worker 1 and Worker 2, which represent technician human resources, are busy almost every hour in every running. Worker 3 and Worker 4, who represent analyst human resources, always have an average idle time of 30 and 28 hours. Representative scheduling of these results is shown in Table 8, which illustrates how work time is assigned for each worker and machine.

2) WOOD SAMPLES: INSTANCE AND RESULTS
a: INPUT DATA SPECIFICATION

The instance used for wood analysis has 90 hours, four workers, seven machines, nine processes, and ten states. The relation between workers, processes, and machines is shown in Table 9. Workers 1 and 2 represent technical human resources. In contrast, Workers 3 and 4 represent analyst's human resources Machines 1,2,3,4, and 5 can process batches with 24 samples, machine 6 with 96 samples for processes 4, 5, 6, and machine 7 with 288 samples for process 9. At the end of process 3 (see Fig. 6), 87.5 % of the processed samples go to state 4, and 12.5% returns to state 2. The model size is 46,774 variables, of which 22,932 are binary and 75,687 constraints.

In Table 10 the processing time (expressed in hours) is shown for each process and the probability that the output of a process will have recirculation or not (follow dashed arrows or straight arrows in Fig. 6). Each state had unlimited capacity, and the start amount of samples in each state (representing the beginning of the season) is 0, except State 1 that has all the samples. State 1 is the initial number of samples that have not gone through any process. State 10 is the number of samples that have gone through all the processes, while the rest are intermediate states.

b: MODEL RESULTS FOR WOOD SAMPLES

Given that during the first two weeks of the season, there is only work for the first steps of the workflow, technicians are mostly busy and analysts are mostly idle. The results shown below are the schedule for the second two weeks taking as input for the number of samples at each state the results obtained by the first two weeks, where there are samples in almost every state. After running the model 20 times for the instance described above, the results are shown below.

For wood analysis, Table 11 is shown the average, standard deviation, the maximum and the minimum value of the objective function, and the total of busy hours for human resources and machines. It shows an average of 88.85 for the objective function, which means the sum of all activities that start and finish within these second 90 hours is around 89 on average, with a standard deviation of 1.9. Also, Worker 1 and Worker 2, the technician human resources, are busy around 85% of the schedule, while Worker 3 and Worker 4, who represent analyst human resources, are busy around 89% of the time. Representative scheduling of these results is shown in Table 12, which illustrates how work time is assigned for each worker and machine.

3) SENSITIVITY ANALYSIS

To evaluate the model's flexibility and assess its performance in different scenarios, some tests were carried out by adding and removing human resources. The model was run for 135 hours in these trials, representing three weeks, with 9 hours per workday. The scenarios are the same as leave analysis, showing the base case (2 technicians and two analysts), adding a technician, adding an analyst, and removing an analyst. Results are shown in Table 13. On average, it can be noticed that the scenario removing a technician takes the

TABLE 9. Relations between Workers, processes, and Machines for wood analysis.

Machines	1	2	3	4	5	6	7
Workers	1 and 2	1 and 2	3 and 4	3 and 4	1 and 2	3 and 4	3 and 4
Process	1	1	2	2	3	4, 5, 6, 7 and 8	9

TABLE 10. Processing time (in hours) for each process and probability for the process to NOT have conditional recirculation (probability to follow black arrows in Fig. 6) for wood analysis.

Process	1	2	3	4	5	6	7	8	9
Process Time	5	4	3	3	3	3	3	3	1
Probability	1	1	1	0.95	0.93	0.93	0.93	0.93	1

TABLE 11. Average and standard deviation for objective function value, execution time and busy hours for each worker and machine for wood analysis.

	Average	%	Max Value	Min Value	Standard Deviation
Execution Time (s)	64.5861	-	276.112	13.655	59.46
Objective Function Value	88.85	-	94	86	1.90
Busy Time periods for Worker 1	76.55	85.06	82	71	2.93
Busy Time periods for Worker 2	74.3	82.56	81	70	3.18
Busy Time periods for Worker 3	79.55	88.39	86	72	3.83
Busy Time periods for Worker 4	79.7	88.56	85	74	3.28
Busy Time periods for Machine 1	45.25	50.28	50	35	4.13
Busy Time periods for Machine 2	43.5	48.33	50	40	3.28
Busy Time periods for Machine 3	49	54.44	64	32	9.35
Busy Time periods for Machine 4	34.8	38.67	52	20	9.09
Busy Time periods for Machine 5	62.1	69.00	69	57	2.59
Busy Time periods for Machine 6	69	76.67	75	63	3.64
Busy Time periods for Machine 7	6.45	7.17	7	6	0.51

shortest execution time while adding a technician takes the largest execution time.

