Agrifood Campaign Planning

Problem presented by

Martin Robinson

Transfaction

Executive Summary

The challenge was to find ways for the players in an agricultural food supply chain to interact in ways that enable the chain to operate more efficiently. What information do they need to exchange, and what incentives need to be in place between them? What software would help the information exchange and responsive actions to take place? The problem was thought about with the UK sugar beet industry as the working example, but similar considerations, with many differences of detail, are expected to apply in other contexts.

The presentation of this problem is supported by NERC Business Policy and Innovation Funding and the PURE Network.

Version 1.1
May 24, 2014
iii+26 pages

Report author

David Allwright (Smith Institute)
Federico Botta (University of Warwick)
Andrew Gibbs (University of Reading)
Karina Piwarska (Centre for Industrial Applications of Mathematics and Systems Engineering, Warsaw)

Contributors

Andreas Asheim (University of Cambridge)
Hendrik Brackmann (University of Oxford)
Elizabeth Buckingham-Jeffery (University of Warwick)
Jeff Dewynne (University of Oxford)
Samuel Groth (University of Reading)
Sam Howison (University of Oxford)
Kamil Kulesza (Centre for Industrial Applications of Mathematics and Systems Engineering, Warsaw)
Dario Paccagnan (Technical University of Denmark)
Agnieszka Radziwon (University of Southern Denmark)
Gonçalo Simões (University of Oxford)
Chang Wang (University of Oxford)
Christoph Winkler (Austrian Academy of Sciences)
Piotr Ziolo (Centre for Industrial Applications of Mathematics and Systems Engineering, Warsaw)

ESGI100 was jointly hosted by
Smith Institute for Industrial Mathematics and System Engineering
The University of Oxford

with additional financial support from
Engineering and Physical Sciences Research Council
European Journal of Applied Mathematics
Oxford Centre for Collaborative Applied Mathematics
Warwick Complexity Centre

Contents

1 Introduction

The UK sugar beet industry involves numerous organisations, companies and individuals operating from different locations and with different objectives and incentives. The aim of the Study Group was to understand the process well enough to model it and to propose ways in which these organisations companies and individuals (collectively described as players) could exchange information and interact, in ways that would result in greater efficiency, and to the benefit of everyone involved in the process. Much of what was done relates specifically to sugar beet, but there are other crops where similar issues arise — though of course each crop will have its own idiosyncrasies.

1.1 Background and scope

(1.1.1) We aim to describe here the way the UK sugar beet industry operates at present. It is presented diagrammatically in Figure 1. In subsection 1.2 we shall describe the perceived inefficiencies in the present system, and in subsection 1.3 the elements of the system that are most easily changeable, and so could be used as controls or incentives.

(1.1.2) Sugar beet is grown mainly in eastern England, and there are perhaps 3000 growers, with widely differing acreages. The beet can be lifted from the fields (*i.e.* harvested) from mid-September onwards. The beet cannot be lifted if the soil is too wet, or if the soil is baked too hard. The beet should be lifted before the first frost. When the beet is lifted it is stored

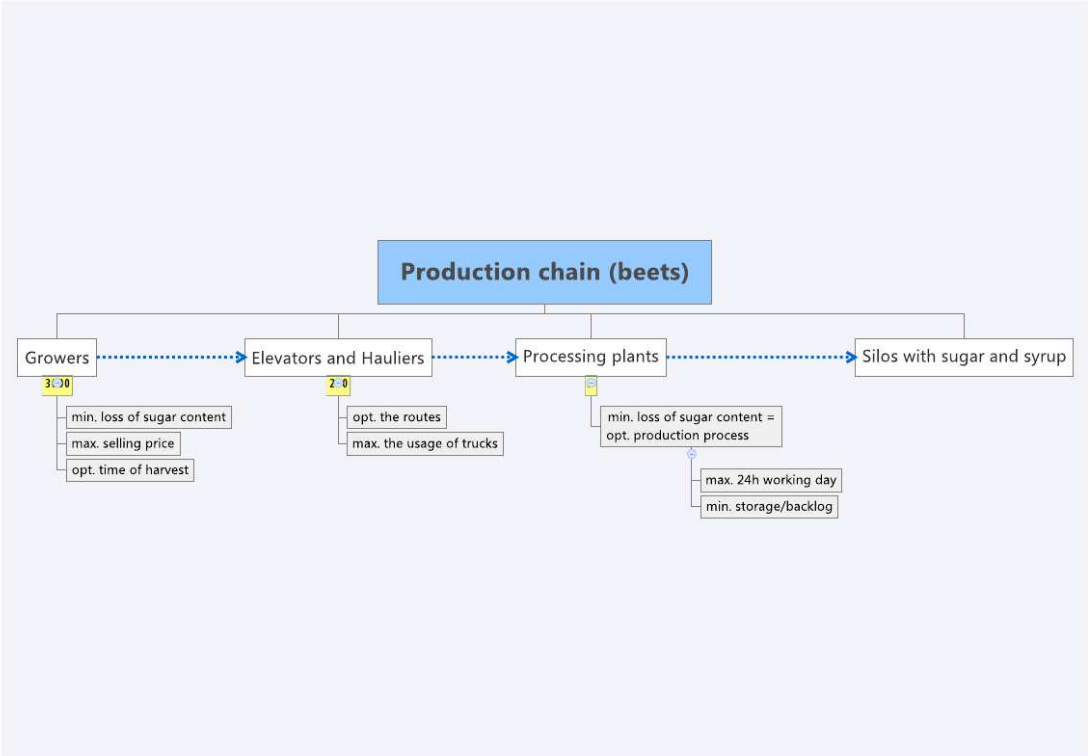


Figure 1: Schematic of the sugar beet production chain.

on a concrete pad on the farm. If there is risk of frost, the beet will need to be stored covered.

(1.1.3) From the pads on the farms, the beet is taken to a processing plant in 30-tonne trucks operated by hauliers. The distance of a grower from the processing plant can be up to 50 miles, with an average of 28. There are about 100 haulier companies, including large firms like DHL, small local haulage companies, and some trucks owned by individual growers. So some have just 1 truck and some have hundreds. The truck drivers can work 9 hours a day, or 10 hours on at most 2 days a week. A large haulier will have a number of drivers so its trucks may be able to operate for more hours a day than a truck belonging to a small company or an individual grower.

(1.1.4) The 5 processing plants in the UK are operated by British Sugar. The beet is unloaded from the trucks onto a large concrete pad (at the plant in Bury St. Edmunds, the pad is 150 m by 40 m and the stack can be up to 6 m high). The plant can process about 800 truckloads a day, and the storage pad can hold about 1400 truckloads. The processing capacity of the plant varies by about ±10% from day to day for various unpredictable reasons. From the stack, the beet is pushed by bulldozers into a water channel that sweeps it along to the processing plant itself. It is processed into syrup and then dry sugar. The beet is sampled to assess its sugar content when it arrives at the plant.

(1.1.5) There used to be more processing plants, but the reduction to 5 has not been accompanied by a proportional increase in capacity so the processing season has been extended, and now runs to February. The whole season, from mid-September when lifting starts, through to February when all the beet has been processed, is called the Campaign. The processing plants operate 24 hours a day during the campaign.

(1.1.6) At present the initial planning of the Campaign is undertaken by British Sugar, and is at the 1-week granularity. So they plan that the beet of certain growers will be lifted in particular weeks, and be brought to a particular plant in particular weeks. They also offer a centralized haulage plan to the growers: in the centralized plan, a grower contracts to provide a certain tonnage of beet at his farm to be ready in a certain week. There are 20 hauliers who supply services to British Sugar as part of this centrally-organised system. Alternatively, a grower can choose to arrange his own haulage, in which case British Sugar pay him a certain allowance per ton-mile for the transport, based on the shortest road distance from his farm to the processing plant. He then uses his own truck or makes his own arrangements with a haulier or another grower who has a truck. His contract then is to provide a certain tonnage at the plant in the specified week.

