

Flower power

Citation for published version (APA):

Hoogeveen, H., Tomczyk, J., & van der Zanden, T. C. (2019). Flower power: Finding optimal plant cutting strategies through a combination of optimization and data mining. *Computers & Industrial Engineering*, 127, 39-44. <https://doi.org/10.1016/j.cie.2018.11.047>

Document status and date:

Published: 01/01/2019

DOI:

[10.1016/j.cie.2018.11.047](https://doi.org/10.1016/j.cie.2018.11.047)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

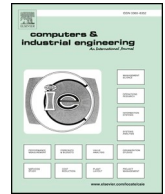
www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.



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

Han Hoogeveen^{a,*}, Jakub Tomczyk^b, Tom C. van der Zanden^a

^a Utrecht University, Netherlands

^b University of Sydney, Australia

ARTICLE INFO

Keywords:

Data mining
Linear programming
Cutting patterns
Column generation

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 per week. This sounds like an easy problem, but working with living material (plants) introduces complications that are rarely encountered in optimization problems: there is no list with possible cutting patterns, describing the average number of cuttings taken from a mother plant per week. More importantly, there is no easy way to find out whether a cutting pattern is feasible, that is, whether the mother plants can keep up delivering the number of cuttings required by the cutting pattern each week: the only alternative to asking for an ‘expert’s opinion’ is to apply a field-test, which takes a lot of time (and there are very many options to check).

We have tackled this problem by a combination of data mining and linear programming. We apply data mining to infer constraints that a feasible cutting pattern should obey, and we use these constraints in a linear programming formulation to determine the minimum number of mother plants that are needed to supply the demand. Due to the linearity of the constraints obtained by data mining, this formulation can be reformulated such that it becomes trivially solvable. Next, we look at the problem of finding the optimal number of mother plants for the case that we can sell a given number of the remaining cuttings on the market for a given price; we show that this problem can be solved efficiently through linear programming.

1. Introduction

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 centers 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. These cuttings are sold to growers, who either place orders beforehand, or place orders during the harvesting weeks. 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 determined on 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 per week, and the mother plants cannot keep up this pace for too many weeks in a row, which results in a shortage in later weeks. 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’ SWI2016.¹ 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

* Corresponding author.

E-mail addresses: j.a.hoogeveen@uu.nl (H. Hoogeveen), J.Tomczyk@maths.usyd.edu.au (J. Tomczyk), T.C.vanderZanden@uu.nl (T.C. van der Zanden).

¹ See < <http://www.ru.nl/math/research/vmconferences/swi-2016/> >

predicted demand.

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

We have looked at these problems for just a single variety of plant in isolation, which is reasonable, since the production processes of the different varieties do not influence each other. Because of a lack of data, we ignored any random disturbances, but these can easily be included in a later stadium, when more data or additional knowledge becomes available. For the variety that we studied, we were provided with the predicted demands and the average number of cuttings harvested per mother plant per week for the years 2006–2015 (160 numbers in total); this was all the information that was available for this problem within the company.

This paper is organized as follows. In Section 2 we analyze the problem. In Section 3 we show how the feasibility constraints can be determined using data mining, and in Section 4 we describe how we can use these feasibility constraints to find the minimum number of mother plants that are required to supply the predicted demand. In Section 5 we present a linear program to find the optimal number of mother plants and the best cutting pattern, given that we know how many cuttings can be sold additionally and at which price. We conclude by providing computational experiments in Section 6 and draw conclusions in Section 7.

Our contribution. We present a simple form of data mining to generate the yet unknown constraints that are needed to verify the feasibility of a cutting pattern; as far as we know, our paper is the first one in which constraints are inferred using data mining. We further show that due to the linearity of these constraints we only need to apply one cutting pattern, which leads to a trivial way to determine the minimum number of mother plants that have to be planted.

2. Finding constraints needed for feasibility of cutting patterns

Since the number of cuttings taken from the mother plants in previous weeks influences the potential yield for the current week in some unknown way, we decided to work with feasible *cutting patterns*. Here, a cutting pattern describes for each one 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;...}, 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 Dümme Orange we found out that the exact time at which the mother plants were planted within the planting interval made no difference with respect to their potential yield of cuttings, and therefore we do not need to let the cutting patterns depend on the time of planting.

