Report on PIMS 2008 GIMMC and IPSW

University of Regina

June 9–20

1 Summary

The Pacific Institute for Mathematical Sciences (PIMS) has sponsored an annual Industrial Problem Solving Workshop (IPSW) since 1997. The aim of the IPSW is to create a mutually beneficial link between researchers in industry and academic mathematical scientists. Faculty and students from the academic community study problems brought by industrial participants during the week-long workshop. Their results are presented at the end of the week. The benefits of an IPSW are numerous.

In order to help train the participants for the IPSW, PIMS hosts the Graduate Industrial Mathematics Modeling Camp (GIMMC) during the week preceding the IPSW. The purpose of the camp is to teach graduate students mathematical modeling methods from experts in the field. A cross-section of relevant industrial problems and modeling techniques are presented.

The GIMMC took place June 9–13 and the IPSW took place June 16–20. Both events were held on the University of Regina campus.

Financial support was an important consideration for these two events because the participants had housing provided, travel support, and a food allowance. In addition, there were five mentors whose expenses were covered. Financial support was provided by the Province of Saskatchewan, President's Office at the University of Regina, Faculty of Science at the University of Regina, Department of Mathematics and Statistics at the University of Regina, MITACS, PIMS, NSERC Prairie Fund, and the Australian Mathematical Sciences Institute.

2 Participant List

Australia:
Roslyn Hickson (Australian Defence Force Acad.), Asef Nazari (Ballarat), Giang Nguyen (South Australia), Melanie Roberts (Western Australia)

Canada:
Taylor Barrett (Regina), Chakra Bakyar (Regina), Irma Elizabeth Diaz Bobadilla (Regina), Michael Cavers (Regina), Shannon Collinson (Guelph), Janice Cotcher (Regina), Bridget Fortowsky (Regina), Ortho Flint (Waterloo), Asef Ganjehlou (Regina), Shaughnessy Hawkins (Guelph), Matt Hennessy (Ontario Institute of Technology), Yuhui Huang (Regina), Parisa Hudson (Western), Stephen Hudson (Western), Harish Kashyap (Regina), Zanin Kavazovic (Laval), Dong Won Kim (Regina), Petko Kitanov (Guelph), Nathan Krislock (Waterloo), Xiaoping Liu (Regina), Heidi Muller (Guelph), Sadia Mwangangi (Regina), Dominic Nelson (Guelph), Fridahus Oloude (Western), Hugo Ordonez (Regina), Sarah Plosker(Regina), George Price (Regina), Mary Jane Richardson (Guelph), Justin Schwark (Regina), Notice Ringa (Guelph), Ryan Tiefenbach (Regina), Ewout van den Berg (UBC), Yongjun Xing (Regina), Sheena Zhang (Waterloo)

Greece:
Nikolas Karalis (Athens)

Mexico:
Angélica Caudillo-Mata (CIMAT), David de la Rosa (CIMAT), Norberto Flores (CIMAT), Javier Vigueras-Gómez (CIMAT)

USA:
Aren Arakelyan (City College), Hernando Bermudez (Montana), David Clark (Michigan Tech), Ha Dang (Long Beach), Yogesh Joshi (NJIT), Manmeet Kaur (NJIT), Harish Krishnamurthy (Northeastern), Lois Kwon (WSU), Rachel Robertson (Michigan Tech), Yegor Sorokin (Iowa)

3 Mentor List

The mentors play a vital role in the GIMMC by providing training on how to approach typical industrial problems. The mentors are expected to play a strong role in the IPSW week as well, but they have no specified duties. Some will stick with a particular problem while others float between several problems providing guidance as appropriate. The mentors for these two workshops were

- Laura Cowen (University of Victoria)
- Ed Doolittle (University of Regina)
- Neville Fowkes (University of Western Australia)
- Don Kreher (Michigan Technological University)
- Roge Mamon (University of Western Ontario)

In addition to the mentors, Michael Kozdron, from the Department of Mathematics and Statistics at the University of Regina, played a vital role in the IPSW. Dr. Amir Amiraslani attended as an observor, was very helpful throughout both weeks, and will be coordinating future workshops from his position in the Department of Mathematics and Statistics at the University of Calgary.

4 Graduate Industrial Mathematical Modeling Camp

In this section we present summaries of what was done in each of the five groups during the GIMMC week.

4.1 Group 1: Laura Cowen, Mentor

Mark Recapture with batch marks but no remarking

Ha Dang, Yuhui Huang, Rachel Robertson, Elizabeth Diaz, David de la Rosa, Flavio Vigueras, Ryan Tifenbach, and Harish Kashyap.

Mark-recapture experiments can be used to estimate the size of a population. However, when tagging individuals, usually unique tags are employed. We look at the complication of marking fish by date (batch marking) and develop methodology to estimate population size.

The joint hypergeometric maximum likelihood estimator (JHE) can be used for batch marked individuals (Neal et al 1993); however, it does not allow for immigration or emigration. Neal et al (1993) extend the JHE to incorporate immigration and emigration (IEJHE) which is implemented in program NOREMARK (White, 1996); however, the number of marked animals in the study area are assumed to be known for each occasion.

Study Design and Data

At a single location in a lake, an initial sample of fish was obtained and fish were batch marked with coloured tags before release. On 5 subsequent occasions, a sample of fish was obtained, the fish currently tagged were counted and a random sample of untagged fish were given a different colour batch tag. A final sample was obtained where fish were counted but tagging did not occur.

Table 1 provides the data from a 2004 study. In addition to the tagged fish, the total number of fish in the sample was recorded. For example, on occasion 2, 922 fish were captured of which 4 had green marks and 403 were given yellow marks. The remaining fish were released with no marks.

Batch marking fish means there is no way to identify the fish individually using the tag. Additional concerns with this study were that the system is assumed to be an open population, thus there may be immigration, emigration, recruitment or mortality. It is assumed that capture probability was homogeneous. The objective of this project was to obtain an estimate of population size.

Table 1: Batch marked data from a 2004 study.

Release Occasion	Number Marked	Tag Colour	Recovery Occasion 2	3	4	5	6	7
1	323	Green	4	8	2	13	3	2
2	403	Yellow		23	4	48	13	12
3	407	Red			2	47	14	16
4	245	Orange				17	5	5
5	373	Green Tail					15	14
6	749	Yellow Tail						14
Total Sampled			922	1760	1035	4190	2650	1820

Notation

Statistics

- k the number of sampling occasions.
- r_{ij} the number of fish tagged and released at occasion i and recovered at occasion j, $i = 1, 2, ..., k-1$; $j = i+1, ..., k$.
- R_{i0} the number of fish tagged and released at occasion i with colour i.
- u_i the number of untagged fish captured and released at occasion i.

Parameters

- p_i the probability of capture at occasion i.
- ϕ_i the probability of surviving and remaining in the population between occasions i and $i+1$, given the fish was alive and in the population at occasion i.
- U_i the total number of unmarked fish in the population at occasion i.
- B_i the number of births (immigration) into the population at occasion i.

Model Development

We modeled the number of recovered fish (r_{ij}) using a binomial distribution, $r_{ij} \sim Binomial(R_{i0}, \phi_i\phi_{i+1}...\phi_{j-1}p_j)$. For example, the probability a fish tagged and released at occasion 1 and recaptured at occasion 2 is $\phi_1 p_2$; the fish survives from occasion 1 to occasion 2 (ϕ_1) and is recaptured at occasion 2 (p_2). Similarly, the probability that a fish tagged and released at occasion 2 is recaptured at occasion 5 is $\phi_2\phi_3\phi_4 p_5$. Thus we can develop the likelihood for the probability of recapture (marked fish) (1).

$$L(\phi, p)_{marked} = \prod_{i=1}^{k-1} \prod_{j=2}^{k} \binom{R_{i0}}{r_{ij}} (\phi_i \phi_{i+1} ... \phi_{j-1} p_j)^{r_{ij}} (1 - \phi_i \phi_{i+1} ... \phi_{j-1} p_j)^{R_{i0} - r_{ij}} \tag{1}$$

Similarly, we can develop a likelihood for the unmarked fish, again modeling the number of unmarked fish with a binomial distribution, $u_j \sim Binomial(U_j, p_j)$ (2).

$$L(p)_{unmarked} = \prod_{j=2}^{k} \binom{U_j}{u_j} p_j^{u_j} (1 - p_j)^{U_j - u_j} \tag{2}$$

As the marked and unmarked fish are considered independent, we can obtain the overall likelihood by multiplying together the two likelihoods: $L = L_{marked} \times L_{unmarked}$.

Parameter Estimates and Model Selection

There are no closed form estimators for the parameters, however they can be estimated analytically using a Newton-Raphson type estimator. We can estimate the total number of unmarked animals in the population at occasion i (U_i) as $\hat{U}_i = \frac{u_i}{\hat{p}_i}$. Further, we can estimate the number of births at occasion i as $\hat{B}_i = \hat{U}_{i+1} - \hat{U}_i \hat{\phi}_i$. Finally, the number of animals alive at time i can be estimated as the estimated number of unmarked fish plus the estimated number of marked fish in the population, $\hat{N}_i = \frac{u_i}{\hat{p}_i} + \frac{\sum_{j=1}^{i-1} r_{ij}}{\hat{p}_j}$. Standard errors for $\hat{U}_i$, $\hat{B}_i$ and $\hat{N}_i$ can be obtained using the delta method.

R statistical software (R Development Core Team, 2008) was used to maximize the likelihood, obtain standard error estimates and calculate AIC (Akaike's Information Criterion) for model selection purposes. Due to limited time, we considered only three models: constant survival and capture probabilities (ϕp), constant survival probabilities and capture probabilities varying over time (ϕp_t), and survival and capture probabilities varying over time with the last survival probability fixed to 1 ($\phi_t p_t, \phi_6 = 1$). In this last saturated model, the last survival probability is fixed to 1 due to non-identifiability of the parameters $\phi_5 p_6$.

Results

Table 2 shows the AIC values for the three considered models and the number of parameters in each model. As the $\phi_t p_t, \phi_6 = 1$ model had the lowest AIC value, it was chosen as the best model. As the other models had large Δ_i values, model averaging was not employed.

Table 2: Akaike's Information Criterion (AIC), the number of parameters (K), and delta AIC values ($\Delta_i = AIC_i - AIC_{min}$) for the three considered models.

Model	AIC	K	Δ_i
ϕp	2478.3	2	118.7
ϕp_t	2392.3	7	32.6
$\phi_t p_t, \phi_6 = 1$	2359.6	11	0

Parameter estimates and standard errors for the $\phi_t p_t, \phi_6 = 1$ model are shown in Table 3. All survival probability estimates were high (> 0.80) and recapture probabilities estimates were low (< 0.15).

Table 3: Maximum likelihood parameter estimates for the model $\phi_t p_t, \phi_6 = 1$, along with the standard error estimates underneath.

	Occasion					
Parameter	1	2	3	4	5	6
p	0.012	0.043	0.009	0.114	0.040	0.031
	0.00004	0.00006	0.00001	0.00006	0.00011	0.00003
ϕ	1.105	1.039	0.817	0.929	0.863	-
	0.000	0.000	0.012	0.042	0.038	-

Population estimates were obtained for each of the sampling occasions. These ranged from 20000 to 120000. Standard errors for these estimates were not obtained due to a lack of time.

Discussion and Future Work

There are several modeling issues that still need to be addressed. First, some of the survival estimates are above 1, this could be alleviated through the use of design matrices and the use of a logit-link function which would constrain the parameter to be between 0 and 1 (Lebreton et al, 1992). Further the interval of time between occasions was not constant, thus survival parameters are defined over different time periods. This could be adjusted for by modeling survival probabilities as a function of time. Finally, variances estimates and confidence intervals need to be obtained for the population size at each occasion.

It is possible that in similar studies, fish may be lost on capture. This can be dealt with by expanding the model to incorporate these losses as was done by Schwarz and Arnason (1996).

Acknowledgements

Treveor Haynes (Department of Geography, University of Victoria) provided the data. Carl Schwarz (Department of Statistics and Actuarial Sciences, Simon Fraser University) provided direction concerning the model.

References

Lebreton, J.-D., Burnham, K. P., Clobert, J., and Anderson, D. R. 1992. Modeling survival and testing biological hypotheses using marked animals: A unified approach with case studies. *Ecological Monographs* **62**, 67-118.

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. url:www.R-project.org.

Neal, A. K. White, G. C., Gill, R. B. Reed, D. F., and Olterman, J. H. 1993. Evaluation of mark-resight model assumptions for estimating mountain sheep numbers. *Journal of Wildlife Management* **57**, 436-450.

Schwarz, C. J. and Arnason, A. N. 1996. A general methodology for the analysis of capture-recapture experiments in open populations. *Biometrics* **52**, 860-873.

White, G. NOREMARK: Population estimation from mark-resighting surveys. *Wildlife Society Bulletin* **24**, 50-52.

4.2 Group 2: Ed Doolittle, Mentor

Snowplowing The Streets Of Regina

David Clark, Mary Jane Richardson, Nikolas Karalis, Nathan Krislock, Giang Nguyen, Ewout Van Den Berg, Notice Ringa, and Ortho Flint.

The problem statement was to design an efficient snow removal plan for the City of Regina. (Several members of the team live in places where there is no snow, but their contributions demonstrate that mathematicians don't need physical experience to work successfully on a problem.) Most of the second day was spent learning about the local snow removal situation and general snow emergency strategies employed in North America. For several participants with limited or no experience with snow, it was fascinating to learn new concepts such as *deadheading* (traversing without plowing) and *sand/salt spreading*. The team quickly agreed on the motto "No Street Left Behind." Coming from diverse research backgrounds, there were very different ideas about how to approach the problem. There was even a heated debate on what *efficient* meant! In the end, the team divided into two groups: cycle decomposition and single commodity network flow.

The commodity flow group adapted a Mixed Integer Linear Programming model [1], to appropriately incorporate their problem specific assumptions and constraints. Then they used CPLEX solver to solve the model and identify which streets to plow or to deadhead during given periods and corresponding frequencies. The final step was to use Fleury's algorithm to find an Eulerian cycle determining the order of visits for these streets. They also created an interactive MATLAB simulation to demonstrate their solution on a simplified map of the city of Regina. Their first presentation was to inform other groups about the problem and intended approach. The second presentation (held on the last day, when the presenter's laptop was still as functional as a brick) was to present a final solution. The team was delighted that a representative from the city of Regina council attended that talk, and so they carefully avoided phrases like *more efficient* or *much better* when presenting their solution.

To go from not knowing that there were street priorities during snow removal to presenting a final route map — over five intensive days and two late nights — was a long, frustrating and exciting journey for all of the team. Even though the model had a lot of room for improvement, they were proud of their achievement.

References

[1] N. Perrier, A. Langevin and C-A. Amaya, "Vehicle Routing for Urban Snow Plowing Operations", *Transportation Science*, **42** (2008), 44–56.

4.3 Group 3: Neville Fowkes, Mentor

There were three separate subgroups working under Neville's supervision. There is a report from each of the subgroups.

