Flower power: Finding optimal flower cutting strategies through a combination of optimization and data mining

Han Hoogeveen
Utrecht University
j.a.hoogeveen@uu.nl *

Jakub Tomczyk
University of Sydney
jstomczyk@gmail.com

Tom C. van der Zanden
Utrecht University
T.C.vanderZanden@uu.nl

Abstract

We study a problem that plays an important role in the flower industry: we must determine how many mother plants are required to be able to produce a given demand of cuttings. This sounds like an easy problem, but working with living material (plants) introduces complications that are rarely encountered in optimization problems: the constraints for cutting such that the mother plant remains in shape are not explicitly known.

We have tackled this problem by a combination of data mining and linear programming. We apply data mining to infer constraints that a cutting pattern, stating how many cuttings to harvest in each period, should obey, and we use these constraints in a linear programming formulation that determines the minimum number of mother plants necessary. We then consider the problem of maximizing the total profit given the number of mother plants and show how to solve it through linear programming.

KEYWORDS: data mining, linear programming, cutting patterns, column generation.

Dümmen Orange is a leading company in breeding and development of cut flowers, potted plants, bedding plants and perennials with over a century of experience in the horticultural industry. In addition to a large marketing and sales network, Dümmen Orange has a strong network of production locations. In these production centra so-called mother plants are planted and grown for a large number of varieties. When these mother plants are ready, cuttings are harvested during a period of approximately 16 weeks, after which the mother plants are removed.

[0] In this paper we report on the project carried out for Dümmen Orange in the context of the study group 'Wiskunde met de Industrie' (Math with Industry). Next to the authors of the report, the group consisted of Yella Klemm, Kevin Laros, Jan Nelissen, and Brian Wismans from Dümmen Orange, Norbert Mikolajewski (RU Nijmegen), and Jagna Wiesniewska (VU Amsterdam).

[*] corresponding author

These cuttings are sold to growers, who either place orders beforehand, or place orders during the harvesting. For each variety, the majority of sales takes place in the 'peak weeks', which is a period of approximately 10 weeks; the company has reasonably accurate demand forecasts per week available.

Dümmen Orange experienced the following problem. For each variety, the number of mother plants to be planted is decided on the basis of sales forecasts to which a buffer of 10% is added. When orders come in, contracts are concluded with the growers guaranteeing that the required number of cuttings will be delivered at the desired time. When the harvesting starts, at some point in time the availability of the buffer of 10% is reported to the sales agents, who then try to acquire orders for selling these additional cuttings. Unfortunately, when they are very successful, too many cuttings are required, and the mother plants cannot keep up this pace for too many weeks in a row, which results in a shortage in later periods. This led Dümmen Orange to the question of when to report the availability of the buffer, and possibly to change its size.

Dümmen Orange posed this problem at the study group 'Wiskunde met de Industrie' SWI2016 . In close contact with Dümmen Orange we figured out that we had to address the following research questions:

1. Model how the number of cuttings harvested in previous weeks influences the potential number of cuttings that can be taken from a mother plant in the current and future weeks.

2. Determine how many mother plants should be planted to meet the predicted demand.

3. Determine how many cuttings to offer for sale in each week (and thus how many to cut).

We have looked at this problem for just a single variety of plant in isolation, where we ignored any random disturbances initially. For the variety that we studied, we were provided with the predicted demands and the average number of cuttings per mother plant for each week from 2005 onwards. Unfortunately, detailed information concerning the effects of taking cuttings on the potential mother plants was not available.

This paper is organized as follows: In Sections 1 and 2, we describe how to answer the second and third question by linear programming, for which we need the answer to the first question, which is solved in Section 3 using techniques from data mining. We conclude by providing computational experiments in Section 4 and draw conclusions in Section 5.