C. FINAL DISCUSSION ABOUT THE CASE STUDY

Some findings of the performance of the optimization model when applied to data from a research lab are commented on as follows.

- All resources have high occupation rates, especially human resources. Worker 1 and Worker 2 (who represent technicians) for leaves analysis and Worker 3 and Worker 4 (who represent analyst human resources) for wood analysis.
- Machines that are the least busy when talking about usage time work on the processes that take the least time to be completed. On the other hand, machines with high occupation rates work on processes that take the longest and execute the most processes. Some machines work on the same process, explaining the difference between the maximum and minimum periods that they were busy. So when a machine was used the most, the other machine that executes the same process was used the least, and vice-versa.

- The problem's random component can explain the difference between the maximum and the minimum number of samples in each state. A group of samples can be re-processed several times in only one process (dashed arrows in Fig. 5 for leaves and Fig. 6 for wood), therefore not advancing in the workflow. On other occasions, the random part indicates that they only need to go through that process once, so they move faster through the workflow. This random component simulates when an error occurs in the process, such as a human error, sample contamination, inadequate pipetting, and a spoiled sample, which can be noticed only at the end of the process.
- In the sensitivity analysis, the best-case scenario in terms of objective function value and number of finished samples seems to be adding a technician to the base case where the three technicians have a high occupation rate. Adding an analyst to the base case makes no significant difference. It only reduces the workload for analysts but cannot achieve greater performance than the base case scenario. Removing an analyst is the worst scenario. In terms of objective function value and the number of

TABLE 12. Representative scheduling of the second 90 hours for wood analysis. Each row represents the scheduling for each worker at the given hour, where empty cells represent idle periods, P1 to P9 process 1 to process 9, M1 to M7 machines 1 to 7.

Hour	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Worker1		P1 M1					P3 M5				P3 M5			P3 M5		P3 M5		
Worker2		P3 M5				P1 M2					P1 M2							
Worker3		P8 M6				P2 M4					P2 M3				P2 M3			
Worker4		P2 M3					P8 M6					P4 M6			P2 M4			

Hour	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Worker1	P3 M5			P3 M5			P3 M5				P3 M5				P1 M1			
Worker2	P1 M2										P1 M1				P3 M5			
Worker3	P2 M4					P2 M4				P2 M3				P2 M4			P9 M7	
Worker4			P8 M6				P2 M3			P4 M6			P8 M6			P5 M6		

Hour	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
Worker1	P1 M2					P3 M5				P1 M1						P3 M5		
Worker2		P3 M5				P1 M1				P3 M5					P1 M2			
Worker3	P2 M3				P2 M4			P9 M7		P4 M6			P5 M6			P6 M6		
Worker4	P6 M6			P7 M6			P8 M6			P2 M4				P2 M4			P9 M7	

Hour	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
Worker1			P3 M5			P3 M5				P1 M1						P3 M5		
Worker2	P3 M5				P1 M1					P3 M5					P1 M2			
Worker3	P7 M6			P8 M6			P4 M6			P5 M6			P6 M6			P7 M6		
Worker4	P2 M3				P2 M3			P9 M7		P2 M3				P2 M4			P9 M7	

Hour	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
Worker1	P3 M5				P1 M1					P3 M5					P1 M2			
Worker2	P1 M2					P3 M5				P1 M1					P3 M5			
Worker3	P2 M4				P2 M3			P9 M7		P2 M4				P2 M3			P9 M7	
Worker4	P8 M6			P4 M6			P5 M6			P6 M6			P7 M6			P8 M6		

finished samples, it increases the workload of the only analyst left and reduces the busy times for machines used by analysts.