Piped Water Cooling of a Concrete Slab

Shannon Collinson, Janice Cotcher, Asef Nazari Ganjehlou
Zanin Kavazović, Heidi Muller, George Price, and Hugo Rodríguez

A First Model

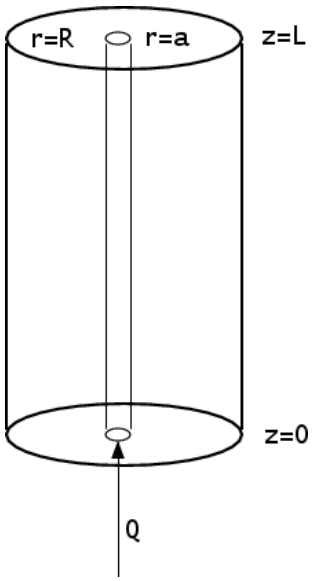

Figure 1: The first model

Concrete is the most widely used building material in the world. It is composed of cement; water; cemetitious materials such as fly ash and slag cement; aggregate—gravel, limestone or sand—and chemical admixtures. After mixing with water and placement, concrete solidifies due to the chemical process known as hydration. Through hydration, the cement bonds the other components together and forms a rock like structure. If the hydration process occurs too quickly the concrete will not form strong bonds so it is desirable to slow down the reaction by cooling. Furthermore. The hydration reaction process is highly exothermic and the cooling of this poses a major

problem in construction since the heating of the material can compromise the structure of the concrete.

One place where the difficulties posed by the exothermic reactions is seen in building a dam. A dam is constructed by pouring concrete slabs of dimension $10m \times 10m \times 3m$. The issue of the exothermic reaction occurs since the conductivity of concrete is very small; it can take many years for concrete to cool, if left to its own devices. The exothermic reaction and the low conductivity of concrete creates an environment such that the temperature within the concrete can change by up to 150K. If the concrete is not cared for, the temperature of the concrete can change up to 323K, at which point thermal stress occurs leading to structural weakening and construction delays. To reduce the build up of temperature and increase the strength of the concrete, cold water is piped through until the end of the curing process; this process takes about 28 days in perfect conditions - controlled temperature and humid atmosphere. After the curing process is complete, the pipes are filled in with concrete. The objective is to determine the length, radius, and temperature and flux of water through the pipes, as well as the required spacing and geometric arrangement of pipes in the water network to achieve a cost effective way to cool the concrete; a complex and difficult initial task. In order to better understand the physics, we first consider the heat exchange between a single pipe of radius a carrying water with a 'sleeve' of concrete of radius R surrounding it; $2R$ can thought to be the spacing between pipes in the network. Eventually R needs to be be determined to optimize the system.

The problem breaks up into two connected parts: a heat transfer problem for the concrete sleeve and a heat transport equation for the water in the pipe.

Heat diffusion within the concrete cylinder

Assuming the heat transfer occurs mainly in the radial direction, the equations governing the heat transfer through the concrete, are given by:

$$\frac{\partial T_c}{\partial t} = \kappa_c\Big(\frac{\partial^2 T_c}{\partial r^2} + \frac{1}{r}\frac{\partial T_c}{\partial r}\Big) + \frac{H}{\rho_c C_c}, \tag{3}$$

with boundary and initial conditions

$$\frac{\partial T_c}{\partial r}(R,t) = 0$$

$$-\kappa_c \frac{\partial T_c}{\partial r}(a,t) = \frac{q}{\rho_c C_c} = -\frac{\gamma_w}{\rho_c C_c}(T_c - T_w),$$

$$T(r,0) = T_0$$

where $T_c(r,z,t)$ is the temperature of the concrete, and $H = H(t)$ is the heat production rate per unit volume associated with hydration, and $-q$ is the heat flux per unit area from the concrete into the pipe given approximately by

$$q = \gamma_w(T_c - T_w), \tag{4}$$

where γ_w is the heat transfer coefficient for (turbulent) flow through the pipe, and $T_w(z,t)$ is the temperature of the water at the $r = a$. We assume the pipe is thin and a good conductor, so the temperature difference between the inside and outside surface of the pipe is negligible. Here $\kappa_c = k_c/\rho_c C_c$, is the diffusivity of concrete, where ρ_c, C_c, k_c are the density, specific heat, and conductivity of the concrete.

The boundary conditions for (3) are Neumann and Robin, respectively. The Neumann condition is imposed on the 'exterior' of the radius of the concrete so that we did not have to consider heat flux out of the concrete cylinder nor did we have to consider heat flux of a theoretical neighbouring cylinder. The Robin condition is necessary for the flux of heat across the cement-water interface. The pipe is considered as part of the water interface because the amount of time that it takes to heat the pipe is irrelevant as the amount of time that it takes for the heat to diffuse through the concrete is extremely long.

Heat transfer along the pipe

The equation for heat transfer for the water running through the pipe is

$$\frac{\partial T_w}{\partial t} + \frac{Q}{\pi a^2}\frac{\partial T_w}{\partial z} = \frac{2q}{a\rho_w C_w},$$

where Q is the volume flux of water through the pipe, assumed constant. We have assumed here that water temperature variations across a section are negligible so that $T_w = T_w(z,t)$, where z is the distance measured along the pipe from the entry point. This becomes

$$\frac{\partial T_w}{\partial t} + \frac{Q}{\pi a^2}\frac{\partial T_w}{\partial z} = \frac{-2\gamma_w}{a\rho_w C_w}(T_c - T_w) \tag{5}$$

using (4). We need to impose the initial condition

$$T_w(z,0) = T_0,$$

and need to ensure that

$$T_w(0,t) = T_{w0},$$

where T_{w0} is the temperature of the water entering the pipe. The pipe thickness is assumed to be negligible so

$$T_c(a,z,t) = T_w(z,t). \tag{6}$$

The describing equations are now in place; the coupling between the two problems occurs at the interface between the regions and is seen in equations (4, 6).

The coupling occurs in this model because of the concrete-water interface: at the point of the interface is the boundary of the concrete, moreover, the concrete-pipe interface provides a boundary for the water pipe. The Robin boundary condition for equation (3), is considered at a, which is the length of the radius of the water pipe.

Scaling

Our primary concern is to determine the time scale t_R for significant changes in the concrete temperature, the length scale z_R for significant changes in temperature along the pipe, and the temperature scale ΔT for variations in the system. To find these we need to scale the system by writing

$$t_R = \text{the time scaling factor}$$

$$z_R = \text{the length scaling factor}$$

$$\Delta T = T_c - T_w;\ \text{the temperature scaling factor}$$

In making the following substitutions we must remember that

$R =$ a fixed value which corresponds to the radius of the concrete cylinder

$\Delta T =$ the temperature scaling factor for the concrete and the water is the same

$T^\infty =$ an arbitrary value of the temperature far away from the concrete cylinder; note that T^∞ is usually taken as ambient temperature.

The substitutions into equations (3) and (5), to non-dimensionalize our model.

$$r = R\,\tilde{r},$$

$$z = z_R\,\tilde{z},$$

$$t = t_R\,\tilde{t},$$

$$(T_c - T^\infty) = \Delta T\, \tilde{T}_c,$$
$$(T_w - T^\infty) = \Delta T\, \tilde{T}_w,$$

where the radius of the concrete sleeve R is assumed known. Since conduction through the concrete drives the process we choose

$$t_R = \frac{R^2}{\kappa_c}. \tag{7}$$

After rearrangement, and using (7), the concrete heat conservation equation (3) becomes Step No 1

$$\left[\frac{\Delta T}{t_R}\right] \frac{\partial \tilde{T}_c}{\partial \tilde{t}} = \left[\frac{\kappa_c\, \Delta T}{R^2}\right] \left(\frac{\partial^2 \tilde{T}_c}{\partial \tilde{r}^2} + \frac{1}{\tilde{r}}\, \frac{\partial \tilde{T}_c}{\partial \tilde{r}}\right) + \left[\frac{H}{\rho_c\, C_c}\right]$$

Step No 2

$$\left[\frac{\Delta T}{t_R}\, \frac{R^2}{\kappa_c\, \Delta T}\right] \frac{\partial \tilde{T}_c}{\partial \tilde{t}} = \left(\frac{\partial^2 \tilde{T}_c}{\partial \tilde{r}^2} + \frac{1}{\tilde{r}}\, \frac{\partial \tilde{T}_c}{\partial \tilde{r}}\right) + \left[\frac{H}{\rho_c\, C_c}\, \frac{R^2}{\kappa_c\, \Delta T}\right]$$

Step No 3

$$\frac{\partial \tilde{T}_c}{\partial \tilde{t}} = \left(\frac{\partial^2 \tilde{T}_c}{\partial \tilde{r}^2} + \frac{1}{\tilde{r}}\, \frac{\partial \tilde{T}_c}{\partial \tilde{r}}\right) + \left[\frac{H}{\rho_c\, C_c}\, \frac{R^2}{\kappa_c\, \Delta T}\right].$$

To properly nondimensionalize our model and determine the correct scales, it is necessary to ensure that the correct balance of terms in the defining equations is achieved. We expect the heat input due to hydration to be balanced by the conductive heat transfer through the concrete; a result that can be achieved by setting $\frac{H}{\rho_c C_c} \frac{R^2}{\kappa_c \Delta T} = 1$, which gives

$$\Delta T = \frac{H}{\kappa_c}\, \frac{R^2}{\rho_c C_c},$$

the temperature scaling that we sought. After all of the above manipulations and finding our specific dimensionless group, the dimensionless equation for equation (3) is

$$\frac{\partial \tilde{T}_c}{\partial \tilde{t}} = \left(\frac{\partial^2 \tilde{T}_c}{\partial \tilde{r}^2} + \frac{1}{\tilde{r}}\, \frac{\partial \tilde{T}_c}{\partial \tilde{r}}\right) + 1.$$

We perform the same type of calculations on the water heat conservation equation (5). After scaling as above we get Step No 1

$$\left[\frac{\Delta T}{t_R}\right] \frac{\partial \tilde{T}_w}{\partial \tilde{t}} + \left[\frac{Q}{\pi\, a^2}\, \frac{\Delta T}{z_R}\right] \frac{\partial \tilde{T}_w}{\partial \tilde{z}} = -\left[\frac{2\, \gamma_w}{a\, \rho_w\, C_w}\, \Delta T\right] (\tilde{T}_c - \tilde{T}_w)$$

Step No 2

$$\left[\frac{\Delta T}{t_R}\frac{\pi\, a^2\, z_R}{Q\,\Delta T}\right]\frac{\partial \tilde{T}_w}{\partial \tilde{t}} + \frac{\partial \tilde{T}_w}{\partial \tilde{z}} = -\left[\frac{2\,\gamma_w}{a\,\rho_w\,C_w}\,\Delta T\,\frac{\pi\, a^2\, z_R}{Q\,\Delta T}\right]\,(\tilde{T}_c - \tilde{T}_w)$$

Step No 3

$$\left[\frac{\pi\, a^2\, z_R}{t_R\, Q}\right]\frac{\partial \tilde{T}_w}{\partial \tilde{t}} + \frac{\partial \tilde{T}_w}{\partial \tilde{z}} = -\left[\frac{2\,\gamma_w}{\rho_w\,C_w}\,\frac{\pi\, a\, z_R}{Q}\right]\,(\tilde{T}_c - \tilde{T}_w),$$

and again we must choose the appropriate dimensionless group and set it equal to one. We expect the heat flow into the pipe to determine the water temperature changes along the pipe so we set

$$\eta = \Big[\frac{2\gamma_w}{\rho_w C_w}\frac{\pi a z_R}{Q}\Big] = 1.$$

From this we are able to determine the required length scale z_R as

$$z_R = \frac{\rho_w C_w}{2\pi}\frac{Q}{a\gamma_w}.$$

At this point, we will substitute t_R and z_R into η, which gives

$$\eta = \Big[\frac{\pi a^2 z_R}{t_R Q}\Big].$$

Finally, the dimensionless equation for water is:

$$\eta\frac{\partial \tilde{T}_w}{\partial \tilde{t}} + \frac{\partial \tilde{T}_w}{\partial \tilde{z}} = -(\tilde{T}_c - \tilde{T}_w).$$

Along with nondimensionalizing the equations in the model, it is necessary to nondimensionalize the initial and boundary conditions. For the initial conditions in concrete and water, respectively, we have

$$\tilde{T}_c(\tilde{r}, \tilde{t} = 0) = 0 \qquad \forall \tilde{r} \in (\epsilon, 1]; \qquad \epsilon = \frac{a}{R}$$

and

$$\tilde{T}_w(\tilde{r}, \tilde{t} = 0) = 0 \qquad \forall \tilde{r} \in [0, \epsilon]; \qquad \epsilon = \frac{a}{R}$$

$$\tilde{T}_w(\tilde{z}, \tilde{t} = 0) = 0 \qquad \forall \tilde{z} \in [0, 1].$$

For the boundary conditions for concrete we have:

$$\frac{\partial \tilde{T}_c}{\partial \tilde{r}}(\tilde{r}=1,\tilde{t})=0 \qquad \forall \tilde{t}\in[0,1]$$

$$\frac{\partial \tilde{T}_c}{\partial \tilde{r}}(\tilde{r}=\epsilon,\tilde{t})=\mu(\tilde{T}_c-\tilde{T}_w) \qquad \forall \tilde{t}\in[0,1];$$

$$\epsilon=\frac{a}{R}$$

where μ is dimensionless group given by

$$\mu=\Big[\frac{\gamma_w R}{\rho_c C_c \kappa_c}\Big].$$

For water, the boundary condition is:

$$\tilde{T}_w(\tilde{z}=0,\tilde{t})=0 \qquad \forall \tilde{t}\in[0,1].$$

The equation set is now in an appropriate form for further analysis either analytic or numeric.

Oil Detection - Analogous Heat Equation Problem

Roslyn Hickson and Melanie E. Roberts

Introduction

Due to the complexity of the electromagnetic waves in the full oil detection problem, an analogous heat flux problem is considered. For this approach, an oscillating heat source is applied at the surface of the medium, $x = 0$. As depicted in Figure 2, the medium has three distinct layers, within which the conductivity and diffusivity are constant. Layers '0' and '2' are saltwater saturated and denoted by the subscript g, whilst layer '1' refers to the 'oil' or 'gas' reservoir and is denoted by the subscript s. Throughout this work we make the approximation that the density variation between the oil/gas and the saline water is able to be neglected, and thus the ratio of conductivities of the oil/gas and the saltwater layer is equivalent to the ratio of diffusivities.

Stage one: Semi-infinite layer of constant conductivity

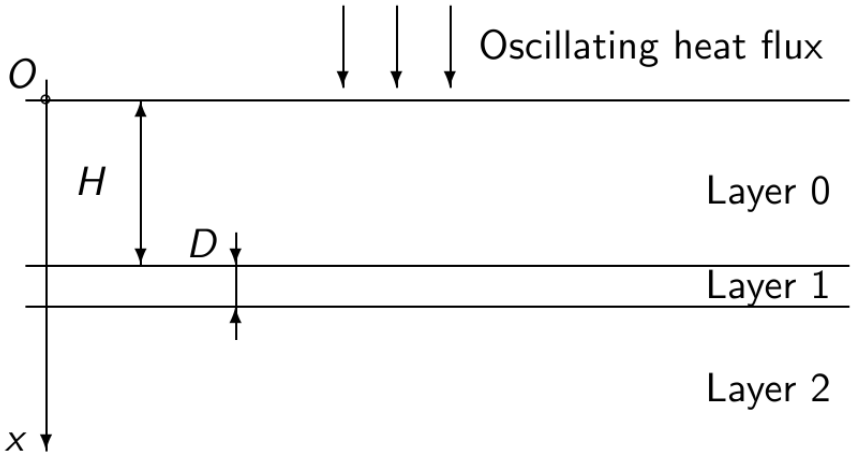

Figure 2: Schematic of the simplified heat flux approach. The heat equation is solved in the three layers and matched at the interfaces using appropriate boundary conditions.