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

$$\begin{aligned} \min M &= \sum_{j=1}^n x_j \\ \text{subject to} \\ \sum_{j=1}^n a_{jt} x_j &\geq b_t \quad \forall t = 1, \dots, T \\ x_j &\geq 0 \quad \forall j = 1, \dots, n \end{aligned}$$

The solution of this LP program yields the minimum number M of mother plants that have to be planted. Dümme 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 variables x_j 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.

The LP formulation above is identical to the LP formulation for the standard cutting stock problem (see for example [Gilmore & Gomory, 1961, 1963](#)), but the cutting patterns correspond to different situations. In the cutting stock problem we consider items with different lengths, whereas we now have identical items that are cut in different weeks. More importantly, in the cutting stock problem, it is trivial to describe the constraints that a cutting pattern must satisfy to be feasible. This makes the cutting stock problem the perfect example for applying the technique of *column generation*, which was invented by [Ford and Fulkerson \(1958\)](#) and [Gilmore and Gomory \(1961, 1963\)](#).

In our problem on the other hand, the domain expert can decide for a given cutting pattern whether it is presumably feasible, but the characteristics of a feasible cutting pattern are unknown, and hence there is no explicit list with constraints that a set of values (a_1, \dots, a_T) must satisfy such that it constitutes a feasible cutting pattern. We infer these constraints by looking at the data, which give the average yield per mother plant 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, Li, & Wu, 2008](#)). We want to apply data mining to *find 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 \(2005\)](#), 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 Dümme Orange gave several indications 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 $t = 1, \dots, T$. After having harvested the maximum of 2.0 cuttings in week t , the mother plants have to recover, which can be formulated by the constraint $a_t + a_{t+1} \leq 3.9$ for all $t = 1, \dots, T - 1$. Furthermore, a pattern that alternates between cutting the maximum amount possible and cutting a much smaller number (e.g. a pattern such as {2.0;1.4;2.0;1.4, ...}) is not feasible either; it turned out later that any cutting pattern must satisfy the constraint to be feasible.

It is apparent that the number of constraints required may be large, and a different set of values is needed for each variety. Obtaining these

values from domain experts would be very time consuming (if possible at all), and hence we need to define a process to automate it. Thereto, Dümme Orange provided us with data specifying the average number of cuttings harvested per mother plant per week in the years 2006–2015. We applied a very primitive form of data mining to derive the constraints; recall that we have no more than $10T$ numbers in our excel file, in contrast to the gigabytes of data that are standard in data mining. Together with the domain expert, we decided that the presumably relevant constraints will be of the form $a_{t_1} + a_{t_2} + \dots + a_{t_k} \leq X$, where $t_1, \dots, t_k \in \{1, \dots, T\}$ and $t_1 < t_2 < \dots < t_k$. To find these constraints, we simply enumerated all possibilities with $k = 1, \dots, 6$ and $t_k - t_1 \leq 10$ and computed for each constraint the right hand side value X as the maximum value observed in the historical data for the expression on 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. Instead of taking the maximum it is also possible to use a bit more conservative constraints by putting X equal to the k -th percentile instead of the maximum of the observed values, especially if in the future a larger training data set may become available.

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 (which data are not available). Hence, the constraints that were inferred might be too restrictive: we may exclude 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; these constraints must hold for all possible t values. Since the majority of cuttings must be harvested during the peak weeks, it is possible to reduce the domain of t ; in this case approximately 1000 constraints of the several thousands of constraints remain.

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

4. Determining the minimum number of mother plants

Now that we have a description of the feasibility constraints, we can solve the linear programming formulation to decide on the number of mother plants that we need. Since the number of feasible cutting patterns is much too large to enumerate, it seems inevitable to use column generation, just like for the cutting stock problem. If we solve the LP for a limited number of cutting patterns, then we find a shadow price π_t ($t = 1, \dots, T$) for each week constraint. Hence, the reduced cost of a cutting pattern described by (a_1, \dots, a_T) is equal to

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

and this must be minimized subject to the feasibility constraints. Since these feasibility constraints are linear, the resulting pricing problem is a linear programming problem, and hence can be solved very efficiently. But it turns out that we do not need the technique of column generation at all. 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. This observation leads to the following result.