An interesting aspect of this research is using the IP formulation proposed to solve actual instances in reasonable times. The results obtained (and already discussed in this section) from applying the solver CPLEX to the leaves and wood samples took 5 minutes or less. Besides, all the experiments carried out in the Sensitivity Analysis also took seven minutes or less. So, the ILP model is the right solution for the size of the problem presented here. However, given the problem's high complexity, exact algorithms can solve at the optimum only instances of limited size. We conducted additional experiments by increasing the number of periods to four weeks. In this case, the personal computer runs out of memory after 27,523 seconds with a gap of 0.5%.

A final but essential issue related to the case study is the satisfaction of the Concha y Toro lab team with the proposed model. The team is used to planning manually, based on previous experiences and intuition, taking into account the availability of equipment and skills of Human Resources. This is evaluated constantly based on the seasonality of the vineyard growing and the availability of samples derived from the analytical requirements of each season (annual). With the "non-static" nature of the season, the demand for analysis varies. It may be necessary to incorporate new analyses that

are different from those already planned. Also, productive requirements can arise within the season, human resources and machines can be suddenly designated to another particular R&D project with higher priority. All these facts make it difficult to planning horizons higher than one month. Since the planning was done manually, it does not incorporate an optimization integrating machines and specialists. Thus, it could cause incorrect estimations of lead times, delays, and sometimes a critical idle time of specialized human resources. In the short to medium term, it will probably be necessary to grow in machines and personnel. Also, the proper authorities have additional regulatory components, and doing the manual planning can not cope. So, the team leader (F. Gainza) decided to explore new approaches, which led to the decision to use optimization models for scheduling in the lab. The lab team actively participated in the different steps of formulating the optimization model, helping to identify the components of the problem (parameters and constraints), defining the aim of the objective function, and giving continuous feedback as the model was built and tested. Based on the results of the case study, the lab decided to hire a new technician.

Also, it is expected that the adoption of the proposed model in the line of analytical processes of the molecular biology laboratory subject of study entails a substantial improvement in the efficiency of use of human resources and equipment. It also allows being used as a tool to evaluate points of

TABLE 13. Average and standard deviation for objective function value, execution time, busy time periods for human resources and machines, and number of samples in the last state for different scenarios.

Type of Sensitivity	Base Case		Adding 1 Analyst		Adding 1 Technician		Removing 1 Analyst	
	Average	Std Dev	Average	Std Dev	Average	Std Dev	Average	Std Dev
Objective Function Value	167.13	0.74	142.47	34.58	213.93	0.26	149.90	0.31
Execution Time (s)	97.55	60.81	80.09	41.62	192.16	140.89	38.06	90.23
Busy Time periods for Technician 1	132.80	0.68	132.79	0.71	133.33	0.90	132.40	0.75
Busy Time periods for Technician 2	132.73	0.88	132.74	0.99	133.20	0.86	132.65	0.75
Busy Time periods for Technician 3	-	-	-	-	133.27	0.70	-	-
Busy Time periods for Analyst 1	71.80	4.07	54.16	6.36	72.80	3.28	114.95	0.22
Busy Time periods for Analyst 2	105.87	4.58	74.11	6.45	107.13	3.04	-	-
Busy Time periods for Analyst 3	-	-	50.89	7.27	-	-	-	-
Busy Time periods for Machine 1	63.33	4.82	59.37	6.15	59.47	8.57	62.60	6.68
Busy Time periods for Machine 2	59.60	4.22	63.89	5.35	64.53	8.57	61.30	6.53
Busy Time periods for Machine 3	40.00	6.77	40.89	8.14	84.20	13.25	46.65	8.00
Busy Time periods for Machine 4	32.80	6.56	32.84	8.03	82.60	14.65	38.55	10.89
Busy Time periods for Machine 5	69.80	1.37	68.53	2.06	100.00	10.12	55.95	9.60
Busy Time periods for Machine 6	108.00	0.00	107.79	0.92	106.13	2.07	49.00	3.15
Busy Time periods for Machine 7	57.80	2.11	59.37	2.14	60.60	2.59	55.95	3.27
Busy Time periods for Machine 8	11.87	0.35	12	0	13.20	0.56	10	0
Samples in State 18 at time 135	3129.6	101.34	3183.16	203.05	3475.20	230.06	2577.60	64.40

TABLE 14. Average running time in seconds (Avg. t) and Average GAP (Avg. GAP) for the additional experiments.