Variations in the temperature profile are investigated at the surface by first considering a semi-infinite layer of constant conductivity and diffusivity, as shown in Figure 3. That is, the 'reservoir' or thin strip of material, layer 1, is neglected. This allows us to determine the characteristic length of penetration of the heat flux. The heat equation

$$\frac{\partial T_g}{\partial t} = \kappa_g \frac{\partial^2 T_g}{\partial x^2} \tag{8}$$

is solved subject to the surface condition of an oscillating heat flux

$$q = q_0 e^{i\omega t}, \qquad x = 0 \tag{9}$$

and the requirement that the temperature achieves the ambient value far from the surface,

$$\frac{\partial T_g}{\partial x} = 0, \qquad x \to \infty, \tag{10}$$

where the heat flux is given by $q = -k_g \, \partial T/\partial x$ with κ_g and k_g the diffusivity and conductivity of the ground respectively, $T_g(x,t)$ is the temperature, q_0 is the magnitude of the heat flux and ω its frequency, x is the vertical extent taken positive down, and t is time.

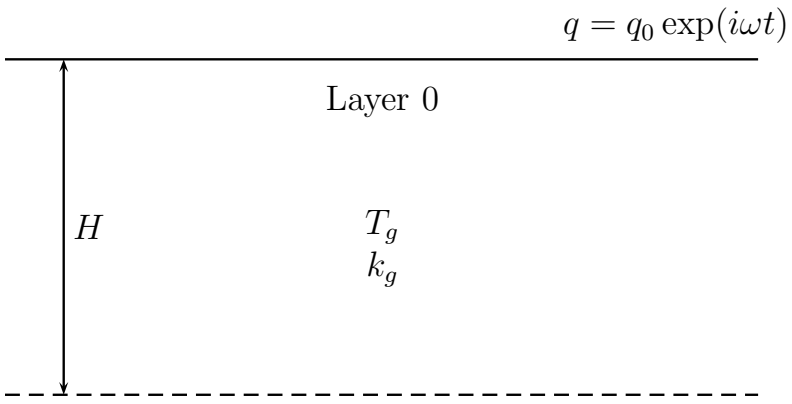

Figure 3: Schematic of the single layer problem.

The temperature profile in the domain, as illustrated in Figure 4, is an exponentially decaying wave of the form

$$T_g(x,t) = F(x)e^{i\omega t}. \tag{11}$$

While the structure of T is of little direct interest, we note the amplitude corresponding to approximately 1% of the maximum temperature variation is

$$4\sqrt{\frac{2\kappa_g}{\omega}}, \tag{12}$$

indicating that the typical depth of penetration of the heat flux is a function of the diffusivity of the ground and inversely related to the frequency of oscillation, ω. The typical detection depth is shown as a function of frequency in Figure 5. It is evident that the detection depth is greater for smaller frequencies, and thus an appropriate detection scheme is to vary the frequency slowly to determine when a variation in the temperature is first detected. This will be further investigated through the second model, which introduces the oil or gas reservoir.

Stage two: Semi-infinite reservoir layer

We investigate the variation in the temperature at the surface relative to the single layer case by introducing a semi-infinite layer of oil/gas at the

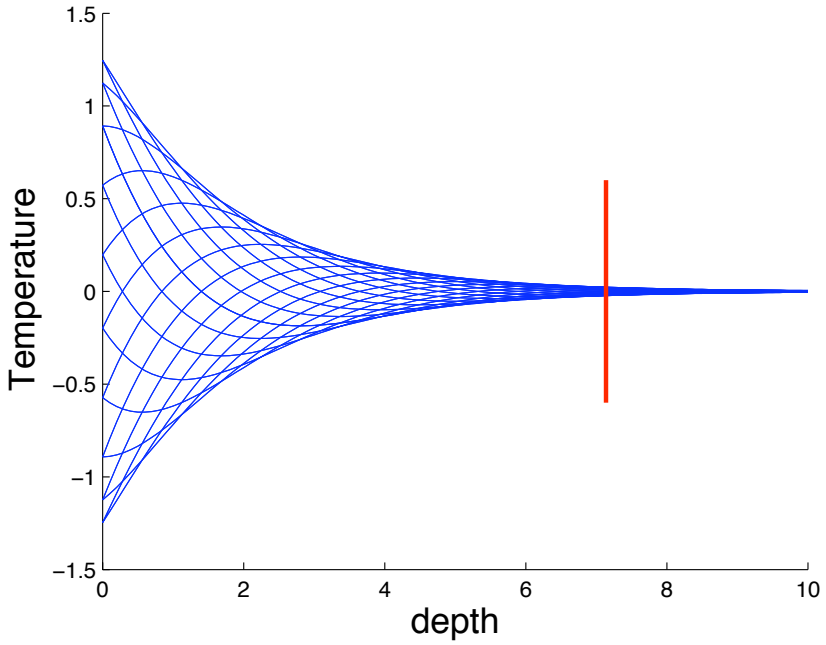

Figure 4: Temperature profiles as a function of depth at time intervals $\Delta t = 0.5$, with $k_g = 1$, and $q_0 = 1$. The typical depth of detection, corresponding to a wave amplitude of approximately $0.01 q_0$, is shown by the red line.

unknown depth $x = H$, as shown in Figure 6. This 'reservoir' layer has a significantly different conductivity and diffusivity. The problem is scaled such that the solutions are relevant to any frequency and amplitude of the heat source, and depth of the 'reservoir' layer. This also allows the solution to easily be related back to the original electromagnetic scenario.

The typical depth of detection $\hat{x} = \sqrt{\kappa_g/\omega}$ is used to scale the x-direction, the heat flux is scaled by the amplitude q_0, and time is scaled as the inverse of the frequency of oscillation, ω. Together with the surface boundary condition, (9), and setting the ratio of conductivities as equivalent to the ratio of diffusivities, this implies the appropriate scale for the temperature is $\hat{T} = q_0 \hat{x}/k_g = q_0/\sqrt{k_g \omega}$. From the temperature scale, $\hat{T}$, it is evident that the variation in temperature at the surface will be a function of both the conductivity of the ground and the frequency of the heat source.

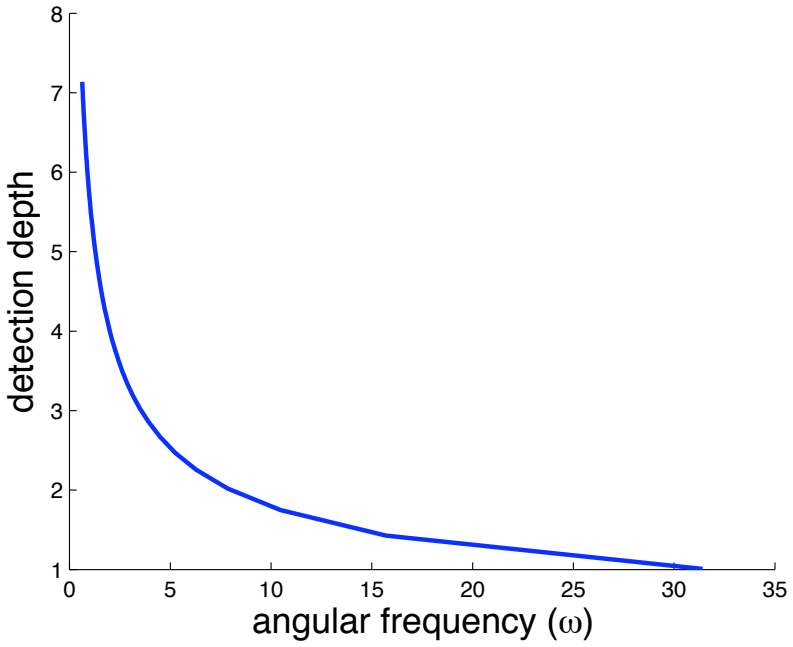

Figure 5: Typical depth of detection as a function of the angular frequency, ω, with $k_g = 1$

The scaled variables are therefore given by

$$x^\star = \frac{x}{\hat{x}}, \quad t^\star = \omega t, \quad q^\star = \frac{q}{q_0}, \quad \text{and} \quad T^\star = \frac{T}{\hat{T}}, \tag{13}$$

where the star is dropped for notational simplicity.

After this scaling, the system of equations governing this model are

$$\frac{\partial T_g}{\partial t} = \frac{\partial^2 T_g}{\partial x^2}, \qquad x < H \tag{14}$$

$$\frac{\partial T_s}{\partial t} = k_r \frac{\partial^2 T_s}{\partial x^2}, \qquad x > H \tag{15}$$

with

$$\left.\begin{array}{l} T_g = T_s \\ q_g = q_s \end{array}\right\} \qquad x = H \tag{16}$$

where $k_r = k_s/k_g$ is the ratio of the conductivities, and equations (16) fulfill the requirement of continuity of temperature and heat flux at the interface $x = H$. To recover the results obtained in Section 4.3, we simply set $kr = 1$.

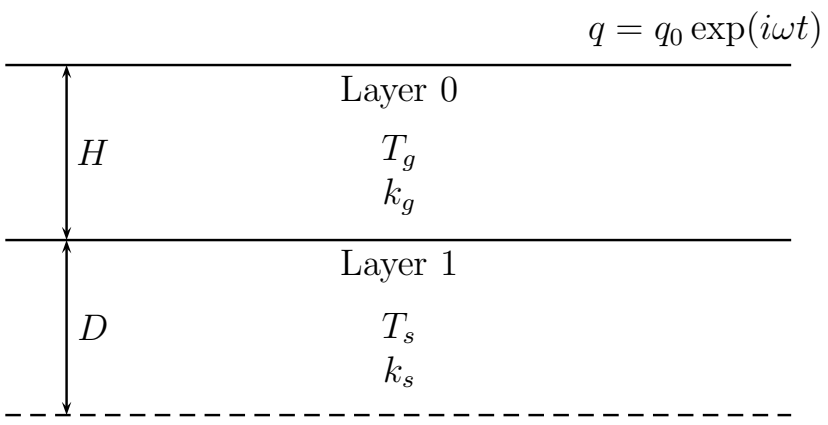

Figure 6: Schematic of the dual layer problem, the heat equation is solved for T_g and T_s in each layer with temperature and heat flux matching conditions across the interface.

The variation of temperature at the surface is then given by a function of the conductivity ratio and depth of the reservoir, as shown in Figure 7. From this figure it is evident that for $H < 1$, the temperature difference is monotonically approaching zero as the depth of the reservoir approaches the typical depth of detection ($H = 1$). Beyond this threshold, the amplitude of the temperature difference is small, and thus of little interest.

Of note is the change in sign of the temperature difference at some reservoir depth beyond the typical depth of heat source penetration ($x = 1$). Beyond this point, the depth of the reservoir is ambiguous, and is not able to be determined uniquely from a given recorded temperature variation. Therefore, a threshold on the minimum recorded temperature variation from which a reservoir depth can be determined is implicit in the model.

Stage Three: Finite depth reservoir

Finally, the thickness of the layer is incorporated into the problem, as shown in Figure 8, to determine its effect on the recorded temperature. The resulting equations are solved using the same method as that employed in Section 4.3. The variation of temperature is then expressed as a function of both the ratio of conductivities and the ratio of the reservoir thickness

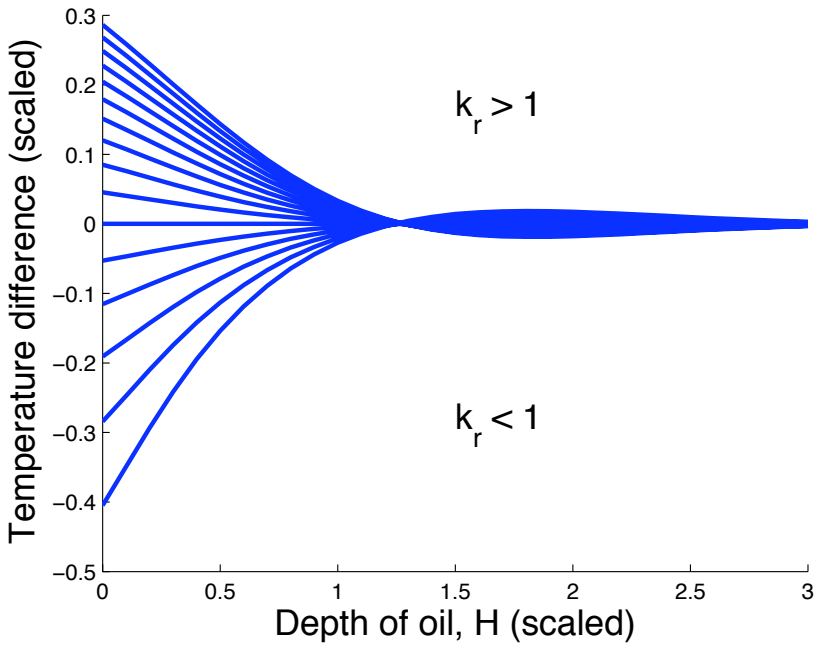

Figure 7: Observed temperature difference as a function of the depth of the oil field for various conductivity ratios.

to its depth below the surface. Rewriting the expression for the recorded temperature variation as a series expansion in the ratio of D to H, that is the thickness of the reservoir to its distance below the surface, gives, to leading order:

$$\Delta T \quad = \quad \frac{D}{2} e^{-(1+i)\sqrt{2}H}(1 \; + \; e^{it}) \; - \; \frac{D}{2k_r}(e^{-(1+i)\sqrt{2}H}(1 \; + \; e^{it})). \quad (17)$$

This indicates that the temperature difference is a function of both the thickness of the layer and the ratio of conductivities. This result is depicted in Figure 9, showing the temperature difference monotonically approaching zero as the ratio of the depth, H, to the thickness, D, increases. This indicates a significant temperature variation will not be detected where the layer is beyond the typical depth of detection, or if the layer is too thin.

Discussion

Solution of the analogous heat flux problem indicates that the signal difference recorded on the sea-floor due to the detection of a layer of oil or

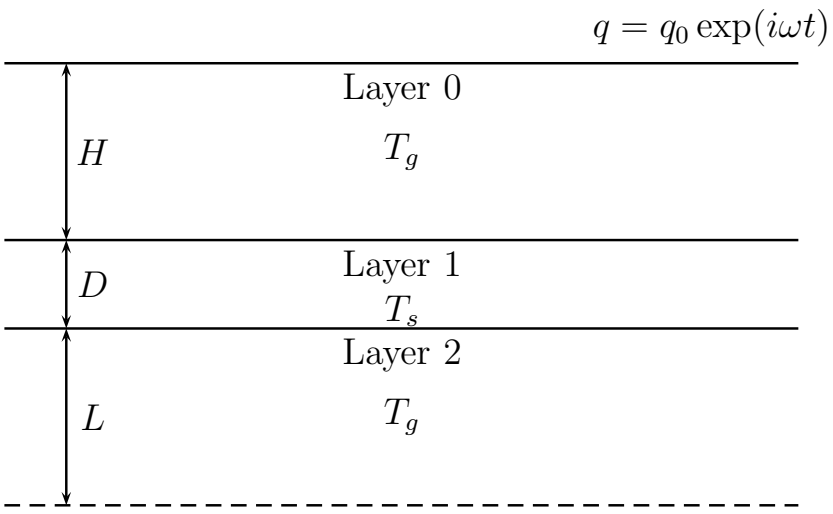

Figure 8: Schematic of the thin layer problem. The heat equation is solved in each of the three layers, with heat flux and temperature conditions across each of the interfaces.

gas is a function of both the ratios of the two conductivities $k_r = k_s/k_g$, and the ratio of the depth of the layer to its thickness. It is therefore not possible to separate the depth of the layer from the conductivities explicitly, and thus a deep layer of relatively low conductivity gradient can replicate a shallow oil or gas reservoir.