(1.1.7) The payment from British Sugar to the grower is based on the sugar content of his crop. When it arrives at the plant, a sample is taken for analysis and the weight of sugar per weight of beet is assessed. The sugar content varies depending on the beet variety, the soil, and the weather conditions during the growing season — sunny days and rain at night are the best. After the end of the growing season (mid-September) the sugar content of the beet in the ground is constant. The sugar content can vary between 15% and 21%. When it is lifted and is waiting on the pad at the farm, sugar content is lost, at a rate of about 0.1% per day. When beet is pushed around by the bulldozers at the processing plant, sugar content is also lost — anything that damages the beet loses sugar content. The rate paid by British Sugar to the growers rises steadily during the Campaign period and is about 15% greater in February than it was as the start of the campaign in mid-September.

1.2 Inefficiencies

(1.2.1) The perceived inefficiencies in the present system are illustrated diagrammatically in Figure 2.

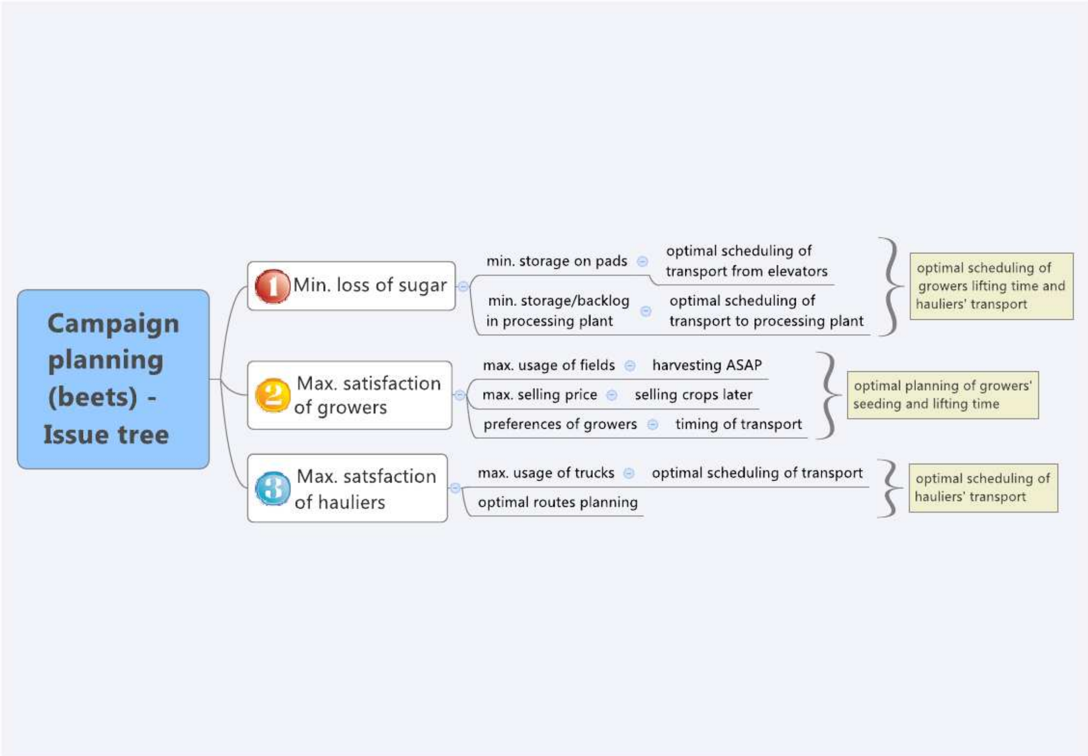


Figure 2: Perceived inefficiencies in the present process.

(1.2.2) At the processing plant, one of the inefficiencies that can arise is if the beet backlog builds up too much — *i.e.* the amount of beet in the stack awaiting processing. The sugar loss from pushing this beet around with the bulldozers is kept smallest if this backlog is kept small.

(1.2.3) For the growers, one of the inefficiencies is the loss of sugar while their beet is waiting on the pad at their farm: the grower wants the interval between lifting the beet and processing it to be small.

(1.2.4) For the hauliers, one of the inefficiencies is the journeys they make with an empty truck at the beginning of a day to their first farm, and at the end of the day from the plant back to the haulage company.

1.3 Possible changes

(1.3.1) One of the possible changes that could be implemented in the system would be to alter the price paid by the processing plant to the growers.

1.4 Literature

(1.4.1) A study of transport efficiency in the sugar beet industry was prepared in 2009 [1]. A study of coupled supply planning and logistics with reference to the sugar cane industry in South Africa is published as [2]. This has quite similar aims to our project.

(1.4.2) We find some very helpful literature, studying sugar-cane industry in Australia and South Africa. They study various optimisation problems arising from different aspects of the industry and implement some simulation tools. Although the majority of the previous study only focus on a single aspect and do not integrate different factors into a single framework, some analysis are well worthy of mentioning: on a short term horizon, [7] studies the the optimisation of harvest schedules, accounting for the geographical and temporal differences in sugar yield; [9] investigates the coordination between transportation and harvest. On a mid-term horizon, works like [8] study sugar production maximization in the context of yearly planning.

(1.4.3) On the other hand, a study of coupled supply planning and logistics with reference to the sugar cane industry in South Africa is published [2]. This has quite similar aims to our project. Their study examines multiple-level planning and adopts a two step simulation to integrate seasonal planning with the short-term logistic. It introduces two simulation tools, MAGI for seasonal supply planning and ARENA for daily supply, to investigate the effects of various factors that could potentially impact the campaign, including harvesting mechanism, vehicles, milling season, and sensitivity to risk. Based on the simulation results their study discusses outcomes under different scenarios, which can facilitate negotiations between different parties.

(1.4.4) Another study, of the sugar cane industry in South America, is in [3] and uses discrete event simulation.

[Christoph, if you write a summary of that it could go here.]

(1.4.5) Some of the possibly-relevant mathematical literature includes that on games with exhaustible resources, for instance the work of Tom Hosking [4]. This could perhaps be developed with say 2 growers, one close to the plant and one far away, so they have different transport costs, and with the plant as another player, having the aim of keeping a steady inflow of beet.

2 Strategy

In this report we first consider various elements of the problem in some isolation, and then consider the issues in putting them together. We first describe the economic models considered, then stochastic models that study the effects of the uncertainties in the system, then some scheduling models that are intended to achieve some of the potential efficiencies better than the current process.

3 Economic models

3.1 Pricing models

(3.1.1) If the price paid by the processing plant to the growers can be chosen in a way that makes the growers neutral between different times for lifting their beet, then that should enable any possible efficiencies in the transport process to be taken advantage of more easily.

(3.1.2) One of the ingredients in modelling this is that it is beneficial to the grower to have his beet lifted early, since he can then reuse that field, preparing it for whatever its next crop is to be. There is therefore a utility function to the grower of lifting the beet at time t, and it is a decreasing function of t. All the beet needs to be lifted by the time of the first frost, so a simple form of the utility function would be

$$U(t) = U_0 \max(1 - t/T_{\text{frost}}, 0), \tag{1}$$

for a suitable constant U_0 and with t measured from the start of the campaign period.

(3.1.3) One way of incorporating this insight into a simplified economic model is as follows. In this model, the growers are aggregated together, and also the beet awaiting processing is aggregated together. Also the model as written here is deterministic, and would need modifications to allow for stochastic effects.

(3.1.4) For the growers, we let $q(t)$ denote their combined production rate, and $p(t)$ be the price paid by the plant to the grower. Then the payoff to the grower is modelled as

$$\int_0^T p(t)q(t)\exp(-rt) - cq(t)\,dt. \tag{2}$$

In this, the discount factor $\exp(-rt)$ is representing the fact that the grower prefers to have his beet collected early. The constant c represents the growers' cost per tonne. The amount of beet initially is some x_0, the total crop, and the lifting process is represented by $dx/dt = -q(t)$ with the constraints that x and q must not go negative.