	4 weeks		6 weeks		8 weeks	
	Avg. t	Avg. GAP	Avg. t	Avg. GAP	Avg t	Avg. GAP
I1 Instance	3302.79	1.60	3600	1.98	3600	2.56
I2 Instance	3600	5.58	3600	37.57	3600	–

TABLE 15. Average running time in seconds (Avg. t) and Average GAP for the new experiments missing Constraint (12).

	4 weeks		6 weeks		8 weeks	
	Avg. t	Avg. GAP	Avg. t	Avg. GAP	Avg. t	Avg. GAP
I1 Instance	3421.93	2.61	3600	4.97	3600	526.99
I2 Instance	3600	401.83	3600	–	3600	–

progress for the current operating scenario. In particular, it is expected to increase the number of samples analyzed per season (increase in analytical capacity) and decrease the "dead times" associated with the hours of dedication of human resources (increase in efficiency person-hours). Additionally, make informed decisions regarding incorporating new specific human resources (for example, new analysts and technicians) and equipment (for example, new centrifuge and automation systems) that strengthen the analysis process.

VI. ADDITIONAL EXPERIMENTS WITH THE MODEL

Additional experiments to those done in the previous section are now conducted to study the model's performance. The experiments used the network structure in Fig. 6 which contains the typical components of the problem. However, intending to construct a set of tested instances, we change the data defining the number of machines and the team group. The skills of the team group were also changed. Additionally, the flexibility of machines was randomly defined, and finally, the planning horizon was increased.

We used two instances to know the impact of incremental resources in the model results. In the first instance (I1), two new human resources were added to the basic instance and, in the second instance (I2), additional to the first instance, we added a new machine.

From results in Table 14, we present the average time and average gap for every planning horizon (four weeks, six weeks, and eight weeks). It is important to mention that a limit of one hour (3600 seconds) was set for each run, so if the average running time shown is less than 3600 seconds, it means that there were runs in which the optimal solution was found before reaching the time limit. On the contrary, if the average time is 3600 seconds, the optimal solution was not found in any of the runs performed. Concerning the results, for I1, with only a planning horizon of four weeks was possible to find an optimal solution in some of the runs made. Therefore, in the I1 instance, the model could obtain optimal or near-optimal solutions when a planning horizon until eight weeks is considered. Since the characteristics of the I1 instance could be found in several research labs, at least

in this field, real problems could be well-solved by using the proposed optimization model. However, also in Table 14, note that the results are different when the I2 instance is considered. None of the planning horizons studied was possible to find an optimal solution within the set time. When the planning horizon increases, the GAP of the solutions also found increases, and for eight weeks, none of the runs were able to find a feasible solution. Therefore, the problem with this configuration of machines and workers starts to see a severe difficulty obtaining optimal solutions. Still, for a planning horizon of four weeks, near-optimal solutions can be achieved.

A new interesting scenario of experimentation was also explored. The possibility that the working day is not only of certain blocks per day but also continuous, emulating the possibility of working in shifts. Such working protocol is used in some environments like for shift scheduling in emergency departments in a hospital [27], for electricity generation of a natural gas combined cycle power plant [28], for call centers scheduling [29], and for police patrol scheduling [30]. This situation can be represented by removing Constraint (12) in the optimization model.

In Table 15 the results for the additional experiments removing Constraint (12) in the base model are shown. For the I1 instance, it can be noted that, as in the previous experiments, when two additional human resources are adding to the base instance, an optimal solution could not be found within the time limit established for the planning horizons of six weeks and eight weeks. Only in the planning for four weeks some of the runs found the optimal solution before the time limit. Note also that the average gap increases exponentially when going from six weeks to eight weeks.

In the case of the I2 instance, where one machine is added to the machines in I1, only for four weeks was it possible to find a feasible solution with a high gap of 401%. So, this variant of the problem, considering a continuous working period, is impossible to solve by an exact method even for four weeks.