Furthermore, it is evident that the signal difference is inversely proportional to the signal frequency, thus small frequencies are able to detect layers at greater depths. This suggests that an appropriate search technique is to slowly decrease the frequency until a signal difference is identified, obtaining an indication of the depth of the oil layer.

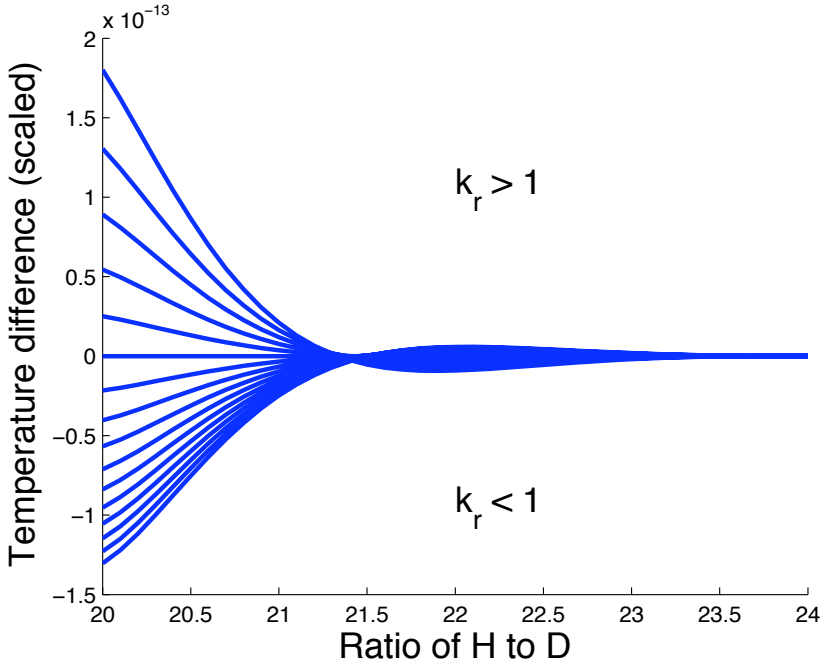

Figure 9: Observed temperature difference at the surface as a function of the ratio of the depth of the oil field and its width.

4.4 Group 4: Don Kreher, Mentor

Bounds On Mixed Covering Arrays

Taylor Barrett, Hernando Bermudez, Shaughnessy Hawkins, Sarah Plosker, Yongjun Xing, and Yuxin (Sheena) Zhang,

Abstract

In this note we discuss the lower and upper bounds on covering arrays which we discovered during the Graduate Industrial Modeling Camp held at the University of Regina as part of the 2008 PIMS industrial Problem Solving Workshop.

Introduction

Covering arrays are used to test multi-parameter software or equipment for possible abhorrent behavior. A *trial* consists of setting the parameters to specific values, running the software, and observing the results. Thus it

may be prohibitively expensive in terms of time and or money to evaluate the behavior over all possible trials. Instead we choose to find a minimal number of trials such that every combination of say t values for every subset of t parameters is tested. Usually t is small, say $t = 2$ or 3, because we believe that if there is a problem it will likely occur in the interaction of a small number of the parameters. Covering arrays may be thought of as an experiment in which the parameters are the *factors* of the experiment. The cardinality of a given factor is the *level* number for that factor and is the number of possible parameter settings for the parameter that the factor represents.

A *strength* t *Covering Array* with k *factors* and N trials on F_1,F_2,...,F_k, is a k by N array whose columns are members of $F_1 \times F_2 \times \cdots \times F_k$ having the property that for any t-tuple $R = [i_1, i_2, \ldots, i_t]^{\mathrm{T}}$ of t rows and every t-tuple $T = [x_{i_1}, x_{i_2}, \ldots, x_{i_t}]^{\mathrm{T}} \in F_{i_1} \times F_{i_2} \times \cdots \times F_{i_t}$ there is a column that contains T on the rows specified by R. The *type* of a covering array is the multi-set $\mathcal{F} = \{\ell_1, \ell_2, \ldots, \ell_k\}$, where $\ell_i = |F_i|$. We use $\mathrm{CA}(N; t, k, \mathcal{F})$ to denote a strength t covering array with k parameters, N trials, and type $\mathcal{F}$. In Figure 10 an example of a $\mathrm{CA}(24; 3, 6, \{2, 2, 2, 2, 2, 4\})$ is provided. If $n_\ell = |\{i : |F_i| = \ell\}|$ be the number of parameters that have ℓ possible values, then it is also our custom to write $\prod_\ell \ell^{n_\ell}$ for the type of the array. A covering array of type n^k is said to be *uniform* and is simply denoted by $\mathrm{CA}(N; t, k, n)$. The non-uniform covering arrays are called *mixed covering arrays* and some authors use the acronym MCA for them although it is unnecessary to do so. The covering array in Figure 10 is in this notation a $(24; 3, 6, 2^5 4^1)$. C. J. Colbourn gives extensive tables of uniform covering arrays on his web-site

`http://www.public.asu.edu/~ccolbou/src/tabby/catable.html`

A useful equivalent set-system formulation to the covering array is the *transverse covering design*. A set S is transverse to the partition

$$\mathcal{H} = \{H_1, H_2, \ldots, H_k\}$$

if $|S \cap H_i| \in \{0, 1\}$ for all $i = 1, 2, 3, \ldots, k$. A *transverse covering design* with parameters $\mathrm{TCD}(t, k, v)$ is a triple $(X, \mathcal{H}, \mathcal{B})$, where X is a v-element set of *points*, $\mathcal{H} = \{H_1, H_2, \ldots, H_k\}$ is partition of X into *holes*, and $\mathcal{B}$ is a collection of k-element subsets of X transverse to the holes called *blocks* such that every t-element subset of X transverse to the holes is contained in exactly one of the blocks. The type of a $\mathrm{TCD}(t, k, v)$ is the multiset of hole sizes and it is again our custom to write the type as $\prod_\ell \ell^{n_\ell}$ when there are n_ℓ holes of size ℓ.

Theorem 4.1 *A* $\text{TCD}(t,k,v)$ *of type* $\prod_\ell \ell^{n_\ell}$ *exists if and only if a* $\text{CA}(N;t,k,\prod_\ell \ell^{n_\ell})$ *exists.*

PROOF. Let $(X,\mathcal{H},\mathcal{B})$ be a $\text{TCD}(t,k,v)$ of type $\prod_\ell \ell^{n_\ell}$, where

$$\mathcal{H} = \{H_1, H_2, \ldots, H_k\}$$

is the set of holes and

$$\mathcal{B} = \{B_1, B_2, \ldots, B_N\}$$

be the set of blocks. For each $j = 1, 2, \ldots, N$, define the j-th trial to be $T_j = [x_1, x_2, \ldots, x_k]^{\text{T}}$, where $x_i = B_j \cap H_i$. Then it is not difficult to see that

$$[T_1, T_2, \ldots, T_N]$$

is a $\text{CA}(N;t,k,\prod_\ell \ell^{n_\ell})$.

Conversely suppose A is a $\text{CA}(N;t,k,\prod_\ell \ell^{n_\ell})$, with parameters sets $\{F_1, F_2, \ldots, F_k\}$. Then let $H_j = F_j \times \{j\}$ and set $X = \cup_{j=1}^{k} H_j$. If $T_j = [x_1, x_2, \ldots, x_k]^{\text{T}}$ is the j-column of A, i.e. the j-th trial, define $B_j = \{(x_i, i) : i = 1, 2, \ldots, k\}$ and let $\mathcal{B} = \{B_1, B_2, \ldots, B_N\}$. Then it is easy to see that $(X,\mathcal{H},\mathcal{B})$ is a $\text{TCD}(t,k,v)$ of type $\prod_\ell \ell^{n_\ell}$. ■

The covering array number $\text{CAN}(t,k,\mathcal{F})$ is the minimum number of trials N such that a $\text{CA}(N;t,k,\mathcal{F})$ exists, and a $\text{CA}(N;t,k,\mathcal{F})$ is an optimal covering array if $N = \text{CAN}(t,k,\mathcal{F})$.

0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	0	0	1	1	0	0
0	1	0	1	1	1	1	0	1	0	0	0	0	1	0	1	1	1	0	1	0	1	1	1
0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	0	1	1	0	0	1
1	1	0	1	0	0	0	0	1	0	1	1	1	1	0	1	0	0	1	1	0	1	0	0
1	0	0	1	0	1	0	1	1	0	1	0	1	0	0	1	0	1	1	0	0	1	0	1
0	0	0	0	0	0	1	1	1	1	1	1	2	2	2	2	2	2	3	3	3	3	3	3

Figure 10: A $\text{CA}(24;3,6,\{2,2,2,2,2,4\}$ or $\text{CA}(24;3,6,2^5 4^1)$

Lower and upper bounds

Let $(X,\mathcal{H},\mathcal{B})$ is a $\text{TCD}(t,k,v)$. Then the *derived transverse covering design* with respect to $x \in H_i$, is $(X \setminus \mathcal{H}_i, \mathcal{H} \setminus \{H_i\}, \mathcal{B}')$, where

$$\mathcal{B}' = \{B \setminus \{x\} : x \in B \in \mathcal{B}\}.$$

Clearly the derived transverse covering design is a $\text{TCD}(t-1,k-1,v-|H_i|)$.

Theorem 4.2 (Lower bound) $\text{CAN}(t,k,\mathcal{F}) \geq \ell \cdot \text{CAN}(t-1,k-1,\mathcal{F}\setminus\{\ell\})$, *for all* $\ell \in \mathcal{F}$.

PROOF. Suppose there is $\text{CA}(N;t,k,\mathcal{F})$ and let $\ell \in \mathcal{F}$. Then by Theorem 4.1 there exists a $\text{TCD}(t,k,v)$ with N blocks and a hole H of size ℓ. For each $x \in H$, the derived transverse covering design with respect to x is a $\text{TCD}(t-1,k-1,v-\ell)$ which has $N_x \geq \text{CAN}(t-1,k-1,\mathcal{F}\setminus\{\ell\})$, blocks, because by Theorem 4.1 it is equivalent to a $\text{CA}(N_x;t-1,k-1,\mathcal{F}\setminus\{\ell\})$. Therefore, summing over the ℓ possible values of x, we obtain the result. ■

If $\mathcal{B}$ is a collection of subsets of a set X and $x \notin X$ is a new point, then $\mathcal{B}+x$ denotes the collection of subsets

$$B + x = \{B \cup \{x\} : B \in \mathcal{B}\}.$$

Theorem 4.3 (Upper bound) *If* $\ell_2 > \ell_1$, *then*

$$\text{CAN}(t,k,\mathcal{F}\cup\{\ell_2\}) \leq \text{CAN}(t,k,\mathcal{F}\cup\{\ell_1\}) + (\ell_2-\ell_1)\text{CAN}(t-1,k,\mathcal{F}).$$

PROOF. Let $(X,\mathcal{H},\mathcal{B})$ be a $\text{TCD}(t-1,k,\mathcal{F})$ with $|\mathcal{B}| = \text{CAN}(t-1,k,\mathcal{F})$ and let $(X\cup J,\mathcal{H}\cup\{J\},\mathcal{B}')$ be a $\text{TCD}(t,k,\mathcal{F}\cup\{\ell_1\})$ in which J is a hole of size ℓ_1 and $|\mathcal{B}'| = \text{CAN}(t,k,\mathcal{F}\cup\{\ell_1\})$. Let $K\cup J$ be a hole of size ℓ_2 and set

$$\mathcal{B}^\star = \mathcal{B}' \cup \left(\bigcup_{x\in K} \mathcal{B}+x\right).$$

Note that $|K| = \ell_2 - \ell_1$. It is an easy exercise to check that

$$(X\cup K, \mathcal{H}\cup\{K\}, \mathcal{B}^\star)$$

is a $\text{TCD}(t,k,\mathcal{F}\cup\{\ell_2\})$, with

$$|\mathcal{B}^\star| = \text{CAN}(t,k,\mathcal{F}\cup\{\ell_1\}) + (\ell_2-\ell_1)\text{CAN}(t-1,k,\mathcal{F}).$$

■

Strength 2 Covering Arrays of type $2^k u^1$

In practical applications most of the parameters have only two possible values: either *on* or *off*. It is reported by Sloane in [5] that the situation for $\text{CA}(N;2,k,2^k)$s was completely solved by Rényi [4] (for N even) and independently by Katona [2] and Kleitman and Spencer [3] (for all N). They report that $N = \text{CAN}(2,k,2)$ if and only if

$$k = \binom{N-1}{\lceil \frac{N}{2} \rceil}$$

The array A is constructed by taking as the first column the zero vector and the remaining $N-1$ columns are taken to be the characteristic vectors of all the $\lceil \frac{N}{2} \rceil$ subsets of an $N-1$ set. Because N grows monotonically as a function of N it is not difficult to show that

$$N \geq 1 + \lceil \log_2 k \rceil. \tag{18}$$

In fact it can be shown that for large k, the minimal N satisfies

$$N = \log_2 k + \frac{1}{2} \log_2 \log_2 k + \cdots$$

Another interesting but non-optimal construction for a $\text{CA}(N; 2, k, 2)$ is to first construct a k by $1 + \ell$ array B by taking a column of all zeros and then filling in the rows by taking the characteristic vectors of any k of the 2^ℓ subsets of an ℓ set where $\ell = \lceil \log_2 k \rceil$. If $\overline{B}$ is the array obtained by making the bit complement of the entries in B, then it is not difficult to see that $[B, \overline{B}]$ is a $\text{CA}(N; 2, k, 2)$ with $N = 2\lceil \log_2 k \rceil + 2$. We find this array interesting because it is self complementary, i.e. the bit-complement of any column is again a column. We use the arrays A and B to establish the following theorem.

Theorem 4.4

1. *If* $u \geq \lceil \log_2 k \rceil + 1$, *then* $\text{CAN}(2, k+1, 2^k u^1) = 2u$.
2. *If* $u < \lceil \log_2 k \rceil + 1$, *then* $\text{CAN}(2, k+1, 2^k u^1) \leq N + u$, *where* N *satisfies* $k = \binom{N-1}{\lceil \frac{N}{2} \rceil}$.

PROOF. First suppose $u \geq 2\lceil \log_2 k \rceil + 1$ and let $m = \lceil \log_2 k \rceil + 1$. We use the array $[B, \overline{B}]$ to construct a $\text{CA}(2u; 2, k+1, 2^k u^1)$ as follows. First set $V_m^{\mathrm{T}} = [0, 1, 2, \ldots, m]$ and $J_k^{\mathrm{T}} = [\underbrace{1, 1, \ldots, 1}_{k \text{ times}}]^{\mathrm{T}}$. Then

$$\begin{bmatrix} B & \overline{B} & 0J_k & 0J_k & \cdots & 0J_k & J_k & J_k & \cdots & J_k \\ V_m^{\mathrm{T}} & V_m^{\mathrm{T}} & m+1 & m+2 & \cdots & u & m+1 & m+2 & \cdots & u \end{bmatrix}$$

is the required $\text{CA}(2u; 2, k+1, 2^k u^1)$. It is optimal, because a factor with u levels must occur with a binary factor in at least $2u$ columns.