(3.1.5) The amount of beet in storage is denoted by $Q(t)$ so $dQ/dt = q - \bar{q}$, where $\bar{q}$ is the rate at which beet is taken from the storage to the factory. Naturally, Q and $\bar{q}$ must also not go negative.

(3.1.6) The payoff function to the plant is modelled in the form

$$\int_0^T -p(t)q(t) - f(Q(t)) + \bar{q}(t)P\,dt. \tag{3}$$

Here the first term is the price the plant is paying to the growers. The second models the cost represented by the amount of beet in storage somewhere in the system, so f is an increasing function of Q. The third term represents the gain for producing sugar, so P is the current price for sugar.

(3.1.7) Solving this model as a game for the growers and plant then consists in the growers choosing $q(t)$ to maximize their payoff, and the plant choosing $p(t)$ and $\bar{q}(t)$ to maximize its payoff.

(3.1.8) For the growers, the solution is by introducing their value function from any point, $V(x(t), t)$, and then the result is that they choose $q(t)$ to maximize

$$\left(p(t) - c - \frac{\partial V}{\partial x}\right) q(t). \tag{4}$$

(3.1.9) For the plant, the optimal point over $\bar{q}$

3.2 Cooperative games

(3.2.1) A cooperative game is one where the cooperation of the players in a coalition generates surplus value. There is then the theory of Shapley value that determines a fair way to distribute that value to the partners in a coalition. Each player receives a value that is the average gain in value that adding him brings, if the coalition is formed sequentially in a random order.

(3.2.2) The Shapley-Gale algorithm is a matching algorithm where the participants have expressed preferences among the possible options available to them. It could potentially be used to implement the matching of growers to time-slots in the hauliers schedule.

(3.2.3) There are 3 elements to the problem,

(a) maximizing the surplus that the process generates (so this involves minimizing sugar loss on the pads in the fields, minimizing transport costs, and minimizing sugar loss at the processing plant); also the surplus is a random variable, so some scalar function has to be chosen, (*e.g.* the mean, the median, the probability of it exceeding some threshold);

(b) allocating the actual surplus to the participants fairly;

(c) transparency — assuring the participants that the process is treating them fairly.

(3.2.4) If objective 1 can be solved, then it results in a surplus that could be allocated among the participants. This surplus then needs to be shared among the participants in a way that is generally perceived to be fair.

(3.2.5) It was suggested that a certain proportion of the payments should be transferred directly from the plant to the growers and from the growers to the hauliers, as at present, with a certain amount kept back. This retained portion of the collective surplus could then be distributed at the end of the campaign period, in a way that reflects each player's contribution to achieving the maximum possible surplus, or penalizes their contribution to failing to achieve the maximum surplus.

4 Stochastic models

4.1 Weather correlation

(4.1.1) If we plan to collect the beet in a certain order, and we wish to keep the supply of beet to the processing plant robust to the effects of weather in delaying the delivery schedule, then it is natural to expect that we should collect from separated areas at the same time: if the plan were to involve collecting from growers in the same area at the same time, then it is not robust to bad weather in that area.

4.2 Summary

(4.2.1) The delivery date of different growers is influenced by the weather. Assuming that the weather hits all growers in a certain region simultaneously and similarly, we analyse how one should sort the growers to lower production's fluctuation. To do so, we analyse a stylised model in which growers are divided into two regions, each with a local weather component. The objective is to optimally mix the growers of these two regions such that the expected excess harvest is minimized. We start by describing the model in detail, then we describe the simulation and we finally conclude.

4.3 Description of Model

(4.3.1) We assume that growers are split into two regions that will be processed over two periods. More specifically, we have N_E growers in the east and N_W growers in the west. Further, we assume that there is only one processing plant, which is able to process C growers' output per period. Let x be the percentage of growers in the east scheduled to be processed in the first period and let y the percentage of growers in the west also scheduled to be processed in the first period.

(4.3.2) If a grower is scheduled to deliver in a certain period, he may not — depending on the weather — be able to lift his sugar beet when intended. The weather $T_i^{R,P}$ at the location of grower i in region $R \in \{E, W\}$ and period P is assumed to be stochastic and in our case modelled as the weighted average of two normally distributed random variables:

$$T_i^{R,P} = \rho^R N^{R,P} + \sqrt{1 - (\rho^R)^2} N_i^P, \tag{5}$$

where $N^{R,P}$ and N_i^P, are assumed to be independent normally distributed. Notice this implies that $T_i^{R,P}$ is again normally distributed. The parameter ρ^R measures the correlation between the weather within a region. Further we assume that the grower i is not able to lift if his local weather is below a certain threshold c_{tresh}. This implies that the total number of growers processed in period one is given by

$$G_1 = \sum_{k=1}^{\lfloor xN_E \rfloor} \mathbb{1}_{(T_k^{E,1} < c_{tresh})} + \sum_{k=1}^{\lfloor yN_W \rfloor} \mathbb{1}_{(T_k^{W,1} < c_{tresh})}. \tag{6}$$

(4.3.3) In the second period, all growers which could not be processed in the first period (there are $(G_1 - C)^+$ of them) are processed and all other growers have another chance to lift their crops. The total number of sugar beet which could be processed in the second period is therefore given by,

$$G_2 = (G_1 - C)^+ + \tag{7}$$

$$\sum_{k=1}^{\lfloor xN_E \rfloor} \mathbb{1}_{(T_k^{E,1} \geq c_{tresh})} \mathbb{1}_{(T_k^{E,2} < c_{tresh})} + \sum_{k=1}^{\lfloor yN_W \rfloor} \mathbb{1}_{(T_k^{W,1} \geq c_{tresh})} \mathbb{1}_{(T_k^{W,2} < c_{tresh})} + \tag{8}$$

$$\sum_{\lceil xN_E \rceil}^{N_E} \mathbb{1}_{(T_k^{E,2} < c_{tresh})} + \sum_{\lceil xN_W \rceil}^{N_W} \mathbb{1}_{(T_k^{W,2} < c_{tresh})}. \tag{9}$$

$$\tag{10}$$

An inefficiency occurs whenever sugar beet is lifted but cannot be processed on the same period.

(4.3.4) Therefore, we propose to minimize the following objective:

$$V(x, y) = \mathbb{E}[(G_1 - C)^+ (G_2 - C)^+] \tag{11}$$

The two terms represent the excess amount of lifted beet over the processing capacity in the first and second period respectively.

4.4 Description of Simulation

(4.4.1) We evaluate the function V using a Monte-Carlo simulation with 1000 iterations. To ensure comparability, we fix a seed for all simulations.

The parameters in our simulation are given in the following table. The parameter c_{tresh} is chosen such that the probability to lift if scheduled is 90 %. The capacity is chosen in such a way that it matches the expected number of lifted beet. To test for robustness, all results are reported for a low correlation and a high correlation regime.

Name	Parameters 1	Parameters 2
N_W	50	50
N_E	50	50
ρ_W	0.2	0.9
ρ_E	0.2	0.9
C	45	45
$\mathbb{P}(W_j^{R,P} < c_{tresh})$	0.1	0.1

Table 1: This table shows the parameters used in the simulation.

4.5 Simulation Results and Interpretation

(4.5.1) Figures 1 and 2 show heat maps of the value function for different strategies. Since the color in the heat maps is mainly arranged by lines, we can infer that the absolute number farmers, which is to be processed in the first week, should be constant. In the case of large correlation, it seems to be optimal to schedule 45 growers to lift, such that all beet in the first period can always be processed. In the case of low correlation, we find that 50 growers in the first period are optimal, such that the expected number of growers is equal to the capacity of the processing plant.

(4.5.2) Further, since the heat map is darker towards the middle, we can infer that it is better to have an equal amount of growers from the east and west scheduled for the first period, compared to a polarized split.