Theorem 4.1. *Let $x^* = (x_1^*, \dots, x_n^*)$ denote an optimal solution to the linear program of minimizing the required number of mother plants. Then there exists an equivalent solution in which we use only one cutting pattern $C = (C_1, \dots, C_T)$.*

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 $j = 1, \dots, n$. 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, then we get the same yield as we get for the optimal solution x^* . \square

Since we want to produce b_t cuttings in week t ($t = 1, \dots, T$), it is optimal to use the cutting pattern $C = (C_1, \dots, C_T)$, where $C_t = b_t/M$, and we determine M as the smallest number such that $(b_1/M, \dots, b_T/M)$ satisfies all constraints identified using data mining. Our first idea was to determine this value M using binary search, but there is an even simpler method. Recall that we have to check whether the values $a_t = b_t/M$ ($t = 1, \dots, T$) satisfy the constraints, such as $a_t + a_{t+1} \leq 3.9$. This is equivalent to checking whether $Ma_t + Ma_{t+1} = b_t + b_{t+1} \leq 3.9M$, which implies that M must be greater than or equal to $(b_t + b_{t+1})/3.9$. For each one of the feasibility constraints we can obtain a lower bound on M in this way; 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 instead.

5. Determining how many cuttings to offer for sale

The solution approach yields a lower bound M on the number of mother plants that must be planted, but even if exactly M mother plants are planted, then in a number of weeks more cuttings will be available than ordered. The exact surplus per week depends on the applied cutting pattern, and we can optimize this as well, given that we know for each week t ($t = 1, \dots, T$) 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. Suppose that the management of Dümme Orange has decided on the number Q of mother plants to be planted; this can be any given value, as long as $Q \geq M$. We can then solve the problem of determining how many cuttings extra to offer for sale per week in a similar fashion by formulating it as an LP again. Next to the decision variables x_j , we introduce decision variables y_t ($t = 1, \dots, T$) that will indicate the number of additional cuttings to be sold in week t . Just like we did for x_j , we ignore the integrality of the y_t variables. We then get the following LP formulation:

$$\begin{aligned}
& \max \sum_t p_t y_t \\
& \text{subject to} \\
& \sum_{j=1}^n a_{jt} x_j - y_t \geq b_t \quad \forall t = 1, \dots, T \\
& \sum_{j=1}^n x_j \leq Q \\
& 0 \leq y_t \leq D_t \quad \forall t = 1, \dots, T \\
& x_j \geq 0 \quad \forall j = 1, \dots, n
\end{aligned}$$

We can use [Theorem 4.1](#) 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 week t ($t = 1, \dots, T$); we must have that $z_t \geq b_t$ and we sell the remainder $z_t - b_t$ at a price of p_t per cutting. Since we use a single cutting pattern, we cut $a_t = z_t/Q$ cuttings per mother plant. Then we can rewrite the LP as

$$\begin{aligned}
& \max \sum_t p_t (z_t - b_t) \\
& \text{subject to} \\
& b_t \leq z_t \leq b_t + D_t \quad \forall t \\
& z_t = Ma_t \quad \forall t
\end{aligned}$$

‘the variables a_t form a feasible cutting pattern’

If Q is a decision variable as well, then the constraints $z_t = Qa_t$ are not linear, but we can rewrite the above formulation to obtain a linear program again by eliminating the a_t variables. To that end, we multiply the constraints that ensure the feasibility of the cutting patterns, for example $a_t + a_{t+2} \leq 3.85$, by Q , such that we obtain the constraint $z_t + z_{t+2} \leq 3.85Q$. Obviously, we have to include the cost of growing Q mother plants in 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.

6. Computational experiments

To make our mathematical formulation more tangible for the domain experts, we have created a graphical user interface around the LP formulation, which allows the user to enter a set of training data (note that we used the historic data to that avail), and then experiment with

various scenarios (see [Fig. 1](#)). 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. In the screenshot the required number of cuttings (for example 5000 for Week 5) have been filled in; which leads to the red line; then the minimum number of mother plants is computed (resulting in the 5793 at the bottom). The numbers in green show the remaining number of cuttings and, in between brackets, the average yield in the corresponding weeks.