If on the other hand $u < \lceil \log_2 k \rceil + 1$, then we use the array A. By Equation 18 we see that N, the number of columns of A, exceeds $\lceil \log_2 k \rceil +$

$1 > u$ and so we set $V = [1, 2, \ldots, u]$, and $U \in \{1, 2, \ldots, u\}^{N-u}$ taken arbitrarily. Now

$$\begin{bmatrix} A & \overline{A}' \\ [V, U] & V \end{bmatrix},$$

where $\overline{A}'$ is the bit complement of the first u-columns of A, is easily seen to be a $\text{CA}(N + u; 2, k + 1, 2^k, u^1)$, where N satisfies $k = \binom{N-1}{\lceil N2 \rceil}$. ■

Acknowledgments

This research is the direct result of the math camp portion of the 2008 PIMS Industrial Problem Solving Workshop and Graduate Industrial Mathematics Modeling Camp held on the campus of the University of Regina. The authors are extremely grateful of the support given by PIMS and are particularly appreciative of the hard work put in by the local organizers, Brian Alspach, Amir Amiraslani, and Shaun Fallat.

References

[1] M.A. Chateauneuf, C.J. Colbourn and D.L.Kreher, Covering arrays of strength 3, *Designs Codes and Cryptography*, **16** (1999), 235-242.

[2] G.O.H. Katona, Two applications (for search theory and truth functions) of Sperner type theorems, *Periodica Math.*, Hung. **3** (1973), 19-26.

[3] D. Klietman and J. Spencer. Families of k-independent sets, *Discrete Mathematics*, **6** (1973), 255-262.

[4] A. Rényi, *Foundations of Probability*, Wiley, New York, 1971.

[5] N. Sloane, Covering Arrays and Intersecting Codes, *Journal of Combinatorial Designs*, **1** (1993), 51-63.

4.5 Group 5: Roge Mamon, Mentor

The quantification of market risk

Chakra Bakyar, Michael Cavers, Yogesh Joshi, Manmeet Kaur, Dong Won Kim, Xiaoping Liu, Fridahus Oloude, and Yegor Sorokin.

Objectives and Motivation

We aim to quantify the total risk in a portfolio of financial assets. The

measurement of such risk, especially risk of large losses, is a central concern to corporate treasurers, fund managers and central bank regulators. Two fundamental questions that must be taken into account in market risk measurement are considered: (i) what sort of mathematical/statistical models accurately and conveniently describe the respective movements and comovement in the individual and multiple sources of risk in a given portfolio? and (ii) how does the portfolio value affected by the changes in the underlying sources of risk? Here, Monte Carlo simulation is employed in the estimation of the profit and loss distribution of a portfolio.

Certain approaches are considered in addressing the problem of estimating the probability of large losses that involve simulating rare but significant events (e.g., market crashes). A common metric in risk measurement is the so-called value-at-risk (VaR). In addition to VaR, the determination of conditional VaR (CVaR) within the framework of rare but significant events is of practical importance. CVaR refers to the expected loss during a period of several days given that we are in the left tail of the distribution.

The Problem

The members of the group took the role of employees working for the treasury department of a firm that has a portfolio of assets and investments such as stocks, options, bonds and perhaps, other derivatives. At the end of the day, they must tell the company "we are $X\%$ confident that the value of the company's portfolio will not drop by more than $\$V$ within the the next N business days. The level V is called the value-at-risk (VaR). For example, when $X = 99$, $N = 10$, it is desired to calculate V.

Typically, we wish to generate the distribution of profits or losses resulting from the market changes, say over a two-week period. VaR is demonstrated graphically in Figure 1 with $X = 99$ and $N = 10$ (the 2-week period is assumed to have 10 business days).

Why calculate VaR and how to calculate it?

Quantifying the total market risk in a portfolio is an important aspect of risk management. VaR is one particular measure in assessing the market risk. In particular, we calculate VaR because of the following important considerations:

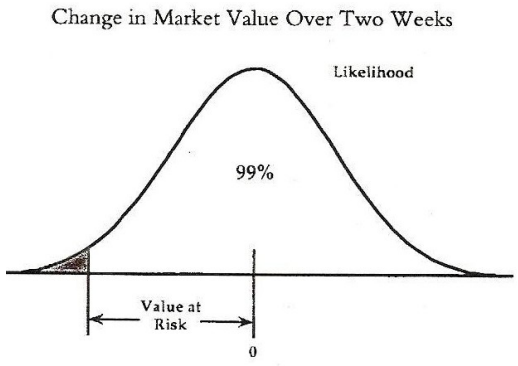

Figure 11: Illustration of the concept of VaR via the Profit and Loss distribution.

- Regulators base the capital they require banks to keep on VaR.
- The market risk capital is k times the 10-day 99% VaR where k is at least 3.0.
- VaR is the loss level that will not be exceeded with a specified probability. Thus, it captures an important aspect of risk in a single number. It is also easy to understand.
- It asks the simple question: "How bad can things get?".

We certainly do not know what will happen tomorrow. However, we can use a probability distribution of portfolio's loss or profit level. To this end, we simulate scenarios to calculate the VaR. The simulation can be done using the following techniques:

- Historical simulation
- Monte Carlo simulation (using a multivariate normal distribution)
- Jump diffusion models
- Heavy-tailed distribution or possibly mixture of normals

During the first two days of the one-week workshop, the group calculated VaR via the historical simulation method. The other techniques were considered in the remaining days of the workshop.

The company's assets and investment holdings
The group split into four different hypothetical departments, with one person designated as Risk Manager. Each Risk Manager for a particular line of business will perform the VaR calculation in his department and this will be consolidated by a designated Over-All Risk Manager. The company is assumed to have investments in the following:

Foreign Currency Options

- long position in 10,000 CAD-USD put options
- long position in 10,000 CAD-AUD put options
- short position in 10,000 CAD-GBP call options
- short position in 10,000 CAD-EUR call options

Commodity

- long position in 20,000 call options on gold price

Equity

- 100,000 shares of Microsoft
- 100,000 shares of IBM
- 100,000 shares of Bombardier
- 100,000 shares of Exxon Mobil Co
- long position in 100,000 call options in NASDAQ index

Fixed Income

- $200,000 investment in 2-year Canadian Treasury notes
- $200,000 investment in 6-month Canadian Treasury bills

Simulation experiment
The mentor provided datasets that participants could use in the simulation analysis. The datasets for the foreign exchange rates and yield rates were downloaded from the Bank of Canada website whilst the datasets for the commodity and shares prices were downloaded from Yahoo Finance website. We wish to calculate VaR for the day 04 Jan 2006 using historical simulation with data from the past five years (2000 Jan – Dec 2005). The choice for the date of 04 Jan 2006 was made in order to have an opportunity to backtest the model using the data in the past.

For historical simulation, we follow the following steps: (*i*) Create a database of the daily movements in all market variables underlying the portfolio. (*ii*) The first simulation trial assumes that the percentage changes in all market variables are as on the first day; the second simulation trial assumes that the percentage change in all market variables are as on the second day; and so on. (*iii*) Suppose we use m days of historical data (say from 5 years of trading) and let v_i be the value of a variable on day i. (*iv*) There are $m-1$ simulation trials and the ith trial assumes that the value of the market variable tomorrow (i.e., on day $m+1$) is $v_m \dfrac{v_i}{v_{i-1}}$.

For simplicity, it is assumed that roughly the portfolio changes on successive days come form independent identically distributed normal distributions. This enables us to calculate the 10-day, 99% VaR from the 1-day, 99% VaR as:

$$\text{10-day VaR} = \sqrt{10} \times \text{1-day VaR}.$$

Here, we provide samples of numerical outputs from two sub-groups of students working in the fictitious commodity and equity departments. In Figure 2, we show a snapshot of the Excel simulation spreadsheet for the Profit/Loss of a long position in 20,000 call options on a gold price. Figure 3 depicts the Profit/Loss distribution for the equity investments together with the computed value for the 1-day and 10-day VaR. In Figure 4, the Profit/Loss distribution for the equity holdings minus the option investment on NASDAQ is displayed. One may observe that without the NASDAQ, the distribution obtained is more tapered and smoother in the tails. Indeed, the presence of NASDAQ investment seems to cause pronounced distortion in the profit/loss distribution. Each group also considered back-testing. This tests how well the VaR estimates would have performed in the past. Essentially, it addresses the question "how often was the actual 1-day loss greater than the 99% per day VaR?".

Example: long position in 20,000 call options on gold price (Today: Gold Price: 597.69; C = 208.364)

	A	B	C	D	E	F	G	H	I
1	**Gold Price - Selected Currencies**								
2	**National Currency Unit per Ounce, based on London pm fix**								
3									
4	**Date**	**Canadian $**	**d1**	**d2**	**v_i/v_{i-1}**	**v_m*v_i/ v_{i-1}**	**C_i**	**P/L**	
5	04-Jan-00	408.12							
6	05-Jan-00	407.69	4.546015	4.44924179	0.99895	597.0687	210.704068	-46797.691	
7	06-Jan-00	406.89	4.536576	4.43980231	0.9980379	596.5235	210.1589022	-35894.376	
8	07-Jan-00	411.41	4.67114	4.57436633	1.0111195	604.3424	217.9777395	-192271.12	
9	10-Jan-00	409.50	4.508738	4.41196458	0.9953528	594.9187	208.5540653	-3797.638	
10	11-Jan-00	410.76	4.588509	4.49173592	1.0030664	599.5291	213.1644427	-96005.186	
11	12-Jan-00	410.96	4.56184	4.46506703	1.000481	597.9838	211.6191541	-65099.413	
12	13-Jan-00	408.68	4.499439	4.40266537	0.9944575	594.3836	208.0189314	6905.03932	
13	14-Jan-00	409.65	4.581458	4.48468427	1.0023821	599.1201	212.7554571	-87825.473	
14	17-Jan-00	414.10	4.66847	4.57169642	1.0108583	604.1863	217.8216122	-189148.58	
15	18-Jan-00	413.81	4.54963	4.45285617	0.9992994	597.2776	210.9129442	-50975.217	
16	19-Jan-00	417.84	4.657033	4.56025995	1.0097402	603.518	217.1533027	-175782.39	
17	20-Jan-00	416.19	4.515957	4.41918359	0.9960484	595.3345	208.9698242	-12112.816	
18	21-Jan-00	415.28	4.534291	4.43751734	0.9978172	596.3916	210.0270117	-33256.566	
19	24-Jan-00	413.86	4.521337	4.42456364	0.9965671	595.6445	209.2798619	-18313.569	
20	25-Jan-00	415.07	4.587049	4.49027585	1.0029247	599.4444	213.0797374	-94311.08	
21	26-Jan-00	411.70	4.47264	4.37586631	0.9918818	592.8441	206.4794391	37694.8859	
22	27-Jan-00	409.48	4.501067	4.40429399	0.9946142	594.4772	208.1126173	5031.32253	
23	28-Jan-00	413.09	4.647656	4.55088279	1.0088243	602.9705	216.6058842	-164834.02	
24	31-Jan-00	410.59	4.494009	4.39723589	0.9939351	594.0713	207.7067081	13149.5064	
25	01-Feb-00	409.44	4.528083	4.43130924	0.9972179	596.0335	209.668821	-26092.753	
26	02-Feb-00	413.55	4.659941	4.56316761	1.0100243	603.6878	217.3231469	-179179.27	

C(S,T) gives the option price under the Black-Scholes model.

$$C(S,T) = S\Phi(d_1) - Ke^{-rT}\Phi(d_2) \quad d_1 = \frac{\ln(S/K) + (r+\sigma^2/2)T}{\sigma\sqrt{T}} \quad d_2 = \frac{\ln(S/K) + (r-\sigma^2/2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}.$$

Figure 12: A spreadsheet snapshot demonstrating the historical simulation for the option investment with gold as the underlier.

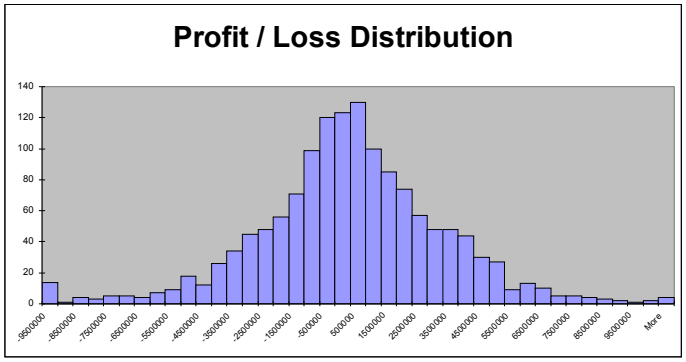

Figure 13: Profit and Loss distribution for the investments linked to equity. The 99% 1-day VaR=$9,514,977 whilst the 99% 10-day VaR=$30,088,999.

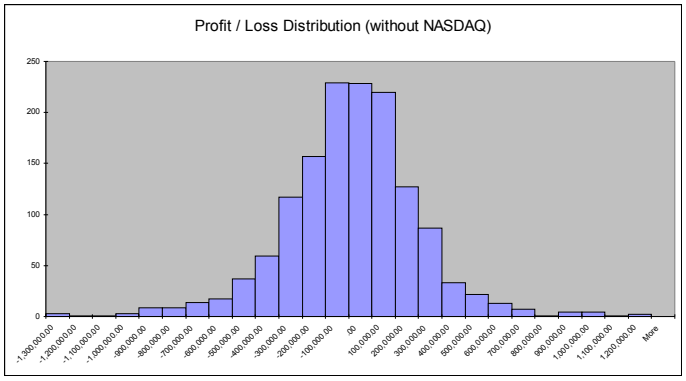

Figure 14: Profit and Loss distribution for all equity investments but without the option linked to NASDAQ. The 99% 1-day VaR=$918,472 whilst the 99% 10-day VaR=$2,904,527.

Model Building

An alternative to historical simulation is model building. Under this framework, we made the assumptions that (i) the change in the value of the portfolio is linearly dependent on percentages in market variables and (ii) the percentage changes in market variables are multivariate normally distributed.

Since not all participants had programming background but everyone had access to Microsoft Office, it was decided that the participation could be maximised if the simulation could be performed using the Data Analysis toolpak in Excel. In the simulation experiment, students typically would only generate 1000 scenarios due to the limited capability of Excel. Surprisingly, whilst this number is not large, it already gives them good results. The mentor's emphasis was on the theoretical underpinnings of the models and the algorithms useful for implementation regardless of the software that one may utilise. Since the company considered has a multi-asset portfolio, simulation from a multivariate normal distribution needs to be carried out. Students were taught on how simulate sample paths of stochastic processes by making appropriate discretisations of stochastic differential equations. The procedures on generating n correlated random samples from normal

distributions were adopted from Hull (2006).

A more realistic setting
Since we are interested in the left tail of the distribution, the VaR calculation will be more accurate if we could capture the impact of rare but extreme events. This could be addressed by involving heavy-tailed distributions. One can simulate from t-distributions, which have fatter tails than those generated by the normal distribution; see Figure 5. A heavier-tailed distribution

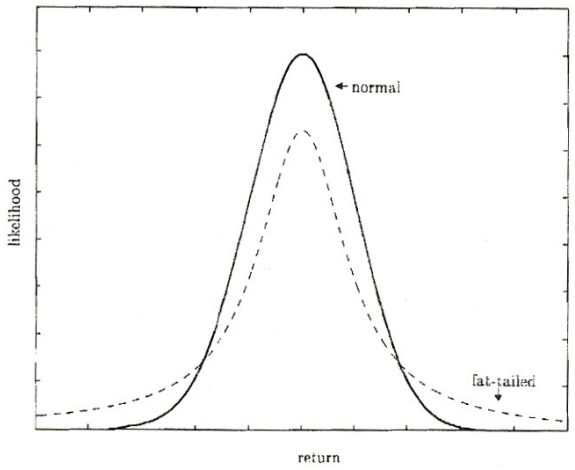

Figure 15: The t-distribution showing heavier tails than those of the normal distribution.

can also be obtained by mixing several normal distributions with different means and variances as put forward in Hamilton (1994). This preludes the use of regime-switching models such as those discussed in Buffington and Elliott (2002), Guo and Zhang (2004), and Erlwein and Mamon(2009), amongst others.