[0] see http://www.ru.nl/math/research/vmconferences/swi-2016/

1 Solution approach: linear programming

Since the number of cuttings taken from the mother plants in previous weeks influences the potential yield for the current week in a way that was unknown to us, we decided to work with feasible *cutting patterns*. Here, a cutting pattern describes for each of the 16 weeks the average number of cuttings that are taken from a mother plant; since it is an average (taken over all mother plants), this number can be fractional. For the variety that we studied the typical yield per week was 2 or a little less; as an example a possible cutting pattern could be $\{2.0; 1.8; 1.9; 2.0; \ldots\}$, which indicates that in the first week on average 2.0 cuttings are taken, in the second week 1.8, etc. To be a bit more general, from now on we use T to denote the number of weeks during which we take cuttings. After consulting the experts from Dümmen Orange we found out that the time at which the mother plants were planted made no difference with respect to their potential yield of cuttings, and therefore we do not need to make the cutting patterns depend on the time of planting. Observe the close resemblance between our cutting problem and the standard cutting stock problem (see for example Gilmore and Gomory (1961) and Gilmore and Gomory (1963)). In the cutting stock problem, however, we consider items with different lengths, whereas we now have identical items that are cut in different periods.

Suppose that we know the set of all n possible, feasible cutting patterns. In that case we can solve the problem of determining the required number of mother plants by formulating it as a linear programming problem. We represent cutting pattern j by the parameters a_{jt} that indicate the average number of cuttings harvested in week t ($t = 1, \ldots, T$), when a plant is cut according to pattern j, for $j = 1, \ldots, n$. Define x_j ($j = 1, \ldots, n$) as the number of mother plants that are cut according to cutting pattern j. If we denote the expected demand in period t by b_t ($t = 1, \ldots, T$), then we can formulate the problem of determining the minimum number of mother plants as a linear programming (LP) problem as follows:

$$\begin{gathered} \min x_1 + \ldots + x_n \\ \text{subject to} \\ \textstyle\sum_{j=1}^{n} a_{jt} x_j \geq b_t \quad \forall t \\ x_j \geq 0 \quad \forall j \end{gathered}$$

The solution of this LP program gives you a lower bound on the number of mother plants that have to be planted. Dümmen Orange can decide to add more (for example to have a buffer to guard against disturbances in the production and/or sales process). Note that, although the x_j variables should attain integral values only since these correspond to numbers, it is sufficient to solve the problem by solving the LP-relaxation (where the integrality constraints are relaxed) and round up the outcome values, since the total of the x_j values is big and at most T of them will get a value different from zero (we will see later that we need only one cutting pattern in an optimal solution). Moreover, if the time of planting the mother plants would make a difference with respect to the yield of cuttings, then this can easily be incorporated in this model by making the x_j variables dependent on the time of planting.

Suppose that the management of Dümmen Orange has decided on the number M of mother plants to be planted. We can then solve the problem of determining how many cuttings to offer for sale per week in a similar fashion by formulating it as an LP again. We assume here that we know for each week t how many cuttings we can sell additionally (which we denote by D_t) and the profit p_t that we gain per cutting sold additionally. Next to the decision variables x_j, we introduce decision variables y_t ($t = 1, \ldots, T$) that will indicate the number of additional cuttings to be sold in period t. Just like we did for x_j, we ignore the integrality of the y_t variables. We then get the following LP formulation:

$$\begin{gathered}\max \textstyle\sum_t p_t y_t \\ \text{subject to} \\ \textstyle\sum_{j=1}^{n} a_{jt} x_j - y_t \geq b_t \quad \forall t \\ \textstyle\sum_{j=1}^{n} x_j \leq M \\ 0 \leq y_t \leq D_t \quad \forall t \\ x_j \geq 0 \quad \forall j\end{gathered}$$

If we solve this LP, then we find the cutting strategy that maximizes the total profit given the number of mother plants M. This LP can also be used to find the value of M that maximizes the total profit; we can then simply make M a decision variable, but we have to include the cost of planting M mother plants in the objective function. Furthermore, we can refine the model in case the profit per additional cutting sold is not constant but decreases when more get sold.

2 Generating cutting patterns

In our derivation of the LP problems of the previous section we have assumed that we know all n possible, feasible cutting patterns. Even if we restrict ourselves to a_{jt} values that are multiples of 0.1, there are so many possible cutting patterns that it is neither feasible, nor efficient to generate them all. Fortunately, we can apply the technique of *column generation*, which was invented by Ford and Fulkerson L. R. Ford and Fulkerson (1958) and Gilmore and Gomory Gilmore and Gomory (1961, 1963). Here we solve the LP problem while taking only a small number of feasible cutting patterns into account; we can start with any subset of the cutting patterns, as long as the feasible region is non-empty. After having solved the current LP, we add variables (which correspond to feasible cutting patterns) that will improve the quality of the solution, until we can guarantee that we have found the optimum of the LP for the entire set of feasible cutting patterns.