(4.5.3) To gain a better understanding of the underlying mechanic, figures 3 and 4 show the distribution of lifted beet for the different regimes. The first thing to notice is that in the high correlation regime, most of the outcomes correspond to all scheduled growers lifting the beet. In the low correlation regime, the distribution is centered around its mean. This might explain the optimal amount of growers to be processed in every week. In both regimes, scheduling growers from different regions reduces the tail of the distribution. The implied reduction in risk makes it optimal to diversify the regions within a given period.

4.6 Conclusion

(4.6.1) We have analyzed how to optimally schedule a number of sugar beet growers, which are influenced by regional effects, to lift their beet in order to

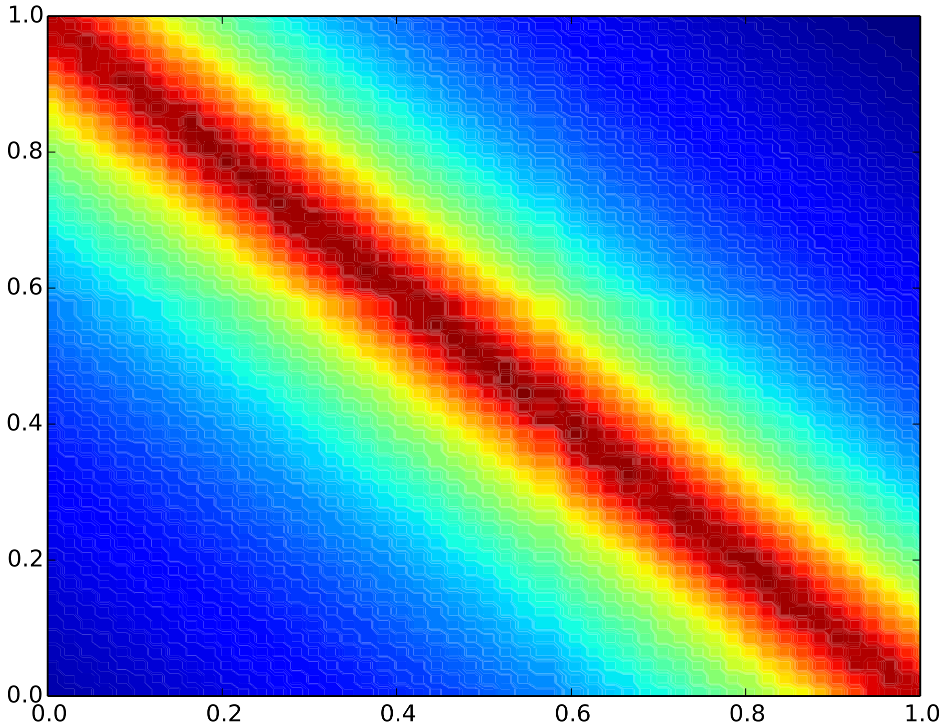


Figure 1: The figure shows a heat map of the value function V for the first set of parameter values. The axis show the proportion in the east (x) and west (y) respectively. A value in the red spectrum corresponds to a lower value function.

be processed by a plant with limited capacity. Two different regimes were considered. In the case of strong comovement within a region, it seems optimal to schedule a number of growers equal to the capacity. In the case of weak comovement, it seems to be better to schedule a number of growers such that the expected number of lifts is equal to the capacity. In any case, better results seemed to be achieved by scheduling growers from different regions within a given period.

4.7 Markov process model

(4.7.1) The collecting and transport and processing of the beet to minimize sugar loss could be considered as a random process — the randomness representing all the uncertainties in the system, including the weather but not limited to that. If the random process is modelled as a Markov process then one way of thinking of the aims would be as minimizing some integrated cost function that is the expected cost integrated over the Cam-

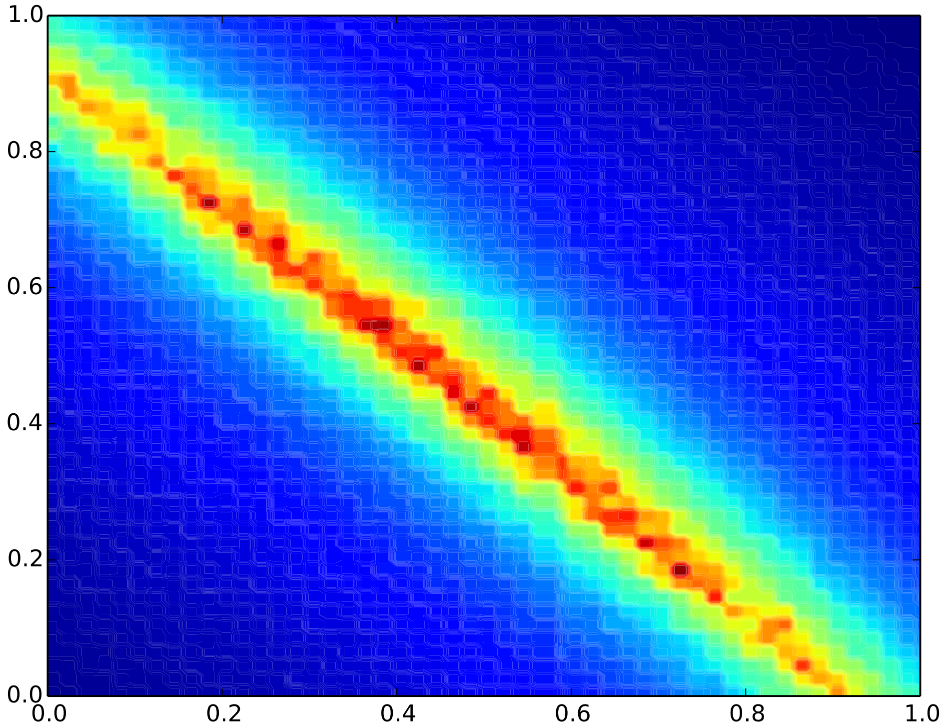
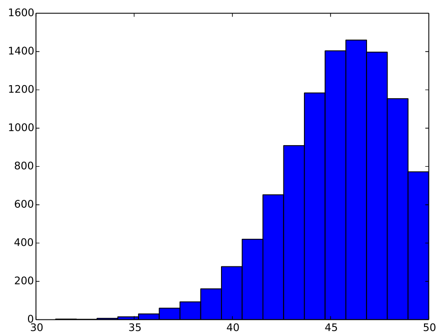
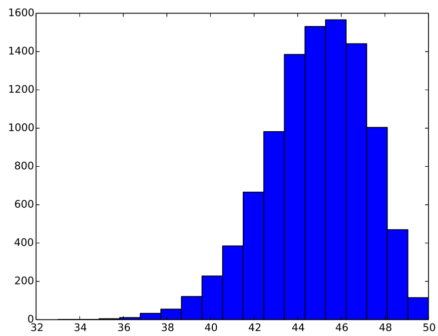


Figure 2: The figure shows a heat map of the value function V for the second set of parameter values. The axis show the proportion in the east (x) and west (y) respectively. A value in the red spectrum corresponds to a lower value function.