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 cuttings that have already been ordered). This is depicted in [Fig. 2](#). For each week, the user can enter how many cuttings have been ordered so far, as well as (an estimate of) the number of cuttings that can be sold additionally. Furthermore, 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 powerful tool for conveying the usefulness of our mathematical model to the domain experts.

6.1. Modeling disturbances

In the model described above we do not take any disturbances into account, due to for example the weather conditions; we are aware of this, but there are simply no data available that can be used in such a model. 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



Fig. 1. GUI for the planting stage.



Fig. 2. GUI for the selling stage.

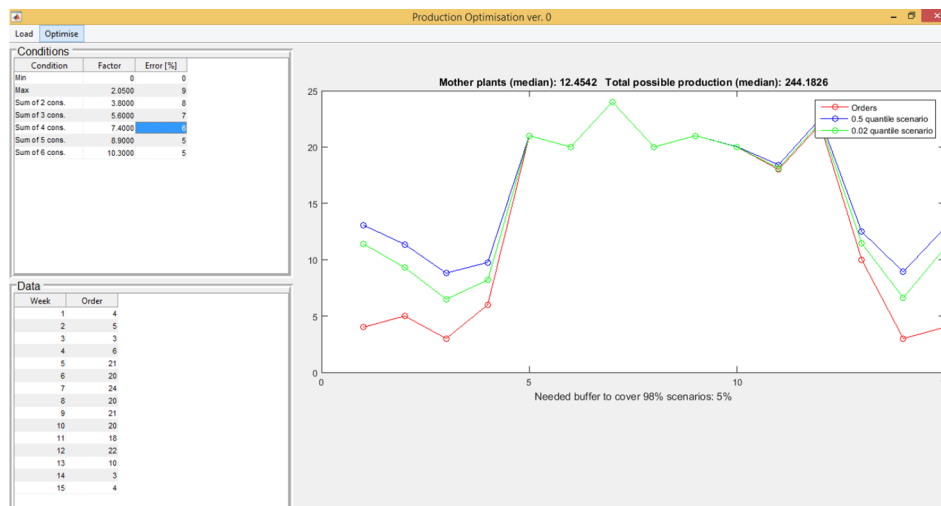


Fig. 3. Print screen of Matlab based GUI which is used to estimate number of mother plants and buffer needed to cover 98% of scenarios.

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.

We designed a Matlab-based GUI in which 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 the 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 Fig. 3.

In order to automate a process of estimating buffer and number of mother plants one has to use historic data and experience of experts at least once at the beginning. This is easily done for one variety, but practically impossible for all of the thousands of varieties. Moreover, estimation of variability from historical data might be challenging. We believe that expert judgement is essential in this approach.

7. Conclusions

The problem of Dümme 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 linear programming, where we use a simple form of data mining to cover the lack of constraints describing the feasibility of the cutting patterns. Especially this latter part seems to be new and very useful for dealing with this kind of problems. The linear programs are very flexible and easily solvable, which offers great potential for future use by Dümme Orange.

References

- Ford, J. L. R., & Fulkerson, D. R. (1958). A suggested computation for maximal multi-commodity network flows. *Management Science*, 5, 97–101. <https://doi.org/10.1287/mnsc.5.1.97>.
- Gilmore, P. C., & Gomory, R. E. (1961). A linear programming approach to the cutting-stock problem. *Operations Research*, 9, 849–859. <https://doi.org/10.1287/opre.9.6.849>.
- Gilmore, P. C., & Gomory, R. E. (1963). A linear programming approach to the cutting stock problem: Part ii. *Operations Research*, 11, 863–888. <https://doi.org/10.1287/opre.11.6.863>.
- Li, X., & Olafsson, S. (2005). Discovering dispatching rules using data mining. *Journal of Scheduling*, 8, 515–527. <https://doi.org/10.1007/s10951-005-4781-0> ISSN 1099-1425.
- Olafsson, S., Li, X., & Wu, S. (2008). Operations research and data mining. *European Journal of Operational Research*, 187, 1429–1448. <https://doi.org/10.1016/j.ejor.2006.09.023>. ISSN 0377-2217 <<http://www.sciencedirect.com/science/article/pii/S037722170600854X>>.