We will work this out for the first LP, in which we minimize the number of mother plants needed to cover the demand. It is well-known from the theory of linear programming that in case of a minimization problem adding a new variable x_j will improve the quality of the solution only if its *reduced cost* is negative. When we solve the current LP, then we find the non-negative shadow prices; let π_t denote the shadow price corresponding to the constraint that we produce at least b_t cuttings. The reduced cost of a variable x_0 that corresponds to using a given cutting pattern $(a_1, \ldots, a_T)$ is equal to

$$1 - \sum_{t=1}^{T} a_t \pi_t,$$

where the 1 corresponds to the cost coefficient of x_0 in the objective function. Instead of just checking for each feasible cutting pattern $(a_1, \ldots, a_T)$ whether its reduced cost happens to be negative (for which we need to know all feasible cutting patterns), we solve the so-called *pricing problem*, the goal of which is to construct a feasible cutting pattern $(a_1, \ldots, a_T)$ with minimum reduced cost. Note that the values a_t $(t = 1, \ldots, T)$ have become decision variables, and we must choose these such that their combination forms a feasible cutting pattern. To that end, we need a way to describe when a set of values $(a_1, \ldots, a_T)$ constitutes a feasible cutting pattern. Moreover, this knowledge should be cast in such a format that we can use it to solve the pricing problem efficiently. To that end, we infer these constraints by applying techniques from *data mining* to the data on the average number of cuttings harvested per week in the years 2006-2015.

3 Data mining

Data mining is used to retrieve relations from the data. There is a large interaction between data mining and operations research, but it is mainly a one way connection: techniques and algorithms from operations research are applied in data mining Olafsson et al. (2008). We want to apply data mining to *learn constraints* that will be incorporated in the model explicitly, after which we can apply the techniques from operations research. As far as we know, such an approach has not been conducted before. For example, Li and Olafsson Li and Olafsson (2005), who use data mining to derive dispatching rules for a complex production scheduling problem, state that *the idea of this data mining approach to production scheduling is to complement more traditional operations research approaches.*

The domain expert at Dümmen Orange gave several constraints on what constitutes a feasible cutting pattern. For instance, for the variety that we consider one can obtain a maximum of 2.0 cuttings per mother plant in a given week; hence, we find the constraint that $a_t \leq 2.0$ for all $t = 1, \ldots, T$. After having harvested the maximum of 2.0 cuttings in week t, the mother plants have to recover, which can be formulated in the constraint that $a_t + a_{t+1} \leq 3.9$ for all $t = 1, \ldots, T - 1$. Furthermore, a pattern that alternates between cutting near the maximum and not cutting very much (e.g. a pattern such as $\{2.0; 1.4; 2.0; 1.4, \ldots\}$) is not feasible either; it turned out later that we must introduce a constraint of the form $a_t + a_{t+2} + a_{t+4} \leq 5.71$.

It is apparent that the number of constraints required is very large, and a different set of values is needed for every species. Since obtaining these values from domain experts would be very time consuming, we experimented with data mining to automatically derive the constraints. To this end, Dümmen Orange provided us with data specifying the average number of cuttings harvested per mother plant in the period 2006-2015. We scanned the data to identify all constraints of the form $a_{t_1} + a_{t_2} + \ldots + a_{t_k} \leq X$, for all k-tuples $(t_1, \ldots, t_k)$ with $t_1 < t_2 < \ldots < t_k$, where $k \leq 6$ and $t_k - t_1 \leq 10$; here X is set to the maximum value that is observed in the historical data for the left hand side.

Even though the bound of each constraint is set to the maximum value observed, the fact that very many such constraints work together ensures that only realistic cutting patterns will satisfy the constraints. The domain expert confirmed that the cutting patterns that we identified in this way appeared feasible. Note that to obtain more conservative constraints one can take X to be the k-th percentile instead of the maximum of the observed values. However, because our data sets were of limited size taking this approach was not necessary, and would have resulted in overly conservative estimates. However, it could be useful in case a larger training data set is used (which may contain more outliers).