Overall, with these experiments, we concluded that the optimization model could efficiently manage problem instances with eight machines, six workers, and two weeks as horizon planning. Over this size, increasing workers and machines, or the horizon planning, the running time achieves the time limit of 3,600 seconds, and increasing gaps are achieved. In the case of a continuous operation, when constraint (12) is missing, the complexity of the problem strongly increases. For the I1 instance, only considering four weeks, optimal solutions were obtained. In contrast, for I2 instances, near-optimal solutions could not find.

VII. CONCLUSION

In this work, a new batch process scheduling problem has been defined and mathematically formulated as an optimization problem. The scheduling problem contributes to the literature considering a framework integrating different machines and a team composed of technicians and analysts. Combining the two components of the problem generates complex and original constraints that model atypical operations to batch processing in the optimization model. A case study from the winery Concha y Toro research lab in Chile has been used to apply the mathematical model. Actual data from the lab was used to analyze the model's capacity to find scheduling of human resources and specialized equipment in such a way as to satisfy the lab requirements.

The results obtained after solving the optimization model were satisfactory for the lab team. The model could be run in different periods. Changes in the technical team and instruments are possible so that replanning could also be executed with the same model and only changing the corresponding parameters. The model is flexible enough to evaluate different scenarios with short execution time, allowing the decision-maker to assess the performance of both the machines and the work team.

Although it is possible to solve real scheduling problems with the characteristics studied in this paper, as future work, (meta)heuristic algorithms could be implemented for solving instances of large size. In particular, for this type of problem, approaches using neighborhoods could be appropriated. So, GRASP and VNS metaheuristics would be recommended.

REFERENCES
[1] M. L. Pinedo, *Scheduling: Theory, Algorithms, and Systems*, 5th ed. New York, NY, USA: Springer, 2016.

[2] B. Afshar-Nadjafi, "Multi-skilling in scheduling problems: A review on models, methods and applications," *Comput. Ind. Eng.*, vol. 151, Jan. 2021, Art. no. 107004, doi: 10.1016/j.cie.2020.107004.

[3] L. Mönch, J. W. Fowler, S. Dauzère-Pérès, S. J. Mason, and O. Rose, "Scheduling semiconductor manufacturing operations: Problems, solution techniques, and future challenges," in *Proc. 4th Multidisciplinary Int. Conf. Scheduling, Theory Appl.*, Dublin, Ireland, 2009, pp. 199–201.

[4] E. Kondili, C. C. Pantelides, and R. W. H. Sargent, "A general algorithm for short-term scheduling of batch operations—I. MILP formulation," *Comput. Chem. Eng.*, vol. 17, no. 2, pp. 211–227, Feb. 1993.

[5] C. A. Méndez, J. Cerdá, I. E. Grossmann, I. Harjunkoski, and M. Fahl, "State-of-the-art review of optimization methods for short-term scheduling of batch processes," *Comput. Chem. Eng.*, vol. 30, nos. 6–7, pp. 913–946, May 2006.

[6] G. Gong, Q. Deng, X. Gong, W. Liu, and Q. Ren, "A new double flexible job-shop scheduling problem integrating processing time, green production, and human factor indicators," *J. Clean. Prod.*, vol. 174, pp. 560–576, Feb. 2018, doi: 10.1016/j.jclepro.2017.10.188.

[7] C. Peng, Y. Fang, P. Lou, and J. Yan, "Analysis of double-resource flexible job shop scheduling problem based on genetic algorithm," in *Proc. IEEE 15th Int. Conf. Netw., Sens. Control (ICNSC)*, Mar. 2018, pp. 1–6.

[8] S. Qu, J. Wang, S. Govil, and J. O. Leckie, "Optimized adaptive scheduling of a manufacturing process system with multi-skill workforce and multiple machine types: An ontology-based, multi-agent reinforcement learning approach," *Proc. CIRP*, vol. 57, pp. 55–60, Jan. 2016.

[9] C. N. Potts and M. Y. Kovalyov, "Scheduling with batching: A review," *Eur. J. Oper. Res.*, vol. 120, no. 2, pp. 228–249, 2000.

[10] A. J. Ruiz-Torres, J. H. Ablanedo-Rosas, and L. D. Otero, "Scheduling with multiple tasks per job—The case of quality control laboratories in the pharmaceutical industry," *Int. J. Prod. Res.*, vol. 50, no. 3, pp. 691–705, Feb. 2012.