(a) $x = 1, y = 0$

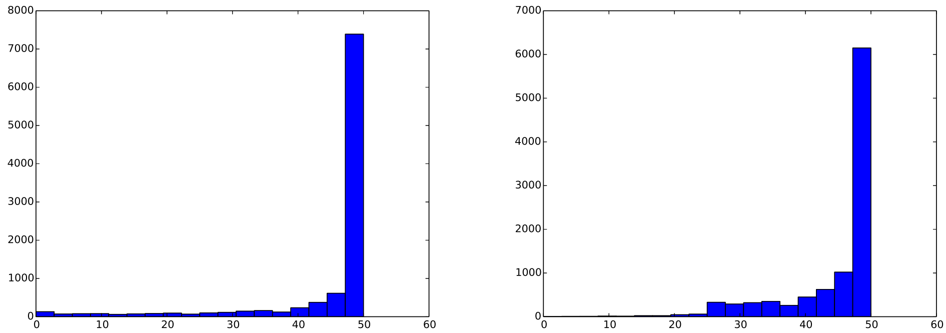


(b) $x = 0.5, y = 0.5$

Figure 3: Distribution of G_1 with correlation $\rho^E = \rho^W = 0.2$ for different mixtures of east and west population.

paign period,

$$\mathbb{E}x\left(\int_0^T c(X(t))\,dt\right), \tag{12}$$



(a) $x = 1, y = 0$ (b) $x = 0.5, y = 0.5$

Figure 4: Distribution of G_1 with correlation $\rho^E = \rho^W = 0.9$ for different mixtures of east and west population.

where $X(t)$ is the underlying Markov process, and $f(X)$ is the cost per unit time incurred when the state is X. We now describe the way that we implemented this approach in a simple case.

(4.7.2) We model the process using a finite state, continuous time Markov chain in order to include random processes affecting the system, generated by individuals changing their plans and decisions due to, for example but not limited to, the weather.

(4.7.3) We considered a simple case where there are two farms, farm A and farm B, one processing plant, one transport system, and one elevator. We assume each farm produces one unit of beet and that multiple farms waiting for transport incurs a storage cost. We assume that the transport system can transport just one unit of beet at a time. We assume additionally that the processing plant is at maximal capacity with one unit of beet, and that any additional units of beet over this one unit incur a storage cost.

(4.7.4) Therefore we consider a 3 dimensional state space, $A \times B \times P$, with one dimension for each farm and a third dimension for the processing plant. Each dimension can be in one of three states. For the farms these are;

- Beets growing (state 0),
- Beets ready for transport (state 1),
- Beets left the farm (state 2).

For the processing plants these are;

- Operating under capacity (state 0),
- Operating at capacity (state 1),
- Operating over capacity, needing to store (state 2).

(4.7.5) In total there are 27 possible states this system can be in, but some are

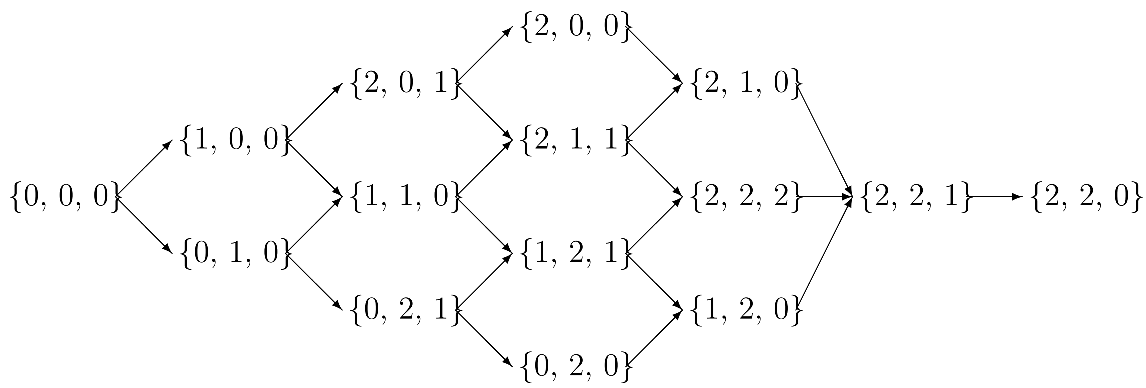


Figure 5: The state space, S, and the acceptable transitions

not permitted. For example, $\{0, 0, 2\}$ (which represents both farms with beets growing and plant operating over capacity) is not a permitted state as we cannot have the processing plant working over capacity prior to any beets leaving the farms. We move between the permitted 15 states with the transitions given by figure 5. We assign a rate, λ_i to each move.

(4.7.6) One way of thinking of the aims in this context would be as minimizing some integrated cost function that is the expected cost integrated over the Campaign period. We assign a cost, f, to each state and this can include a penalty for being in a wasteful or inefficient state. f in state i is the cost per unit time that the chain incurs by remaining in this state. In this way the cost function should force the system to avoid paths which include higher cost states. From this approach we can find a set of rates which minimise the integrated cost function, indicating the rates that would lead to the system with smallest cost. The ratio of the two optimum rates leaving the same state indicates which direction in the state diagram will be favoured, indicating the preferred method to minimise the cost. A more detailed explanation of the technical aspects is given in the Appendix.

(4.7.7) With the tools just presented, the expected value of this cost can be estimated as a function of the parameters (*i.e.* the transition rates and the cost function). In principle, one could simply minimise the expected cost with respect to the parameters, but this would gives us the trivial result that the rates should be as large as possible (*i.e.* if all the rates are large enough, everything happens so quickly that the chain incurs almost no cost, regardless of the states it visits). Therefore, we minimise the following "overall" cost function, rather than just the expected cost:

$$E[\Gamma | X(0) = 0] + \sum_i \lambda_i$$

(4.7.8) In this scenario, large rates are penalised because they have a large contribution in the "overall" cost. Bearing in mind our problem this is also a fair assumption, since large rates might represent, for instance, a faster processing rate at the plant which will be more expensive.

(4.7.9) In order to simplify the model sufficiently so we can visualise the solutions, we initially consider just a few different rates; λ_1 the rate at which the beets become ready on all of the farms, λ_2 the rate at which the beets get delivered from all of the farms to the processing plant, and λ_3 the rate the processing plant processes beets. We assign the same cost of 1 to each state, except for states we have identified as wasteful states to be in. These are $\{1,1,0\}$; the state where all farms have beets awaiting transportation and becoming less sugar-rich, and $\{2,2,2\}$; the state where the processing plant is operating at over-capacity so more beets wait in storage there. We assign these states the cost $1+p$. Additionally, the states $\{2,0,0\}$, $\{2,1,0\}$, $\{0,2,0\}$, $\{1,2,0\}$ reflect the processing plant operating under capacity and these will also be penalised, but with penalty $p_1 < p$.

(4.7.10) In this simple model the allowed transitions are fully specified by the following generator matrix:

$$\begin{pmatrix}
-2\lambda_1 & \lambda_1 & \lambda_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -\lambda_1-\lambda_2 & 0 & \lambda_2 & \lambda_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -\lambda_1-\lambda_2 & 0 & \lambda_1 & \lambda_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -\lambda_1-\lambda_3 & 0 & 0 & \lambda_3 & \lambda_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -2\lambda_2 & 0 & 0 & \lambda_2 & \lambda_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -\lambda_1-\lambda_3 & 0 & 0 & \lambda_1 & \lambda_3 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -\lambda_1 & 0 & 0 & 0 & \lambda_1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_2-\lambda_3 & 0 & 0 & \lambda_3 & \lambda_2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_2-\lambda_3 & 0 & 0 & \lambda_2 & \lambda_3 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_1 & 0 & 0 & \lambda_1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_1 & 0 & 0 & \lambda_1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_2 & 0 & 0 & \lambda_2 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_3 & 0 & \lambda_3 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_2 & \lambda_2 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_3 & \lambda_3 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}$$

(4.7.11) From this simple scenario we can generate plots like that depicted in figure 6. This tells us that in the optimal case once a farm is ready to transport its beets this transportation should occur prior to other farms becoming ready ($\lambda_2 \approx 4\lambda_1$).

(4.7.12) We then add one extra level of complexity to this model by considering 5 different rates; $\lambda_{1,A}$, $\lambda_{1,B}$ the rate at which the beets become ready on each of the farms, $\lambda_{2,A}$, $\lambda_{2,B}$ the rate at which the beets get delivered from each of the farms to the processing plant, and λ_3 as before. This takes into account, for example, the differences in distances between farms and the processing plant. This can give results like that depicted in figure 7. This tells us, among other things, that one farm should have a slower rate of production than the other ($\lambda_{2,B} < 1 = \lambda_{2,A}$).