Since jumps in the level of market variables such as foreign exchange rates, stock indices, commodity prices, interest rates, etc are common, jump-diffusion models (e.g., Merton (1976)) can be employed in conjunction with Monte Carlo simulation. The jump component can be modelled by a Pois-

son process and consequently, the unconditional distribution of the market variable, such as the stock price, is a Poisson mixtrue of lognormal distributions. The discretisation and simulation of such a process can be found in Glasserman (2004).

When a portfolio contains options, the change in the portfolio value is not linearly related to the percentage changes in the market variable. An approximate quadratic relationship can be derived between the change in the portfolio value and percentage changes in the market variables. This relationship can be used to calculate the moments of the portfolio changes. The skewness of the distribution for the portfolio changes can then be incorporated. For example, the Cornish-Fisher expansion can be applied to estimate percentiles of the probability distribution from the moments.

Concluding remarks

Throughout the one-week workshop, participants learn one fundamental aspect of financial modelling, which is to quantify the market risk associated in a portfolio of assets. They were equipped with the basic theoretical background and mathematical tools on how to discretise and simulate models based on Brownian motion. When returns from a basket of investments or assets are considered and perhaps, these returns are more likely to be correlated too, participants learn how to perform simulation from a multivariate normal distribution. Finally, participants were exposed to further directions in improving VaR estimates by considering simulation from fat tailed distributions by a simple procedure. In particular, suppose one wishes to simulate a random variable of zero mean and unit variance, but with a given degree of tail-fatness (fourth moment). Sticking to the more or less bell-curved shapes for the probability density of returns (and ignoring skewness), one can utilise the idea that a random variable has fat tails if it can be expressed as a random mixture of normal random variables with different variances.

Overall, the participants were pleased to learn certain aspects of measuring risk of market fluctuations affecting the company's portfolio. In addition, from the standpoint of practical importance, they learn how to implement stochastic models using simulation-based techniques within a financial modelling context using only the spreadsheet platform, which continues to be ubiquitous in the industry sector. It has be noted that majority of the students in the group knew nothing about finance to begin with, let alone its jargons and the various financial terminology. But in the end, they were able to appreciate the concepts and pick up gradually some of the core and

related-issues that are of prime considerations in quantitative finance.

References:

Buffington, J. and Elliott, R. (2002), "Regime switching and European options", in *Stochastic Theory and Control, Proceedings of a Workshop*, Lawrence, K.S., October 2002, Springer Verlag.

Erlwein, C. and Mamon, R. (2009), "An on-line estimation scheme for a Hull-White model with HMM-driven parameters, *Statistical Methods and Applications*, 18(1), 87-107.

Glasserman, P. (2004), *Monte Carlo Methods in Financial Engineering*, Springer, New York.

Guo, X and Zhang, Q. (2004), "Closed-form solutions for perpetual American put options with regime switching", *SIAM Journal on Applied Mathematics*, 64(6), 2034-2049.

Hamilton, J. (1994), *Time Series Analysis*, Princeton University, New Jersey.

Hull, J. (2006), *Options, Futures and Other Derivatives*, Prentice-Hall, New Jersey.

Merton, R. (1976), "Option pricing when underlying stock returns are discontinuous", *Journal of Financial Economics*, 3, 125-144.

5 Industrial Problem Solving Workshop

In this section we present summaries of what was done for each of the five problems presented during the IPSW week.

5.1 An Analysis of Blackjack and Other Table Games for the Saskatchewan Gaming Corporation

Solution Team: Ha Dang, David de la Rosa, Bridget Fortowsky, Irma Elizabeth Diaz Bobadilla, Yuhui Huang, Michael J. Kozdron, Donald L. Kreher, Sarah Plosker, Yegor Sorokin, and Sheena Zhang.

5.1.1 Opening remarks

The problem under study here was proposed by Les Cloutier, the Director of Table Games Development for the Saskatchewan Gaming Corporation (SGC), who wished to develop an algorithm to help analyze the table games at Casinos Regina and Moose Jaw, enabling him to detect any anomalies that might require management attention. Ideally, the casino would like to make a profit as close to the theoretically expected house profit as possible (computed using the industry-accepted house advantage) with few large wins or large losses. Ultimately, they would like to detect such "rare" wins or "rare" losses at each gaming table on a daily basis. Blackjack was the particular table game on which this investigation focused. The primary reason for analyzing blackjack is that blackjack is the most common table game played at Casinos Regina and Moose Jaw and so blackjack is the Saskatchewan Gaming Corporation's table game with the most available data. We believe that our analysis and conclusions for blackjack can be applied with only minor modifications to the other table games such as baccarat, craps, roulette, caribbean stud poker, red dog and three card poker.

Blackjack is an inherently random game. (Indeed, this is true of all casino table games.) Calculations can determine the probability of a win or a loss, but cannot, with 100% certainty, predict the outcome of a particular game. Expected values can help determine the house advantage, but again this only provides a theoretical framework for what the house is *likely to win*. It does not offer any guarantee for what the house *will win*. Furthermore, it does not negate the possibility that the house will suffer an extremely severe loss on a given day. Conversely, however, it does not negate the possibility that the house will garner an extremely large profit on a given day. Thus,

our solution to the above problem of analyzing the table games at Casinos Regina and Moose Jaw will necessarily be probabilistic (i.e., no absolute solution is possible; there will always be randomness involved). In fact, the main conclusions that we reach are that daily large losses a! nd daily large gains are to be expected. We attempt to give a sense of how likely such "rare events" are to occur, and show that witnessing a few rare events every month or so is not at all unexpected.

5.1.2 The theoretically expected house profit for blackjack

It is well-known that the *house advantage* is the result of the casino not paying winning wagers according to the true odds of the game. The house advantage is formally defined as the casino's expected profit expressed as a percentage of the player's original bet.

Example 5.1 For instance, in double-zero (or American) roulette, there are 38 numbers (18 black, 18 red, and 2 green zeros) but payouts are based on 36 numbers. That is, if a player makes a \$1 bet on a "red" and a "red" appears, then the player receives \$2 (his original \$1 plus a profit of \$1). However, there is only a 18/38 chance of that number appearing. This means that the player's *expected profit* on such a \$1 bet is

$$\left(\frac{18}{38}\times(+1)\right)+\left(\frac{20}{38}\times(-1)\right)=\frac{18}{38}-\frac{20}{38}=-\frac{2}{38}=-0.0526=-5.26\%.$$

Thus, the house advantage is said to be 5.26%.

It is widely accepted that for a player following blackjack basic strategy[1], the house advantage is 0.55%. In other words, for every \$1000 bet at blackjack by players following basic strategy, the casino expects to earn \$5.50. Basic strategy is the strategy which maximizes the player's average gain playing one hand against a complete shoe. This calculation of house advantage is based on a number of assumptions and may not necessarily apply at every casino since different casinos all use slightly different house rules. Furthermore, the calculation is based on a complete shoe and, of course, the cards remaining in the shoe changes from hand-to-hand. In addition, not every blackjack player follows basic strategy. As a result, the casino's true (but unknown) house advantage is often much higher. Casinos Regina and Moose Jaw estimate that their true house advantage is 1% and use this as

[1] Upon request by a player, Casinos Regina and Moose Jaw will provide a wallet-sized "Blackjack Basic Strategy Card."

their figure for determining complementary compensation (a.k.a. *comps*) for their *Player's Club* rewards program.

5.1.3 Current practice

Currently, casinos use hold percentage as a management tool for determining the performance of games. *Table Games Hold Percentage* is the percentage of money that is held by the table when measured against the total money spent at the table. Formally, we have

$$\textbf{Hold percentage} = \frac{\textbf{amount of money the casino wins back}}{\textbf{amount of chips bought at the table}}.$$

The amount of chips bought at the table is also known as the *drop*. It is important to note that chips are NOT required to be used at the table at which they are bought. That is, suppose that a player enters the casino wanting to play in the poker room, but the poker room is full. While waiting, the player buys \$100 worth of chips at the blackjack table BJ01. After playing a few hands at BJ01 and winning \$30, that player learns that a spot has opened up in the poker room and decides to leave the blackjack table. The player does not need to sell back his remaining chips at BJ01, but rather can use those same chips at the poker table. Suppose that the player then loses his \$130 at poker and leaves the casino. In this example, although the casino has experienced a net profit of \$100 from this player, the corresponding hold percentage at BJ01 is negative (and equals $-30/100 = -3\%$) whereas the corresponding hold percentage for the poker room is positive (in fact, i! t is "infinite" since the player did not drop any "new" money at the poker room).

Thus, substantial deviations in hold percentage may not represent real problems for the casino (such as cheating at play, fraud by a staff member, or malfunctioning equipment), but rather the movement of players between tables. Since there is no easy way to track the movement of chips between tables, this "false drop" renders hold percentage as an inadequate measurement tool for analyzing particular table games.

5.1.4 Our measurement tool

Instead of using hold percentage, we have decided that a more reliable measure is results per table-hour. By *table-hour* (TblHr), we mean one hour of blackjack played at one table. For example, if there are two blackjack tables open for three hours each, then there are six total table-hours. Denote by

result the net profit for the blackjack tables under consideration. That is,

$$\text{result} = \text{total money in} - \text{total money out}$$

and so

$$\frac{\text{result}}{\text{total table-hours}} = \text{result/TblHr}.$$

There are three categories of blackjack tables at Casino Regina; those with a \$5 table minimum, those with a \$10 table minimum, and those with a \$25 table minimum. On any given day, there are a variable number of blackjack tables open. Hence, our four statistics of interest will be

- result/TblHr for all \$5 tables,
- result/TblHr for all \$10 tables,
- result/TblHr for all \$25 tables, and
- result/TblHr for all tables combined.

We have chosen to omit those tables with zero table-hours since this indicates that the table was not open for play.

The data provided by the Saskatchewan Gaming Corporation consisted of blackjack table data for 2004. From this data, we computed estimators for the mean of our four statistics of interest as shown in Table 4 below. We have also included the sample standard deviations (SD), as well as the sample range. Notice that the sample sizes near 365 indicate that blackjack tables were open for at least one hour on most of the days of the year. The smaller sample size for the \$25 tables indicates that there were two days in 2004 on which no \$25 tables were open.

It should be noted that we followed standard statistical procedure and trimmed the most extreme 0.5% of our data to eliminate outliers caused by potentially unexplainable events such as clerical errors in data entry.

Table	Sample Mean	Sample SD	Sample Range	Sample Size
\$5	118.19	198.66	(−431.28 , 593.48)	362
\$10	397.35	728.52	(−1876.18 , 2352.52)	362
\$25	165.85	528.880	(−1312.00 , 1848.00)	360
Overall	683.34	915.47	(−2120.21 , 3073.46)	362

Table 4: Sample statistics for result/TblHr based on 2004 data

It is reasonable to approximate the daily result/TblHr distribution by a normal distribution. Formally, if we let Y denote a daily result/TblHR, then Y has a normal $N(\mu, \sigma)$. As neither the true mean μ, nor the true standard deviation σ is unknown, we approximate this from the data. In particular, our data yields

- \$5 tables: Y is approximately $N(118.19, 198.66)$,
- \$10 tables: Y is approximately $N(397.35, 728.52)$,
- \$25 tables: Y is approximately $N(165.85, 528.880)$, and
- overall: Y is approximately $N(683.34, 915.47)$.

Having found the distribution of daily result/TblHr to be approximately normal, it is now possible to use a normal model to determine *prediction intervals* of non-rare wins or losses for daily result/TblHr (categorized by table grouping). These prediction intervals are shown in Table 5 below.

For example, if Y denotes the overall daily result/TblHr, then we are 98% confident that Y falls between $-\$1446.37$ and \$2813.04. If Y is outside of this interval, then it is categorized as a rare event that may require further review by management.

Table	92%	96%	98%
\$5	(−229.60 , 465.97)	(−289.80 , 526.17)	(−343.95 , 580.33)
\$10	(−878.05 , 1672.75)	(−1098.84 , 1893.54)	(−1297.43 , 2092.13)
\$25	(−760.06 , 1091.77)	(−920.34 , 1252.05)	(−1064.52 , 1396.22)
Overall	(−919.37 , 2286.04)	(−1196.81 , 2563.49)	(−1446.37 , 2813.04)

Table 5: Prediction intervals of non-rare events

We end this section by reviewing how to compute prediction intervals. Suppose that we label the data as $y_1, \ldots, y_n$. The sample mean is given by

$$\overline{y} = \frac{y_1 + \cdots + y_n}{n} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

and the sample standard deviation is given by

$$s = \sqrt{\frac{(y_1 - \overline{y})^2 + \cdots + (y_n - \overline{y})^2}{n-1}} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (y_i - \overline{y})^2}.$$

A level C prediction interval is given by

$$\overline{y} \pm z^* s\sqrt{1+(1/n)} = \left(\overline{y} - z^* s\sqrt{1+(1/n)}\,,\, \overline{y} + z^* s\sqrt{1+(1/n)}\right)$$

where the critical value z^* is chosen so that there is area $(1-C)/2$ in each tail of the normal density curve.

Table 6 below shows the value of value of the critical value z^* corresponding to the levels 92%, 96%, and 98%.

Prediction level C	value of z^*
92%	1.751
96%	2.054
98%	2.326

Table 6: Critical values for various prediction levels

For example, in the case of the \$5 tables, the sample mean is $\overline{y} = 118.19$ and the sample standard deviation is $s = 198.66$. Since there are $n = 362$ sample points, a 92% prediction interval is given by

$$\begin{aligned}\overline{y} \pm z^* s\sqrt{1+(1/n)} &= 118.19 \pm (1.751)(198.66)\sqrt{1+(1/362)} \\ &= (118.19 - 348.33, 118.19 + 348.33) \\ &= (-230.14, 466.52).\end{aligned}$$

The reason for the slight discrepancy between this result and the result presented in the first row, first cell of Table 5 is accounted for by the fact that in this explicit computation we have rounded our sample mean and sample standard deviation to two decimal places and our critical value has been rounded to three decimal places. The computer software used to construct Table 5 had much more accurate precision.

5.1.5 Are rare events really that rare?

It is seemingly paradoxical that rare random events actually happen since, by their very nature, rare events are "rare." However, as we will attempt to explain below, just because an event has a very small probability of occurring at a *given fixed time*, that does not prevent that event from happening at *over some period time*. Of course, in the present context, a rare event corresponds to either a big win or a big loss for the casino.

For each of the following charts, the first column denotes the "probability of a rare event on a particular day," and the remaining three columns denote

the "probability of at least k rare events in N days" for $N = 30$, 90, and 365.