[11] R. B. Chase and S. Rajagopalan, "A production planning and scheduling system at a chemical testing laboratory," *Int. J. Prod. Econ.*, vol. 29, no. 2, pp. 125–138, Mar. 1993.

[12] S. Hartmann, "Scheduling medical research experiments—An application of project scheduling methods," Christian-Albrechts-Universität zu Kiel, Institut für Betriebswirtschaftslehre, Manuskripte aus den Instituten für Betriebswirtschaftslehre der Universität Kiel, Kiel, Germany, Tech. Rep. 452, 1997. [Online]. Available: https://ideas.repec.org/p/zbw/cauman/452.html

[13] O. P. Mejia, M. C. Anselmet, C. Artigues, and P. Lopez, "A new RCPSP variant for scheduling research activities in a nuclear laboratory," in *Proc. 47th Int. Conf. Comput. Ind. Eng. (CIE)*, Lisbonne, Portugal, 2017, p. 8.

[14] J. C. Boyd and J. Savory, "Genetic algorithm for scheduling of laboratory personnel," *Clin. Chem.*, vol. 47, no. 1, pp. 118–123, Jan. 2001.

[15] N. Shah, C. C. Pantelides, and R. W. H. Sargent, "A general algorithm for short-term scheduling of batch operations—II. Computational issues," *Comput. Chem. Eng.*, vol. 17, no. 2, pp. 229–244, Feb. 1993. [Online]. Available: http://www.sciencedirect.com/science/article/pii/009813549380016G

[16] N. Rakovitis, N. Zhang, and J. Li, "A novel unit-specific event-based formulation for short-term scheduling of multitasking processes in scientific service facilities," *Comput. Chem. Eng.*, vol. 133, Feb. 2020, Art. no. 106626, doi: 10.1016/j.compchemeng.2019.106626.

[17] A. Naber and R. Kolisch, "MIP models for resource-constrained project scheduling with flexible resource profiles," *Eur. J. Oper. Res.*, vol. 239, no. 2, pp. 335–348, Dec. 2014, doi: 10.1016/j.ejor.2014.05.036.

[18] L. Leaven and X. Qu, "A two-stage stochastic programming model for phlebotomist scheduling in hospital laboratories," *Health Syst.*, vol. 7, no. 2, pp. 100–110, May 2018.

[19] A. G. Leeftink, R. J. Boucherie, E. W. Hans, M. A. M. Verdaasdonk, I. M. H. Vliegen, and P. J. van Diest, "Batch scheduling in the histopathology laboratory," *Flexible Services Manuf. J.*, vol. 30, nos. 1–2, pp. 171–197, Jun. 2018.

[20] Washington Wine Industry Foundation. (2004). *Field Guide to Clean Plants & Quarantines for Grapes in Washington State.* [Online]. Available: https://washingtonwinefoundation.org/wp-content/uploads/2016/01/Field-Guide-to-Clean-Plants-and-Quarantines-for-Grapes-in-WA-FINAL-2014.pdf

[21] H. Waite, M. Whitelaw-Weckert, and P. Torley, "Grapevine propagation: Principles and methods for the production of high-quality grapevine planting material," *New Zealand J. Crop Horticultural Sci.*, vol. 43, no. 2, pp. 144–161, Apr. 2015, doi: 10.1080/01140671.2014.978340.

[22] H. Waite, J. Armengol, R. Billones-Baaijens, D. Gramaje, F. Halleen, S. Di Marco, and R. Smart, "A protocol for the management of grapevine rootstock mother vines to reduce latent infections by grapevine trunk pathogens in cuttings," *Phytopathol. Mediterranea*, vol. 57, no. 3, pp. 384–398, 2018.

[23] M. F. Basso, T. V. M. Fajardo, and P. Saldarelli, "Grapevine virus diseases: Economic impact and current advances in viral prospection and management," *Revista Brasileira de Fruticultura*, vol. 39, no. 1, pp. 1–22, 2017.