Another possible shortcoming of our data mining model might be that we do not have the data available that we need. We used the data concerning the number of cuttings that were actually harvested instead of the maximum number of cuttings that could have been harvested. Hence, the constraints that were inferred might be too restrictive: it might not consider a certain feasible cutting pattern, simply because this cutting pattern has not been used before. We leave these issues to the experts, who if necessary can perform some experiments to test cutting patterns.

Below, we have listed a small excerpt of the list of constraints that we obtained using data mining.

$$\begin{array}{lcl}
a_t & \leq & 2.0 \\
a_t + a_{t+1} & \leq & 3.9 \\
a_t + a_{t+2} & \leq & 3.85 \\
a_t + a_{t+1} + a_{t+2} & \leq & 5.75 \\
a_t + a_{t+2} + a_{t+4} & \leq & 5.71 \\
a_t + a_{t+1} + a_{t+2} + a_{t+3} & \leq & 7.6 \\
a_t + a_{t+1} + a_{t+3} + a_{t+4} & \leq & 7.61 \\
a_t + a_{t+1} + a_{t+2} + a_{t+3} + a_{t+4} & \leq & 9.46 \\
a_t + a_{t+1} + a_{t+3} + a_{t+4} + a_{t+5} & \leq & 9.21 \\
a_t + a_{t+1} + a_{t+2} + a_{t+3} + a_{t+4} + a_{t+5} & \leq & 11.15 \\
a_t + a_{t+1} + a_{t+3} + a_{t+4} + a_{t+5} + a_{t+6} & \leq & 10.9
\end{array}$$

Note that we have linear constraints only. Hence, the resulting pricing problem of finding the minimum reduced cost, which was equal to

$$1 - \sum_{t=1}^{T} a_t \pi_t,$$

subject to the constraints we identified using data mining is just another linear program, and hence can be solved very efficiently. Since the feasible region described by the constraints is convex, we have that each convex combination of a set of cutting patterns satisfies these constraints, and hence corresponds to a feasible cutting pattern again.

Theorem 3.1. *Let $x^* = (x_1^*, \ldots, x_n^*)$ denote an optimal solution to the linear program of minimizing the number of mother plants. Then there exists an equivalent solution in which we use only one cutting pattern $C = (C_1, \ldots, C_T)$.*

Proof. Define $M = \sum_{j=1}^n x_j^*$. We construct this cutting pattern C by taking the weighted average of all cutting patterns, where we use x_j^*/M as our weight function, for $j = 1, \ldots, n$. Hence, we have that

$$C_t = \sum_{j=1}^{n} a_{jt} x_j^* / M.$$

Since all weights are non-negative and add up to 1, this is a convex combination, and therefore C is a feasible cutting pattern. If we cut all M mother plants according to this cutting pattern, we get the same yield as we get for the optimal solution x^*. □

As a result, we can solve the LP of minimizing the required number of mother plants in a more efficient way using binary search. In each iteration we test whether harvesting b_t cuttings in period t ($t = 1, \ldots, T$) from a given number Q of mother plants corresponds to a feasible cutting pattern. The resulting cutting pattern has $a_t = b_t/Q$ ($t = 1, \ldots, T$), and all that is left is to check whether it satisfies the constraints. If this is the case, then we decrease Q, and if it fails the test, then we increase Q.

In fact, we do not even need binary search. Recall that we have to check whether the values $a_t = b_t/Q$ ($t = 1, \ldots, T$) satisfy the constraints, like $a_t + a_{t+1} \leq 3.9$. This is equivalent to checking whether $Qa_t + Qa_{t+1} = b_t + b_{t+1} \leq 3.9Q$, which implies that Q must be greater than or equal to $(b_t + b_{t+1})/3.9$. For each constraint from data mining we can obtain a lower bound on Q in this way, from which we find that the minimum number of mother plants required is equal to the maximum of these lower bounds.

Note that this approach works only if we can guarantee that a convex combination of a set of cutting patterns is feasible. If we would need additional non-linear constraints to describe a feasible cutting pattern, then we have to resort to column generation again. The pricing problem would then not be solvable as an LP any more, but we could apply an approach such as Constraint Programming.