Prob of a rare event on a given day	Prob of at least 1 rare event in		
	$N = 30$ days	$N = 90$ days	$N = 365$ days
$p = 0.08$	0.9180	0.9994	1
$p = 0.04$	0.7061	0.9746	1
$p = 0.02$	0.4545	0.8377	0.9994
$p = 0.01$	0.2603	0.5953	0.9745

Prob of a rare event on a given day	Prob of at least 2 rare events in		
	$N = 30$ days	$N = 90$ days	$N = 365$ days
$p = 0.08$	0.7042	0.9951	1
$p = 0.04$	0.3388	0.8795	1
$p = 0.02$	0.1205	0.5396	0.9947
$p = 0.01$	0.0361	0.2273	0.8804

Prob of a rare event on a given day	Prob of at least 3 rare events in		
	$N = 30$ days	$N = 90$ days	$N = 365$ days
$p = 0.08$	0.4346	0.9785	1
$p = 0.04$	0.1169	0.7030	1
$p = 0.02$	0.0217	0.2688	0.9773
$p = 0.01$	0.0033	0.0619	0.7074

Prob of a rare event on a given day	Prob of at least 4 rare events in	
	$N = 90$ days	$N = 365$ days
$p = 0.08$	0.9359	1
$p = 0.04$	0.4874	1
$p = 0.02$	0.1067	0.9345
$p = 0.01$	0.0129	0.4960

Prob of a rare event on a given day	Prob of at least 5 rare events in	
	$N = 90$ days	$N = 365$ days
$p = 0.08$	0.8555	1
$p = 0.04$	0.2920	0.9990
$p = 0.02$	0.0348	0.8553
$p = 0.01$	0.0022	0.3028

The calculations in the previous tables are particular cases of the following general result. If a rare event occurs with probability p, then the probability of no rare events in N days is

$$\Pr\{\text{no rare events in } N \text{ days}\} = (1-p)^N.$$

That is, on a given day, no rare event occurs with probability $1-p$, and so the only way for no rare events to occur in N days is for no rare event to occur on any single day. By the multiplication principle, this happens with probability

$$(1-p) \times (1-p) \times \cdots \times (1-p) \quad (N \text{ times}).$$

Hence, the probability of at least 1 rare event in N days is

$$\Pr\{\text{at least 1 rare event in } N \text{ days}\} = 1 - \Pr\{\text{no rare events in } N \text{ days}\} = 1-(1-p)^N.$$

In fact, this formula can be generalized. For $k = 0, 1, 2, \ldots$, as a result of the binomial theorem, the probability of exactly k rare events in N days is

$$\Pr\{\text{exactly } k \text{ rare events in } N \text{ days}\} = \frac{N!}{k!(N-k)!} p^k (1-p)^{N-k}$$

where $N! = N \times (N-1) \times \cdots \times 2 \times 1$ (and $0! = 1$ by convention). Thus, the probability of at least $k+1$ rare events in N days is

$$\begin{aligned}
&\Pr\{\text{at least } k+1 \text{ rare event in } N \text{ days}\} \\
&\quad = 1 - \Pr\{\text{no rare events in } N \text{ days}\} - \Pr\{\text{exactly 1}\} - \Pr\{\text{exactly 2}\} - \cdots \\
&\quad - \Pr\{\text{exactly } k\} \\
&\quad = 1 - (1-p)^N - Np(1-p)^{N-1} - \frac{N!}{2!(N-2)!} p^2 (1-p)^{N-2} - \cdots \\
&\quad - \frac{N!}{k!(N-k)!} p^k (1-p)^{N-k} \\
&\quad = 1 - \sum_{j=0}^{k} \frac{N!}{j!(N-j)!} p^j (1-p)^{N-j}.
\end{aligned}$$

This formula can easily be implemented by computer software to generate more extensive tables than the ones presented here.

5.1.6 Combining rare events and our measurement tool

In Table 5 we computed prediction intervals an in the previous section we explained that rare events will happen (with alarming frequency). For instance, the 92% prediction interval for all tables combined is

$$(-919.37\,,\,2286.04).$$

Thus, if we define a "rare event" as having an overall result/TblHr loss of at least \$919, then this is a rare event that happens with probability $p = 0.04$. (Recall that a 92% prediction interval corresponds to a 4% chance of a large profit and a 4% chance of a large loss.) As shown in the previous charts, we see that there is a probability of 0.292 of seeing at least 5 such losses in a given month, and a 0.9990 chance of seeing at least 5 such losses in a given year.

In other words, the casino must realize that simply by chance occurrence there will be days when it realizes a large loss. Conversely, however, there is exactly the same chance that the casino will see 5 days per year for which there is a result/TblHr of at least \$2286 (profit).

5.1.7 Other concerns

A number of other issues were raised by the SGC Director of Table Games Development who initially thought that, perhaps, an analysis of the blackjack table data might provide clues as to other, more serious issues including money laundering, player theft, dealer theft, and collusion. Unfortunately, the available data is aggregate daily data. As such, it is not possible to use this "global" table to detect "local" phenomenon, and so we cannot deal with these issues directly using the data provided. Our best advice is to suggest that, if any of the above are suspected, the casino send additional staff to the floor for observation.

5.2 Supply Chain Optimization for the North American Operations of the Evraz Group

Solution Team: Hernando Bermudez, Mike Cavers, Yogesh Joshi, Manmeet Kaur, Zanin Kavazovic, Dong Won Kim, Nathan Krislock, Asef Nazari, Fridahus Oloude, and Rachel Robertson.

5.2.1 Introduction

The Evraz group makes a monthly scrap metal purchase consisting of ten main grades of scrap that are sold by a number of suppliers. Each month, it is determined how much of each grade of scrap will be needed for the next month. Then the company must buy what it needs. However, each supplier has a limited quantity of certain grades of scrap to sell. The suppliers are located throughout Canada and the United States. The price of scrap varies between suppliers, as does the transportation costs to get the scrap to the plant. There are some contracts and gentlemen's agreements already in place which determine the purchases of a certain amount of scrap that the company must buy, but there is also a certain amount of scrap each month that can come from any of the suppliers. It is also important that the company spreads around their purchases in order to maintain good business contacts with many suppliers.

Evraz's plant in Regina, SK, needs approximately 35–40 million dollars worth of scrap each month to keep running. Scrap costs about $450 per ton and transportation costs are somewhere between $20–45. These are the two main costs associated with this process. There is an inventory cost as well. Evraz likes to keep about three weeks worth of scrap at their facility, however, this cost is hard to calculate, and the shipments do not necessarily arrive when scheduled.

The main consideration is to buy enough of each grade of scrap to keep the plant running at the lowest cost, while still satisfying contracts that are already in place. Evraz would also like a spreadsheet tool that they can use monthly to determine the cheapest way to purchase what they need.

The remainder of the report is as follows: The problem is stated in Subsection 2; the model is described in Subsection 3; we present the results in Subsection 4 and finally in Subsection 5 we conclude and discuss of some future work.

5.2.2 Problem Statement

Evraz Regina Steel manufactures approximately 100 kms. of pipe and 3,000 tons of steel on a daily basis. To keep this monster operation flowing, its Supply Chain/Logistics Team procures over 80,000 tons of scrap metal every month. To avoid both high scrap inventories and the risk of running short on supply, the monthly scrap buy is a balancing act between cost and delivery times.

Analyze all relevant supply chain variables such as freight cost, scrap costs and transit times. Evraz wanted us to optimize their monthly scrap purchase as a function of the price differential and associated transportation costs. They wanted a spreadsheet-based tool that takes all supply chain variables into account—such as costs and transit—to optimize their monthly scrap buy. This tool, combined with experience and market trends, will be used to efficiently plan the scrap purchasing.

5.2.3 Model

Based on the information we have from the problem description and the company, in the simplest case we are dealing with a transportation problem. There is a certain number of providers and certain amount of scrap they can provide in different types. Actually, the amount of different types of scraps from each provider would change by time or season, but at the first stage we think they are fixed. If i ($i = 1, 2, \ldots, n$) is for cities and j ($j = 1, 2, \ldots, m$) for types of the scrap, we have the following information:

- n number of providers,
- m number of types of scrap,
- x_{ij} the decision variable that the company is going to buy that amount of scrap type j from provider i,
- c_{ij} the total cost to bring scrap type j from provider i. This cost includes purchase cost, transportation, etc.,
- m_{ij} the least amount of scrap type j to buy from provider i,
- M_{ij} the maximum amount of scrap type j available from provider i.

Using this information, one mathematical model for this problem is:

$$\min \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} x_{ij}$$

subject to

$$0 \leq m_{ij} \leq x_{ij} \leq M_{ij} \quad i = 1, \ldots, n \quad j = 1, \ldots, m$$

As mentioned, this is the simplest possible case and we can solve this problem by deciding about each scrap grade independently. However, if there are some more constraints, for example, if there is a limitation in company's budget, this model is not valid anymore. If we want to extend this model, we would account for some more consideration of the company, and some more stochastic nature of scraps and providers.

5.2.4 Results

The primary goal of this project is to develop an Excel document which will allow EVRAZ to compute the minimum cost monthly scrap metal purchase under some given constraints easily and quickly. Since the overall minimum cost can be obtained by finding the minimum cost to purchase the necessary amount of each grade separately, we created individual spreadsheets to optimize the purchase of each grade. We also included a summary sheet which allows easy entry of the required amount of each grade of scrap metal and displays the minimum cost purchase computed by the optimization solver from the individual grade spreadsheets.

Each spreadsheet lists all the suppliers from which the grade of metal may be purchased. The variables over which we optimize are the number of rail cars of scrap metal to order from each supplier. It is convenient to use the number of rail cars ordered as the variables since this is typically how the order is placed. Furthermore, we now have the option to insist that these variables are integer if this becomes an important factor in the future. Given the average amount in tons that a rail car holds, we can now compute the total amount of scrap metal to order from each supplier. This total amount is then compared to both the minimum amount of scrap EVRAZ has agreed to purchase from the supplier, and the maximum amount of scrap available from that supplier. Finally, we compute the cost based on the freight price and the scrap metal price provided by the supplier. Summing this over all the suppliers gives us our total cost, which is the amount we would like to minimize.

This mathematical model is a linear optimization problem and can be efficiently solved by the Tools→Solver command in Excel. In the solver, we need to specify the lower/upper bound constraints on the amount of scrap metal to order, and the demand constraint on the total amount of scrap to

order. Choosing the option in the solver to assume a linear model, we let the solver minimize the total cost by varying the number of cars of scrap metal to order from each supplier.

In addition to being a useful tool for EVRAZ to compute the minimum cost scrap metal purchase quickly and to investigate different scenarios with ease, this Excel document has also been seen to compute purchase plans which provide significant savings over historical purchase plans. We compared the purchase plans for a past month and found that the savings obtained were roughly 2%; since this purchase plan may need to be slightly modified, we expect the savings to be on the order of 1%. However, given the monthly cost of scrap metal, even a 1% savings is of great benefit for EVRAZ. Given the many advantages mentioned, we feel this Excel document will greatly assist EVRAZ in making informed business decisions and enhance their scrap metal purchasing techniques.

5.2.5 Conclusion/Future work

The model presented is a simple linear program that minimizes the total cost of EVRAZ's monthly scrap purchase by taking both scrap and freight costs into account. However, there are many factors that the current model does not consider. We feel it would be beneficial to the company to look at similar models to optimize other facets of the operation while at the same time keeping costs at a minimum.

Currently, the linear program presented is to be solved each month in Excel by using the scrap prices for each supplier for that month. In the future, we would like to add multiple time periods to the model. This will allow us to incorporate future prices and demands into the model to further reduce the total long term costs the company incurs. However, by doing this we must also incorporate inventory capacities and inventory costs into the model. This is because it may be in the company's best interest to order more scrap during the current month for future use, so long as the cost to store the scrap is not too high.

Since future prices and demands are generally unknown, we may consider the problem as having a stochastic element present in the data. Thus, it may be best to consider the problem as a stochastic program, from which can obtain a solution that is optimal over a set of scenarios.

Also, by adding multiple time periods we can account for the seasonal nature of suppliers, and further, ensure the proper amount of stock is in inventory so that the company can continue its operations even if a shipment is missed.

Finally, the reliability of the current model needs to be checked. Some tweaking may be necessary by the company since the optimal solution provided by the program may not be realistic.

5.3 Plow, Plow, Plow the Field with Accutrak

Solution Team: Ewout van den Berg, David Clark, Melanie E. Roberts, Edward Doolittle, Roslyn Hickson, Norberto Flores Guzmán, Javier Flavio Vigueras.

5.3.1 Introduction

Our problem was presented by Dr. Ron Palmer of Accutrak. Accutrak is a Regina-based company that produces autosteering machines for farms, which assist farmers by automatically steering tractors within a field. The system reduces overlapping of plow lanes and optimizes plow routes, which reduce the resource consumption and distance traveled by each tractor.

To identify a tractor's position, the Accutrak autosteering systems use radio beacons mounted around the perimeter of each field. These beacons communicate with a receiver on the tractor, using wireless communication which is subject to multipath errors. All beacons transmit at the same frequency (usually 300 MHz) using time-division multiplexing.

The position of a tractor within a field is calculated using the approximate distance to each beacon from the tractor. This distance is expressed as a number of complete wavelengths, plus a fractional number of wavelengths, between the beacon and the tractor.

For example, suppose we have a wavelength of 0.50 meters. If the signal from a beacon requires 6 complete wavelengths plus 0.20 wavelengths to travel to the tractor, then the tractor is $0.50 \times 6.2 = 3.1$ meters away from the beacon. However, the radio equipment in use reports only the fractional part of the number of wavelengths. In the example above, we only know that the beacon is $n.2$ wavelengths away, for some integer n. The major part of this problem is to identify the value of n for each beacon. Typically, 10 such "fractional" readings per second are taken from each beacon.

If the integer and fractional parts of the number of wavelengths to each beacon are known, then the position of the tractor can be calculated using a linear least-squares fit. Once this is known, existing software uses further readings to keep track of the tractor's position to a high degree of accuracy.

Unfortunately, there are several limitations and sources of error.

- Each beacon has an unknown amount of systemic bias, resulting in an incorrect (but consistent) error in the reading of the fractional part of the wavelength. This is caused by various parts of the system, such as the antenna length, cables, and (to a certain extent) the terrain in

the area. Experimentally, this error is at most 1/8 of the wavelength, or about 12.5%.

- The positions of the beacons are known to within approximately 10 cm.
- Unpredictable errors in the fractional portion of the wavelength may occur as a result of atmospheric interference and similar effects. These are typically transitory.
- The fractional portion of the wavelengths are quantized at the receiver into 256 levels. This introduces measurement errors and effectively limits the accuracy of all calculations. However, this level of quantization is changeable with some effort.
- Accutrak wishes to minimize the amount of time and effort necessary to initially determine the values of n for each beacon. It also prefers to avoid any special driving patterns or detailed instructions that drivers must follow to properly calibrate the system. In the future, this system may be used on unmanned tractors or mowers. Thus, determining each n should be entirely automatic and require at most a few minutes of time.
- Accutrak does not wish to add any complexity to the radio hardware used. Computational resources are available, although Accutrak would prefer that a solution be implementable in integer arithmetic.

Given this information, we have identified the following problem to solve.

Problem statement: Our problem is to identify the integer number of wavelengths between each beacon and the tractor, which is located at an unknown starting position. We must account for several sources of error and respect the limitations put forth above. From this information, the position of the tractor can be calculated separately.

Accutrak also has requested that we develop a method to double-check an existing set of integer wavelength values and ensure that they correspond to the tractor's expected location within a field.

We make certain assumptions about the situation:

- The systemic errors in measuring fractional wavelengths are constant. This is not strictly true, as they may vary depending on changes in

topography (for example, readings taken at the top of a hill may produce slightly different errors than readings taken at lower elevations). However, these changes are very small and occur very slowly.

- The initial position of the tractor can be determined to within a 10 meter radius using GPS or other methods. This radius is not essential; any level of precision helps as long as it is known.

- We ignore transient errors in the fractional wavelengths. To a certain extent, the quantization of readings into 256 levels already simulates this sort of noise.

These assumptions are generally acceptable, and are not significant barriers to real-world implementation of our solutions.

The problem was approached in several ways that are described in subsequent sections. The solution which is most useful from an implementation point of view is the nonlinear programming approach which is described next.