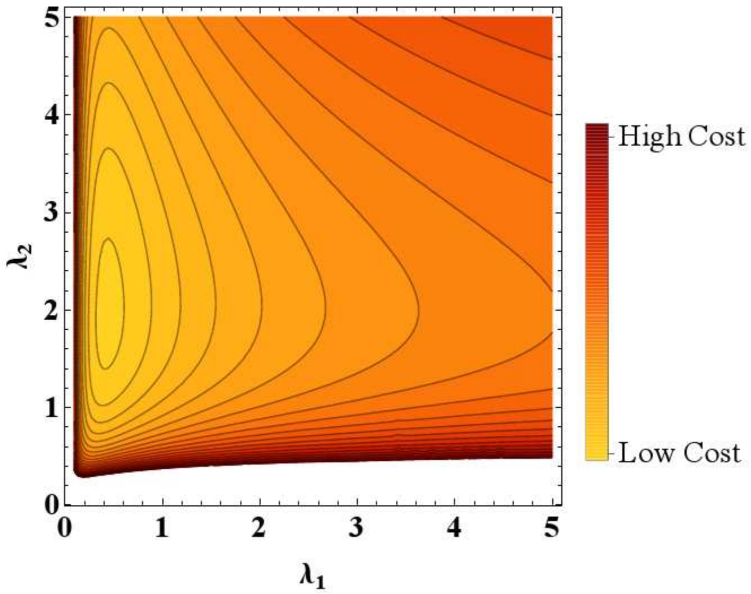


Figure 6: The contour plot of the overall cost function, with fixed parameters $\lambda_3 = 1$, $p = 100$, $p_1 = 10$.

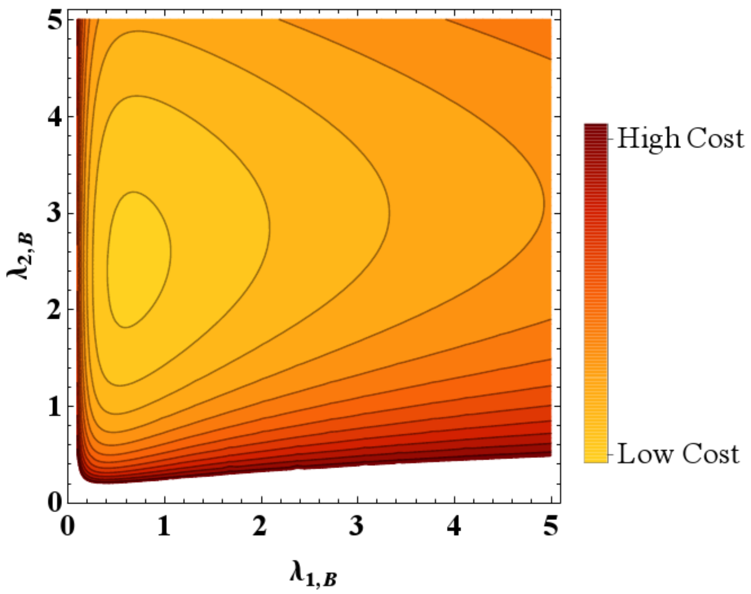


Figure 7: The contour plot of the overall cost function, with fixed parameters $\lambda_{1,A} = \lambda_{2,A} = \lambda_3 = 1$, $p = 100$, $p_1 = 10$.

4.8 Limitations and Extensions

(4.8.1) From this simple model, we have seen that some general qualitative features of the system can be described. A step towards a description closer to reality would require us to define a Markov process on a larger state space, including more farms, transportation companies and processing plants. Due to time limitations, this larger model has not been implemented but the method would extend very easily. For very large state spaces, the generator matrix would be sparse (due to the limited amount of allowed transitions) and a sparse linear solver could be used to obtain the solution.

(4.8.2) Due to its simplicity, however, this model has some limitations as well. In particular, it is not meant to provide a comparison to real data or to estimate the monetary cost of the Campaign. Moreover, the Markov nature of this model is meant to take into account stochastic effects (such as weather conditions) which might play a relevant role in the Campaign, but the validity of the assumption of a Markovian process would need a deeper assessment.

5 Scheduling models

Although the aim of the Study Group is not to produce scheduling methods, we did consider that scheduling methods having the required efficiency measures as part of the cost function would be one of the ingredients needed in any final system.

The scheduling problem we would like to address concerns the ordering of beet transport from numerous farms to a single beet processing plant in order to optimise the processing rate whilst ensuring that the time beets are left to accumulate outside the plant is kept to a minimum.

5.1 Problem Outline

(5.1.1) Consider a single beet processing plant surrounded by N farms F_i each located at a distance d_i from the beet plant. The quantity of beet in tons at each farm is Q_i and the number of trucks working moving between to that farm and the plant at time t_i is N_i. We assume that all the trucks travel at the same speed v and have the same capacity of C_t tons. So the rate $R_k(t_i)$ at which farm F_k can deliver to the plant is given by

$$R_k(t_i) = \frac{N_i Q_k v}{2d_k} ton\ s^{-1}. \tag{13}$$

The 2 in the denominator is due to the fact that the trucks must make a round trips, each of which is twice d_i. We also assume that only a portion of all the farms can operate for 24 hours. Finally we assume that the beet plant can process beets at a constant rate of R_P tons per hour for 24 hours.

(5.1.2) We shall consider one day and aim to deduce close to the optimal ordering of beet collection from the farms in order to optimise the rate of beet processing at the plant whilst minimising the accumulation of beets outside the plant. In order to do this we shall decide upon a preferred (near optimal) profile for the rate of beet delivery at the plant throughout the day based on heurisitic arguments. Then we shall employ a least squares approach which shall order the deliveries from the farms in order to get at close to this profile as possible.

5.2 Beet delivery rate

(5.2.1) The preferred beet delivery rate profile can be chosen. Here we employ some heuristic arguments to choose an example profile. We are given that some, but not all farms, operate for the entire 24 hour period. We suppose that there are not enough 24 hour farms to achieve the beet processing rate R_P so we require an accumulation of beets before the end of the working day so that the plant can achieve R_P at night. This suggests that a profile as shown in Figure 8 would be sensible.

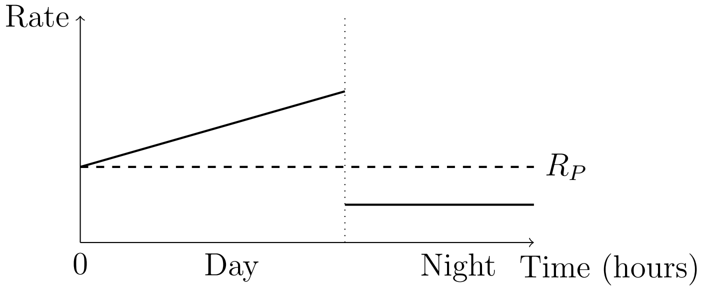


Figure 8: Solid line is the proposed profile for rate of beet arrival at plant. Dashed line is the processing rate of the plant.