Now we consider the second LP, in which we optimize the choice of the number of cuttings that must be harvested in period t. The LP-formulation is as follows:

$$\begin{gathered}
\max \textstyle\sum_t p_t y_t \\
\text{subject to} \\
\textstyle\sum_{j=1}^n a_{jt} x_j - y_t \geq b_t \quad \forall t \\
\textstyle\sum_{j=1}^n x_j \leq M \\
0 \leq y_t \leq D_t \quad \forall t \\
x_j \geq 0 \quad \forall j
\end{gathered}$$

We can use Theorem 3.1 again to show that we can use a single cutting pattern C. As a consequence, we can once again solve this problem without generating cutting patterns. We introduce the variables z_t that indicate the number of cuttings that we harvest in period t ($t = 1, \dots, T$); we must have that $z_t \geq b_t$ and we sell the remainder at a price of p_t per cutting. Since we use a single cutting pattern, we cut $a_t = z_t/M$ cuttings per mother plant. Then we can rewrite the LP as

$$\begin{gathered} \max \textstyle\sum_t p_t z_t \\ \text{subject to} \\ b_t \leq z_t \leq D_t \quad \forall t \\ z_t = M a_t \quad \forall t \\ \text{'the variables } a_t \text{ form a feasible cutting pattern'} \end{gathered}$$

Notice that we have to subtract the constant $\sum p_t b_t$ from the objective value to make the outcome values equal. Even in the case where M is a decision variable, the problem can still be reformulated so as to avoid column generation by working with the variables $z_t = M a_t$ only. To that end, we multiply the constraints describing the cutting patterns like $a_t + a_{t+2} \leq 3.85$ with M, such that we obtain the constraint $z_t + z_{t+2} \leq 3,85M$. Obviously, we have to include the cost of growing M mother plants to the objective function. We further remark that our approach can also be used in case we refine the model by offering the possibility of selling up to $b_{t,1}$ cuttings for price $p_{t,1}$, up to $b_{t,2}$ cuttings for price $p_{t,2}$, etc.

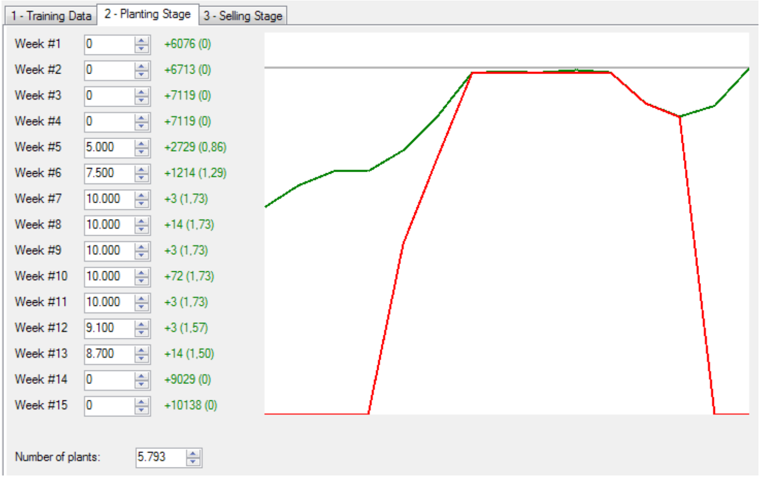
4 Computational experiments

4.1 First approach

To make our mathematical formulation more tangible for the domain experts, we created a graphical user interface around the LP formulation, which allows the user to enter a set of cutting patterns as training data (note that we used the historic data to that avail), and then experiment with various scenarios. The user can specify a number of mother plants and the (predicted) demand levels for each week, and then see whether the demands can be met given this number of mother plants, and how much (if any) additional capacity there is in each week. The software can also calculate the minimum number of mother plants required to meet a specific set of demands.

The red line shows the demands entered by the user, while the green line shows the maximum number of cuttings we could take each week, while still being able to meet the demands. The gray line shows the absolute maximum number of cuttings available in a single week, but note that it is never feasible to take this many cuttings, except for in the last week (when the demand has dropped to zero).