[24] R. Billones-Baaijens and S. Savocchia, "A review of Botryosphaeriaceae species associated with grapevine trunk diseases in Australia and New Zealand," *Australas. Plant Pathol.*, vol. 48, no. 1, pp. 3–18, Jan. 2019.

[25] H. Waite and P. May, "The effects of hot water treatment, hydration and order of nursery operations on cuttings of *Vitis vinifera* cultivars," *Phytopathologia Mediterranea*, vol. 44, no. 2, pp. 144–152, 2005.

[26] G. P. Martelli. (2014). *Virus Diseases of Grapevine.* American Cancer Society. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470015902.a0000766.pub3

[27] O. EL-Rifai, T. Garaix, V. Augusto, and X. Xie, "A stochastic optimization model for shift scheduling in emergency departments," *Health Care Manage. Sci.*, vol. 18, no. 3, pp. 289–302, Sep. 2015.

[28] E. H. Özder, E. Özcan, and T. Eren, "Staff task-based shift scheduling solution with an ANP and goal programming method in a natural gas combined cycle power plant," *Mathematics*, vol. 7, no. 2, p. 192, 2019. [Online]. Available: https://www.mdpi.com/2227-7390/7/2/192

[29] S. Liao, C. van Delft, and J.-P. Vial, "Distributionally robust workforce scheduling in call centres with uncertain arrival rates," *Optim. Methods Softw.*, vol. 28, no. 3, pp. 501–522, Jun. 2013, doi: 10.1080/10556788.2012.694166.

[30] A. Ingolfsson, M. A. Haque, and A. Umnikov, "Accounting for time-varying queueing effects in workforce scheduling," *Eur. J. Oper. Res.*, vol. 139, no. 3, pp. 585–597, Jun. 2002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0377221701001692

GUIERDY CONCHA received the B.S. degree in bioinformatics engineering and the M.Sc. degree in operations research from the Universidad de Talca, Chile, in 2015 and 2021, respectively. From 2015 to 2019, she worked in research laboratories at the Center for Bioinformatics and Molecular Simulation and the Center for Medical Research, Universidad de Talca, emphasizing molecular biology and biophysics of ion channels, where she has participated in several articles. Since 2019, she has been working with the Logistics Technology Center, Universidad de Talca.

FELIPE GAINZA received the degree in biotechnology engineering from the Universidad Tecnológica de Chile, in 2005, and the Ph.D. degree in plant genetic engineering from the Instituto de Ciencias Biologicas, Universidad de Talca, in 2011. During his Ph.D., he focused on the molecular aspects of Vitis vinifera L. fruit development, including characterizing genes encoding micronutrient transporters and examining how their expression and nutrient availability impacts reproductive development. From 2010 to 2012, he carried out his postdoctorate at the University of California at Davis, which focused on understanding the symbiotic behavior of microorganisms in a wild corn breed under nitrogen deficiency conditions using omic and physiological approaches. From 2012 to 2016, he directed the genetic breeding program for Prunus spp. rootstocks at the Centro de Estudios Avanzados en Fruticultura (CEAF), Chile. He currently works as the R+D+i Leader in Molecular Biology at the Viña Concha y Toro Research and Innovation Center, leading initiatives that aim to improve and strengthen the company's plant material through biotechnological tools that add value and seek applied solutions for the industry.

ALFREDO CANDIA-VEJAR received the B.S. degree in mathematics from the Universidad de Concepción, Chile, and the M.Sc. and Ph.D. degrees in computer science and systems engineering from the Federal University of Rio de Janeiro, Brazil, in 1988 and 1991, respectively. From 1991 to 1997, he was an Associate Professor with the Department of Mathematics, Universidad de Tarapaca. He is currently an Associate Professor with the Department of Industrial Engineering, Universidad de Talca, Chile. From 2001 to 2002, he was the President of the Chilean Institute of Operations Research (ICHIO). He has published several articles in top journals, like the *European Journal of Operational Research*, *Computers & Operations Research*, *Information Sciences*, IEEE Access, and the *International Journal of Production Economics*. He has also served as a supervisor for several theses. From November 2016 to June 2020, he was the Director of the Logistics Technological Center (CETLOG), University of Talca. His current research interests include network optimization, humanitarian logistic, and logistic problems in supply chain management.

● ● ●