5.3 Re-scheduling

(5.3.1) We can assume our original schedule, in which each haulier delivers beet to the plant at a constant rate throughout a period of length T time intervals (of some desired granularity), to be of the form

$$M = \begin{pmatrix} R_1(t_1) & R_2(t_1) & \dots & R_N(t_1) \\ R_1(t_2) & R_2(t_2) & \dots & R_N(t_2) \\ \vdots & \vdots & & \vdots \\ R_1(t_T) & R_2(t_T) & \dots & R_N(t_T) \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix} = \begin{pmatrix} f(t_1) \\ f(t_2) \\ \vdots \\ f(t_{24}) \end{pmatrix},$$

Here the vector a_k corresponds to the distance of farm k, so $a_k = 1/(2d_k)$, whilst $R_k(t_i)$ corresponds to the rate at which beets are hauled from farm k at time t_i, so as in (13). The default case is where each $R_k(t_i)$ and

$f(t_i)$ is constant over time t_i. Denote this default schedule as R_0. This will in turn result in a constant RHS vector F_0. Suppose we require some alternative RHS vector F, given that the vector a is fixed, how can we reschedule so that, for example, there is an increase in beet rate towards the end of the day? Suppose that our desired schedule matrix $R = R_0 + \tilde{R}$, where $\tilde{R}$ corresponds to the changes made to the default schedule, which gives us our new RHS rate vector $F = F_0 + \tilde{F}$. We now have a new system derived from this,

$$\tilde{R}a = \tilde{F}.$$

This system corresponds to the reshuffling of the schedule, if $\tilde{R} = 0$ and $\tilde{R} = 0$ then no rescheduling has taken place. But suppose we have a desired new schedule F; then we need to find $\tilde{R}$ which gives us $\tilde{F} = F - F_0$. Whilst the system is under determined, it will typically have infinitely many solutions. Intuitively, the rate will increase when more of the trucks are moving from nearby farms, as the travel time from the farm to the plant is shorter. Also, there are certain properties about $\tilde{R}$ which must hold.

(a) As the number of trucks working at any time t_i is constant, all entries of each row of $\tilde{R}$ must add up to zero. In other words, if an extra truck is working at one farm, it means one fewer is working somewhere else.

(b) As the total number of beets needing to be hauled over the whole day is fixed, all entries of each column of $\tilde{R}$ must add up to zero. In other words, if less beet is going to be hauled at one hour, then more will have to be hauled later on to make up for this.

(c) For every entry of the matrix $-Q_i \leq R_k(t_i) \leq Q_i$, i.e. no more than all of the beets of a single farm can be moved in one go.

(5.3.2) We shall call every type of matrix that satisfies these properties a 'beet matrix'. Due to the first and second constraints, an $N \times M$ beet matrix has $(N-1)\times(M-1)$ degrees of freedom. Note that F is also a beet matrix, as an increase in rate at one time will mean a corresponding decrease in rate at another. Note also that a linear combination of beet matrices is also a beet matrix. Define a 'simple beet matrix' to be a matrix with only four non-zero entries, which lie in a square. For example, consider the simple beet matrix

$$A_{(m,n,x,y)} = \begin{cases} a_{m,n} = -1 \\ a_{m+x,n+y} = -1 \\ a_{m+x,n} = 1 \\ a_{m,n+y} = 1 \\ \text{(All other entries zero)} \end{cases}$$

(5.3.3) Although this has not yet been attempted, it is suspected that $\tilde{R}$ can be constructed by taking a linear combination of these. The proposed method is as follows

(a) Sort the vector a from low to high.

(b) Use a multiple of the beet matrix $A(1, T-1, N-1, 1)$ to ensure that the bottom entry of $\tilde{F}$ is met. This is essentially taking the trucks from the farthest farm and shuffling them to the nearest farm at the end of the day to boost the rate at this time.

(c) Moving up one row of F at a time, use the row above in $\tilde{A}$ to help correct for the previous step. The second beet matrix added will be $A(1, T-2, N-1, 1)$.

(d) If the third beet matrix constraint is ever violated, move in to columns modify columns 2 and $N-1$ instead, moving in additional columns if necessary.

(e) There will be no remaining row to correct the top row - but this will not matter. We know that the desired RHS $\tilde{F}$ is also a beet matrix, so provided $\tilde{F}_2$ to $\tilde{F}_T$ are as required then $\tilde{F}_1$ must be as required, as there are only $T-1$ degrees of freedom in the beet matrix $\tilde{F}$, so the entry $\tilde{F}_1$ must be as required.

Unfortunately there was not enough time to test this method during the week, but hopefully this or something similar could be used.

6 Prediction markets for campaign planning

6.1 Description of prediction markets

(6.1.1) We will start with the definition given by Leigh and Wolfers: *prediction markets are markets where participants trade contracts whose payoff depends on unknown future events. The defining feature of a prediction market is that the price of these contracts can be directly interpreted as a market-generated forecast of some unknown quantity.* [10] Their mechanism relies on the efficient markets hypothesis: *the price of a financial security or prediction market contract reflects all available information.* [10] Therefore prediction markets are an example of efficient crowdsourcing — aggregating dispersed, and often contradictory, knowledge from a group of people to obtain very precise information about the outcome of a future event.

(6.1.2) Modern approach to prediction markets began in 1988, when three economists of the Iowa University created a market to predict the outcome of the presidential election (Bush vs Dukakis). It was observed that in any given moment in time such market gave much better forecast than all major polls. The experiment has been carried on for many other elections and the comparison shows that it beats all polls in about 75% of the times. The advantage of markets is even bigger when the time to election is long. [11]

(6.1.3) After the success of Iowa Electronic Markets the interest in prediction markets grew rapidly. Currently markets are used by many large corporations (Google, Microsoft, IBM, Lockheed Martin, *etc.*) as a tool to assess the probability that a project will end as planned, that a sales goal will be achieved or as a tool to estimate the market potential of innovative products. [12] For instance, General Electric has been running markets for new ideas and products originated by employees. Eli Lilly, a large pharmaceutical company, ran a prediction market to support choice of new drugs for further development, primary decision factor being market potential. Further, in the BRAIN[1] project — an internal research at Hewlett-Packard, it was shown how to run prediction markets with small numbers of participants (up to 10 people) and still obtain meaningful results.

(6.1.4) Apart from internal corporate applications there are also many publicly available commercial markets (*e.g.* Hollywood Stock Exchange, Intrade), where operators often profit from fees or selling complex analyses derived from the market data. Furthermore, even DARPA[2] and IARPA[3] have implemented prediction markets [13], mainly to obtain accurate predictions important for the American military or intelligence community.

6.2 Reliable information source for campaign planning

(6.2.1) Prediction markets are primarily a source of information that can be aggregated efficiently from their participants. This property allows to use them to reduce the uncertainties in the process of campaign planning. For instance, the information that could possibly be gathered through a prediction market encompass the main time points of the campaign (start, peaks, end), predictions about the quality of the soil and about yield in different regions and time periods.

(6.2.2) Another very important benefit from the use of prediction markets is the fact that all parties, including farmers, would be involved in the campaign planning process, everyone could feel that his voice is heard and that he can have impact on the entire process. Such a prediction market could (or even should) be incorporated into a larger system for campaign planning, which in part would work as a public consultation platform.

(6.2.3) For instance, take one of the biggest uncertainties in the planning process — the yield from fields. On one hand it depends on such hard to predict

[1] Behaviorally Robust Aggregation of Information in Networks

[2] Defence Advanced Research Projects Agency — an American government agency supporting large scientific projects that might be useful for military purposes

[3] Intelligence Advanced Research Projects Agency — the counterpart of DARPA devoted to intelligence purposes

factors as weather. The knowledge of the team involved in the campaign planning will probably rely to a great extent on the data from weather forecasts. However due to experience using market for crowdsourcing farmers might give better results. For example the futures market for orange juice concentrate predicts Florida weather better than the National Weather Service does. [11] There are also additional factors best known to the people that are in the field (figuratively and literally). These include: the quality of the soil (variable and dependent on recent usage history), historical yields, information from the current season, such as the schedule of all agricultural tasks that have been done or are to be done (sowing, fertilization, irrigation). All this data is of great significance for predicting the yield throughout the campaign.

(6.2.4) All this information can be easily aggregated into yield forecasts by a properly set prediction market involving farmers. Others participants are also welcome, as they increase diversification of information that in turn can enhance accuracy of a prediction market ([10] and [12]). The questions on the market, that have to be binary, could ask about several levels of yields for every district or county separately. As an example we provide a set of questions for Uttlesford district in Essex county for one given week would take the form [4]:

> The yield in Uttlesford district from 6 to 12 October to be below 3000 tons.
>
> The yield in Uttlesford district from 6 to 12 October to be over 3000 and below 6000 tons.
>
> The yield in Uttlesford district from 6 to 12 October to be over 6000 below 9000 tons.
>
> The yield in Uttlesford district from 6 to 12 October to be over 9000 tons.