We also implemented an interface for the harvesting stage, which aids in determining how many additional cuttings to offer for sale (on top of the amounts that have already been (pre-)ordered). This is depicted in Figure 2. For each week, the



Week #1	0	+6076 (0)
Week #2	0	+6713 (0)
Week #3	0	+7119 (0)
Week #4	0	+7119 (0)
Week #5	5.000	+2729 (0,86)
Week #6	7.500	+1214 (1,29)
Week #7	10.000	+3 (1,73)
Week #8	10.000	+14 (1,73)
Week #9	10.000	+3 (1,73)
Week #10	10.000	+72 (1,73)
Week #11	10.000	+3 (1,73)
Week #12	9.100	+3 (1,57)
Week #13	8.700	+14 (1,50)
Week #14	0	+9029 (0)
Week #15	0	+10138 (0)

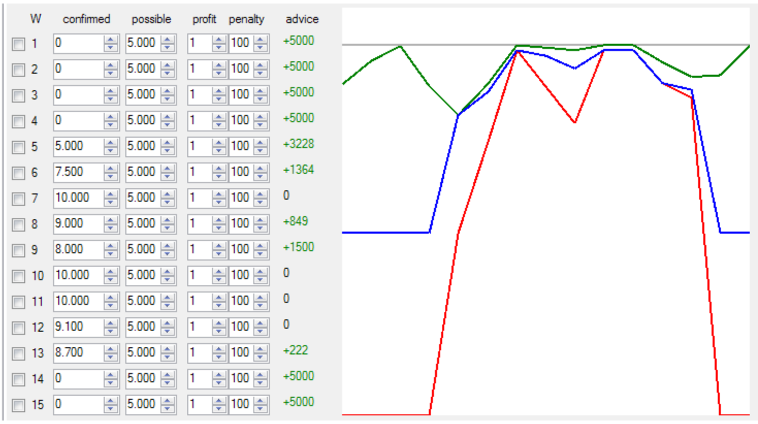
Number of plants: 5.793

Figure 1: GUI for the planting stage.

user can enter how many cuttings have been ordered so far, as well as (an estimate of) the number of cuttings for which there is additional demand. Additionally, the user can enter (for each week) a profit for each additional cutting sold, and a penalty for not delivering cuttings that have already been ordered. Given these values, the program calculates an optimal strategy for selling the additional cuttings.

The red and green lines have the same meaning as before, while the blue line represents our program's advice on selling additional cuttings.

We found that this implementation was a quite powerful tool for conveying our mathematical model to the domain experts.



	W	confirmed	possible	profit	penalty	advice
□	1	0	5.000	1	100	+5000
□	2	0	5.000	1	100	+5000
□	3	0	5.000	1	100	+5000
□	4	0	5.000	1	100	+5000
□	5	5.000	5.000	1	100	+3228
□	6	7.500	5.000	1	100	+1364
□	7	10.000	5.000	1	100	0
□	8	9.000	5.000	1	100	+849
□	9	8.000	5.000	1	100	+1500
□	10	10.000	5.000	1	100	0
□	11	10.000	5.000	1	100	0
□	12	9.100	5.000	1	100	0
□	13	8.700	5.000	1	100	+222
□	14	0	5.000	1	100	+5000
□	15	0	5.000	1	100	+5000

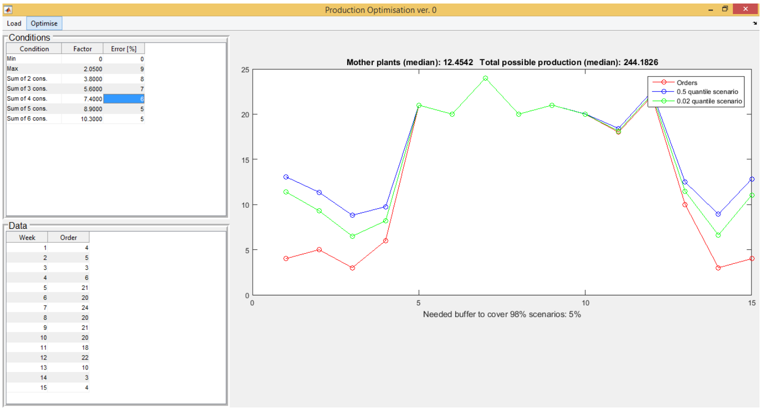
Figure 2: GUI for the selling stage.