Similar sets of questions could be posed for every district or county for every week (or even every day) of interest.

(6.2.5) Such a prediction market could be incorporated into a larger software tool that would use different algorithms and solutions to help in the campaign planning process and would allow for fast and efficient information exchange between all parties involved in the campaign.

7 Interaction models

In this section we describe some of the possible mechanisms that were discussed for how the different participants in the system could usefully interact with each other, following the information-gathering that can be effected by a prediction market.

[4] Number of toms and dates are arbitrary.

Broadly speaking the mechanisms discussed were for how they can exchange bids and offers.

7.1 Bidding and offering processes

(7.1.1) We have discussed various potential bidding and offering processes. Some are based on the idea of a limit order book. This is used in financial markets to match bids and offers in buying stock. In the simplest case, imagine that people wanting to buy a stock make offers that they will buy certain amounts at certain prices. On the other side, people wanting to sell make offers that they will sell certain amounts at certain prices. Then the market-maker clears the market by allocating first the highest bid to the lowest offer, then the next highest and so on. If there are ties, which is usually the case because the allowed bids and offers are discretized) then we will reach a situation where there is more demand for the cheapest offer than the amount available. In this case, there are two ways of making the allocation.

(a) Proportional: the bidders each receive a particular proportion of their bid, the proportion being the supply:demand ratio.
(b) First-come-first-served: the supply is allocated to the bidders in the time-order their bids came in.

(7.1.2) If this kind of process were applied to the grower-haulier allocation process, then growers would enter bids of what they would pay for transport of their beet in a particular time-slot. They could make multiple bids, expressing (for instance) their preference for day or night,their preference between different days. But only *one* of their bids will be accepted.

(7.1.3) On the other side, hauliers make offers of what haulage capacity they can supply in each time-slot, and at what price. Then the bids and offers are stacked up and handled in a similar way to the outline above. In the beet context it seems that the second method of dealing with ties will fit better with the way the industry operates, since it tends to ensure that more growers are collected in consecutive time-slots.

(7.1.4) If such a system is to operate in the beet industry, it needs to include not just growers and hauliers but the processing plants too. This introduces complications but a potential approach was discussed and is outlined here. The information that the growers and hauliers enter will be as mentioned above. But the processing plants will also need to state what price they are prepared to pay for beet delivered in particular time-slots. This may vary from plant to plant. Also the time-slots may have limited amounts of beet that can be booked into them.

(7.1.5) The process of clearing the market then could proceed by a grower accepting a price and amount from a processing plant, and then having a certain time during which he accepts a certain offer, or offers, of transport, and then confirms the whole arrangement when the different elements are in place.

(7.1.6) It is important to avoid the situation of booking in to the processing plant but then not being able to arrange transport.

(7.1.7) In this process, which is effectively an auction, the participants need to have an incentive to bid their true values. This is done (in more conventional auctions) by a Vickrey auction, in which the item is sold to the highest bidder but at the price offered by the second-highest bidder. A similar scheme would be needed in the beet market, but may have complications because of the 3-participant nature.

A Expected cost integrals

We give a more detailed explanation of the method used in Subsection 4.7.

A.1 Distribution of path integrals

(A.1.1) Let $X(t)_{t\geq 0}$ be a continuous-time Markov chain, which takes values in the set $S = \{1, 2, 3, \dots\}$ of allowed states and consider A to be a subset of S containing all the states except the final one. We want to evaluate the distribution of path integrals given by:

$$\Gamma = \int_0^\tau f_{X(t)} dt$$

where f is a non-negative real cost function and $\tau = \inf\{t > 0 : X(t) \notin A\}$ is the hitting time of the final state.

(A.1.2) The function f_i has the interpretation of cost per unit time of staying in state i and, therefore, Γ is the total cost over the period spent in A (with the assumption that A does not contain any absorbing state). The Laplace transform of the distribution of path integrals defined above is given by:

$$y_i(\theta) = E_i \left[e^{-\theta\Gamma}\right]$$

with the understanding that $y_i(\theta) = 1$ for $i \notin A$.

(A.1.3) The following theorem provides a simple way of calculating this.

Theorem 1. *For each $\theta > 0, y(\theta) = (y_i(\theta), i \in I)$ is the maximal solution to the system of equations:*

$$\sum_{j\in I} q_{ij} z_j = \theta f_i z_i, \quad i \in A$$

with $0 \le z_j \le 1$ *for* $j \in A$*, and* $q_{i,j}$ *the elements of the generator matrix and* $z_j = 1$ *for* $j \notin A$ *in the sense that* $y(\theta)$ *solves this system of equations and, if* $z = (z_i, i \in I)$ *is any solution, then* $y_i(\theta) \ge z_i, \forall\ i \in I$*.*

(A.1.4) The Laplace transform of the distribution is closely related to the moment-generating function (via a minus sign in the exponential), so by formal differentiation of the system of equations given in the theorem we can obtain all the moments of the distribution. In particular, one formal differentiation gives us the expected value of the path integral (conditional on the chain starting at $i \in A$).

(A.1.5) This methodology is very similar to the potential theory for Markov chains (as presented, for example, by Norris [6]) and indeed gives the same exact results. In such a context, one in principle could also consider discount factors but we have not included them in our model. See also the description by Pollett et al.[5].

(A.1.6) The basic idea behind this technique is that we can define a continuous-time Markov process on a state space with an absorbing state; the process will, then, spend some time in the bulk of the state space, where it incurs in a certain cost per unit of time spent in it, and then it will eventually hit the boundary, where it stays forever (in more generality, there could also be a cost when hitting the boundary, but this would just be a constant added to the overall cost).

Bibliography

[1] Sugar Beet Transport Efficiency Study. Final Report. NFU/British Sugar. 7 September 2009. Chris Douglas. Supplied by Martin Robinson.

[2] Coupled modelling of sugarcane supply planning and logistics as a management tool. Computers and Electronics in Agriculture, **68**, 168–177 (2009). P.-Y. Le Gal, J. Le Masson, C.N. Bezuidenhout, L.F. Lagrange.

[3] A discrete event simulation of sugar cane harvesting operations. Roberto Semenzato, Sebastián Lozano, Ricardo Valero. Journal of the Operational research Society, **46**, 1073–1078, 1996.

[4] Tom Hosking. Games with exhaustible resources.

[5] Pollett, P.K., Stefanov, V.T., Path integrals for continuous-time Markov chains. Journal of Applied Probability **39** No. 4, pp. 901–904, Dec 2002.

[6] Norris, Markov Chains, Cambridge University Press.

[7] Assessing the potential benefits of alternative cane supply arrangements in the Australian sugar industry. Agricultural Systems, **76**, 623638 (2003). A.J. Higgins, R.C. Muchow.

[8] Optimizing cane supply decisions within a sugar mill region. Journal of Scheduling, **2**, 229-244 (1999). A.J. Higgins.

[9] Reducing harvest costs through coordinated sugarcane harvest and transport operations in Louisiana. Journal Association Sugar Cane Technologist, **28**, 32-41 (2008). M.E. Salassi, F.G. Barker.

[10] Leigh, Wolfers, Prediction markets for business and public policy, The Melbourne Review, **3** (1), May 2007.

[11] Stix, Super Tuesday: Markets predict outcome better than polls, Scientific American, 4 February 2008.

[12] Ho, Chen, New Product Blockbusters: The Magic and Science of Prediction Markets, California Management Review, **50** (1), 2007.

[13] `http://www.goodjudgmentproject.com/`
and `http://hanson.gmu.edu/policyanalysismarket.html`