4.2 Second approach

A limitation of the model proposed so far is that it does not take into account unpredictable effects, such as the effects of weather or disease. We performed an experiment where the right hand sides of the constraints were perturbed randomly as a potential approach to getting more robust solutions. However, even though it is possible to determine a good relation between environmental conditions (sun, rain, etc.) and the condition of the mother plants, we cannot use this in our computation of the number of mother plants to plant, since the mother plants have to be planted in advance, and it is impossible to give a reasonable prediction of the environmental conditions at the time of harvest. On the other hand, when we get more data, then we could estimate the fluctuations in outcomes. Therefore, we propose an expert-based approach as well.

In the following Matlab based GUI, the user can infer constraints from data mining results or experience and estimate their variability (error). The estimated variability is used to generate scenarios, which are essential to provide confidence intervals. To generate scenarios we sample from a uniform probability distribution, where the range depends on the estimated variability. Intuitively, variability should decrease with number of summed values in constraints. Obviously, we do not take into account rare, but possible events like wars, droughts, volcanic eruptions (which may block deliveries) or plant diseases.

Based on the provided constraints and variability data, possible scenarios are simulated. They are used to calculate the number of mother plants and a buffer (in percentages) needed to cover, for instance 98% of scenarios. See Figure 3.



Condition	Factor	Error [%]
Min	0	0
Max	2.0500	9
Sum of 2 cons.	3.8000	8
Sum of 3 cons.	5.6000	7
Sum of 4 cons.	7.4000	6
Sum of 5 cons.	8.9000	5
Sum of 6 cons.	10.3000	5

Week	Order
1	4
2	5
3	3
4	6
5	21
6	20
7	24
8	20
9	21
10	20
11	18
12	22
13	10
14	3
15	4

Mother plants (median): 12.4542 Total possible production (median): 244.1826

Needed buffer to cover 98% scenarios: 5%

Figure 3: Print screen of Matlab based GUI which is used to estimate number of mother plants and buffer needed to cover 98% of scenarios.

In order to automate a process of estimating the buffer and number of mother plants one has to use history in combination with the advice of experts at least once at the beginning. Due to high number of varieties (thousands) it would be very time consuming. Moreover, estimation of variability from historical data might be challenging. We believe that expert judgement is essential in this approach.

5 Conclusions

The problem of Dümmen Orange is quite different from other applications because of the laws of nature that have to be obeyed: the output is not constant, but decreases over time if you require too much in the beginning. We have attacked this problem by techniques from mathematical programming, where we use techniques from data mining to cover the lack of technical constraints. Especially this latter part seems to be new and very useful for dealing with these kinds of problems. The linear programs are very flexible and easily solvable, which offers great potential for future use by Dümmen Orange, especially when looking at combinations of varieties.

Finally, we want to thank the organizers of the study group 'Wiskunde met de Industrie' for their work, which has resulted in our collaboration with Dümmen Orange.

References

P. C. Gilmore and R. E. Gomory. A linear programming approach to the cutting-stock problem. *Operations Research*, 9(6):849–859, 1961. doi: 10.1287/opre.9.6.849. URL http://dx.doi.org/10.1287/opre.9.6.849.

P. C. Gilmore and R. E. Gomory. A linear programming approach to the cutting stock problem: Part ii. *Operations Research*, 11(6):863–888, 1963. doi: 10.1287/opre.11.6.863. URL http://dx.doi.org/10.1287/opre.11.6.863.

J. L. R. Ford and D. R. Fulkerson. A suggested computation for maximal multi-commodity network flows. *Management Science*, 5(1):97–101, 1958. doi: 10.1287/mnsc.5.1.97. URL http://dx.doi.org/10.1287/mnsc.5.1.97.

X. Li and S. Olafsson. Discovering dispatching rules using data mining. *Journal of Scheduling*, 8(6):515–527, 2005. ISSN 1099-1425. doi: 10.1007/s10951-005-4781-0. URL http://dx.doi.org/10.1007/s10951-005-4781-0.

S. Olafsson, X. Li, and S. Wu. Operations research and data mining. *European Journal of Operational Research*, 187(3):1429 – 1448, 2008. ISSN 0377-2217. doi: http://dx.doi.org/10.1016/j.ejor.2006.09.023. URL http://www.sciencedirect.com/science/article/pii/S037722170600854X.