5.3.2 Nonlinear Programming

The technology currently in use tracks *changes* in the fractional wavelengths very accurately. However, the fractional values themselves may not be absolutely correct. This model primarily uses the change between subsequent measurements to identify the integer wavelength values.

The following concept is central to our model. Suppose we take a reading from beacon i located at (x_i, y_i) and obtain the fractional wavelength f_i. If the position of the beacon and the value of f_i are exactly correct, then the tractor must lie on a circle centered at the beacon, with radius $\lambda(n_i + f_i)$ for some nonnegative integer n_i (see Figure 16), and where λ is the wavelength used by the beacons.

However, there is some error a_i in the measurement of x_i, and some error b_i in the measurement of y_i. Thus in reality, the position (x, y) of the tractor satisfies

$$(x - x_i - a_i)^2 + (y - y_i - b_i)^2 - \lambda^2(n_i + f_i)^2 = 0$$

for some a_i, b_i, and n_i. This process can be repeated for each beacon, producing a set of k circles, one for each beacon. Note that because the beacons use time-division multiplexing, the readings from all beacons arrive essentially simultaneously. Details of the calculation of f_i are discussed later in this section.

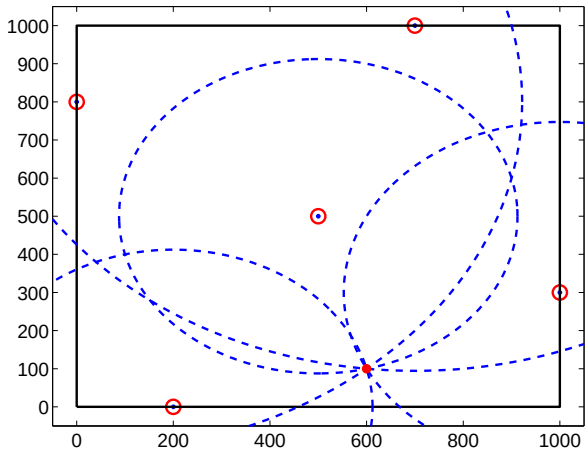

Figure 16: Schematic depicting the problem setup with beacons (red and blue dots) and a tractor (red dot). The dotted blue circles indicate all positions within a certain radius of each beacon.

If we take enough readings, several from each beacon, we will be able to solve these equations for x, y, a_i, b_i, and n_i. In order to avoid creating too many more unknowns, we proceed as follows. We instruct the tractor operator to move in a straight line, at a constant speed, for a short period of time. Typically this is only a few seconds, as we need only to collect about 6 readings. Movement in a straight line at a constant speed is made possible by existing Accutrak technology. Recall that readings from each beacon happen at regular intervals, approximately every 1/10 of a second. Thus after each reading, the tractor has moved the same distance in the same direction. We denote the distance traveled d. The x and y components of this movement are Δx and Δy respectively. Thus on the second reading, we have moved from (x, y) to $(x + \Delta x, y + \Delta y)$, and so on. This allows us to write additional equations while introducing only two new unknowns.

We add one additional equation:

$$(\Delta x)^2 + (\Delta y)^2 - (\alpha\lambda)^2 = 0,$$

where λ is the wavelength used by the beacons and $0 < \alpha < 1/2$. The term $\alpha\lambda$ represents the distance traveled between two readings. This indicates that the rate of movement of the tractor is slow enough that between readings, we have moved less than half of a wavelength. This is important for determining our direction. Note that at the usual wavelength of 1 meter,

a tractor traveling at 8 km/h easily satisfies this constraint.

With k beacons and r measurements over time, we have a total of $4 + 3k$ unknowns and $1 + kr$ equations. We will have more equations than variables as long as $r \geq 4$ and $k \geq 3$. In an ideal world, this set of equations would be consistent. Unfortunately, the various systemic errors preclude this. However, it is still possible to find a least-squares solution to this nonlinear system.

Here is a summary of the variables involved in our model. All distances are measured in meters unless otherwise specified.

Known	
(x_0, y_0)	Estimated initial position of the tractor (± 10 m)
(x_i, y_i)	Position of beacon i (± 10 cm)
k	Number of beacons
λ	Wavelength used by beacons
r	Number of measurements taken during calibration
d	Distance traveled during calibration
$f'_{i,j}$	Fractional number of wavelengths from tractor to beacon i during measurement j. Assumed to include systemic bias.
$f_{i,j}$	Total of all fractional changes from $f'_{i,1}$ up to $f'_{i,j-1}$ (see below).

Unknown	
(x, y)	Exact initial position of the tractor.
(a_i, b_i)	Error in position of beacon i.
$(\Delta x, \Delta y)$	Change in position after one measurement.
n_i	Integer number of wavelengths from tractor to beacon i at the beginning of the measurement.

We first compute the $f_{i,j}$ from the $f'_{i,j}$. For beacon i, we have a sequence of fractional values $f'_{i,j}$, $j = 1, \ldots, r$. So that the value of n_i need not be changed if we "roll over" from n_i to $n_i \pm 1$, we calculate a sequence $f_{i,j}$, in which $f_{i,j}$ represents the sum of all fractional changes from $f'_{i,1}$ up to $f'_{i,j-1}$. There are several minor issues here. For example, suppose $f'_{i,j} = 0.8\lambda$ and $f'_{i,j+1} = 0.1\lambda$. It appears that our change in fractional values was -0.7λ. However, our assumption about the speed of the tractor indicates that we could not move more that half of a wavelength between readings, and so we must have moved from $(n_i + 0.8)\lambda$ to $((n_i + 1) + 0.1)\lambda$. Thus, the actual change in fractional values was $+0.3\lambda$. This effectively tells us which "direction" we are moving compared to beacon i.

Our problem can be formulated as this nonlinear program:

$$\text{minimize } \|f\|$$

where f is a vector consisting of these functions:

$$(x - x_i - a_i + (j-1)\Delta x)^2 + (y - y_i - b_i + (j-1)\Delta y)^2 - \lambda^2(n_i + f_{i,j})^2$$

for $i = 1, \ldots, k$, and $j = 1, \ldots, r$

$$(\Delta x)^2 + (\Delta y)^2 - (\alpha\lambda)^2$$

Subject to:

$$\begin{array}{rcll} x_0 - 10 \leq & x & \leq x_0 + 10 & \\ y_0 - 10 \leq & y & \leq y_0 + 10 & \\ -0.1 \leq & a_i & \leq 0.1 & i = 1, 2, \ldots, k \\ -0.1 \leq & b_i & \leq 0.1 & i = 1, 2, \ldots, k \\ & n_i & & \text{bounded based on initial box, } i = 1, 2, \ldots, k. \end{array}$$

This is a nonlinear program with bounded constraints. Because of the errors in measurement, there is not likely to be an exact solution. However, there will be a least-squares solution.

A solution to the above linear program gives the following information:

- The n_i value for each beacon, as requested.
- The position of the tractor, which may be more useful.
- The direction of the tractor's travel.
- An approximation of the measurement error for each beacon.

As previously mentioned, Accutrak has one additional problem: to be able to validate a set of n values, given an expected physical location. This check is relatively simple, and is implemented as follows. We expect to be given (x, y), which the tractor's tracking software believes is our current location. We are also provided with a set of n_i and f_i, which are the current readings for each beacon. We also require ϵ, measured in meters, giving the maximum number of meters of inaccuracy we expect as a result of beacon positioning error and systemic errors. For our typical wavelength of $\lambda = 1$ meter, we have $\epsilon \approx 0.2$ meters.

For each beacon i, we calculate $R_i = (x - x_i)^2 + (y - y_i)^2 - \lambda^2(n_i - f_i)$. If $-2r\epsilon - \epsilon^2 \leq R_i \leq 2r\epsilon + \epsilon^2$, then the current n_i and f_i are within ϵ of our current position, and thus valid.

We arrive at this idea by again looking at circles. Let $r = \lambda(n_i + f_i)$ be the radius of the circle that our observed n_i and f_i indicate we are on. If our current position falls on a circle of radius at most $r + \epsilon$ and at least $r - \epsilon$, then we are within ϵ of the expected location, and so our data is assumed to be valid. Substituting our (x, y) into the equation for a circle of radius r centered at (x_i, y_i), the residual must be at most $(r + \epsilon)^2 - r^2$ and at least $r^2 - (r + \epsilon)^2$, which give the stated results.

To implement this, we suggest the following algorithm should be applied approximately every minute (600 measurements):

1. For each beacon, calculate $d = 2r\epsilon + \epsilon^2$.
2. If all beacons satisfy $-d \leq R \leq d$, all data is valid. Continue.
3. If there is only one beacon for which $-d \not\leq R \not\leq d$:
   (a) Replace n_i with $n_i - 1$ and $n_i + 1$, and re-check the inequality using each.
   (b) If exactly one of the above is within bounds, change n_i to $n_i - 1$ or $n_i + 1$ as appropriate and continue.
   (c) If neither or both of the above are within bounds, stop and recalibrate.
4. If more than one beacon does not satisfy $-d \leq R \leq d$, stop and recalibrate.

5.3.3 Implementation

We use the MATLAB package `bcnls` to find a least-squares solution to this problem. Unfortunately, we experimentally found that our objective function appears to have many local minima. Since we are not using a global optimizer, the solver tends to locate a minimum value very near to its initial position.

To alleviate this problem, we implement a multilevel mesh-based heuristic search. This search works as follows. Begin with the search region $x_0 - 10 \leq x \leq x_0 + 10$, $y_0 - 10 \leq y \leq y_0 + 10$.

1. Divide the search region into a 5×5 grid.
2. For each grid point, run a reduced version of the nonlinear program and record the objective value.

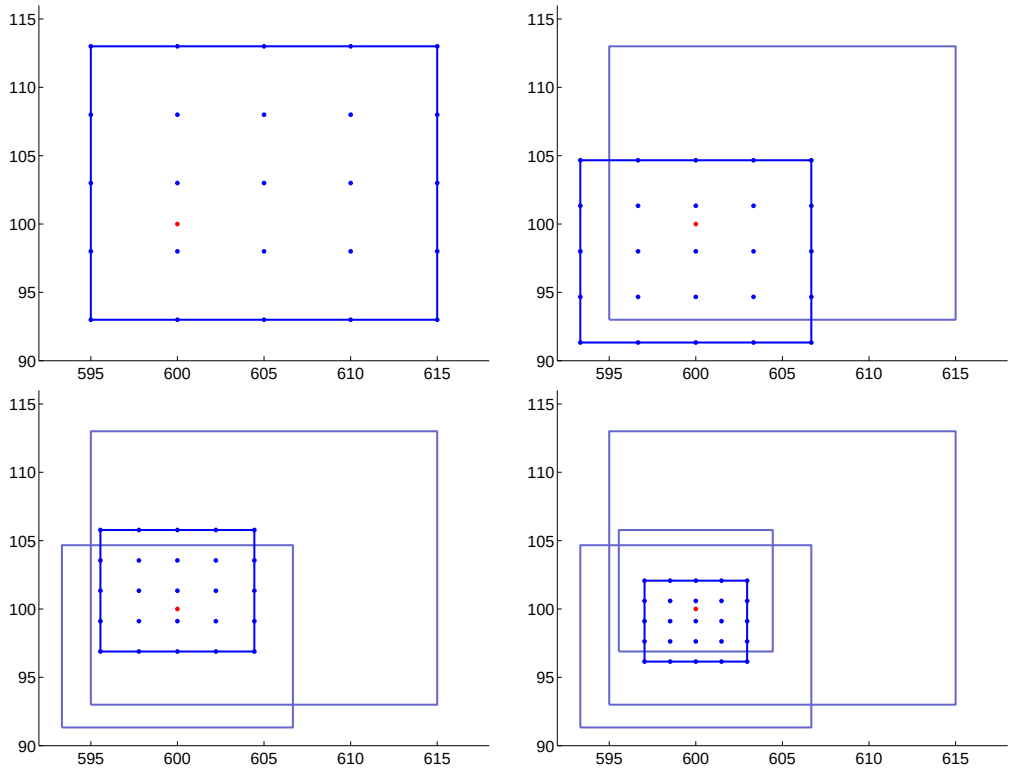

Figure 17: Multigrid mesh focusing in on a tractor (the red dot)

3. Replace the search region with a new region whose sides are half the length of the old one, and which is centered at the point giving the smallest objective value.

4. If we have iterated fewer than 12 times, go to 1. Otherwise run the full nonlinear program on the point giving the smallest objective value, and return the result.

This method effectively "zooms in" on the best locations, identifying the (approximate) area containing the globally best solution. The final step ensures that we find the locally best point, giving a good approximation for the global minimum. Figure 17 shows four steps of this process.

The reduced nonlinear program is a version of the full nonlinear program in which the error tolerance level of the solver is decreased, and a_i and b_i are removed, to increase speed.

With regard to our problem constraints, note that this method could be implemented entirely using C on simple processors. Our current implementation uses MATLAB, but `bcnls` is based on `bcls`, which is already written as a C library.

5.3.4 Experiments and Results

Our tests have demonstrated that our method is effective and requires relatively little time. A tractor's positioning system could easily be calibrated during the time in which the tractor is driving down a driveway into a field. Experiments have also shown that taking additional readings (increasing r in our model) does *not* typically increase the accuracy of our model, beyond $r = 4$. We believe this makes sense because that is the r value at which there are more equations and variables. Any additional readings essentially give the same information, with the same systemic errors and beacon position errors.

Experiment 1: Quantization. Quantization levels significantly affect error levels. We use a five beacon field with a typical arrangement, as depicted in Figure 18. Four beacons are located around the outside of the field, one per edge, and one beacon is located in the center (such as at a rock pile).

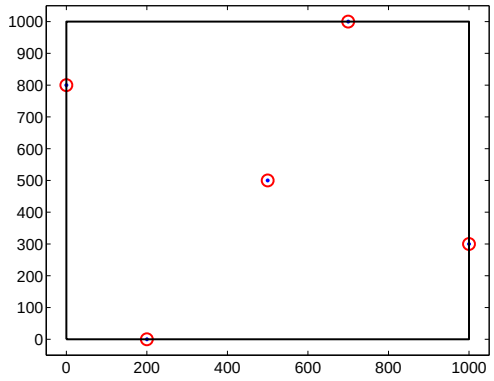

Figure 18: Five beacons in a typical field.

Figure 19 demonstrates a significant benefit in increasing the quantization level to 11 bits, which corresponds to 2048 levels of quantization.

Experiment 2: Number of Beacons. We place n beacons in a circular arrangement, with n ranging from 1 to 20, as shown in Figure 20. While

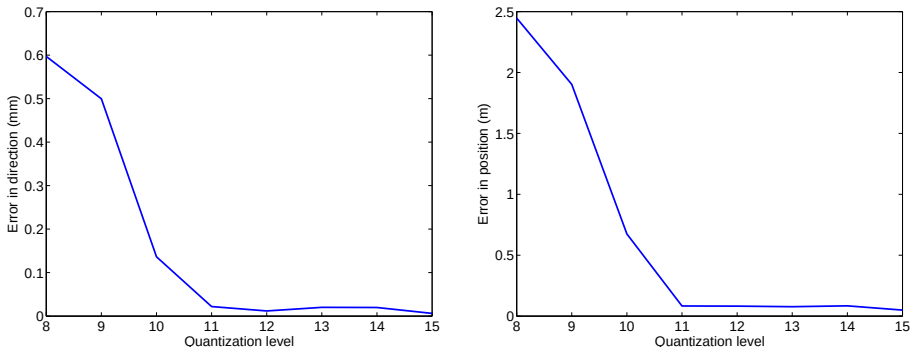

Figure 19: Error in direction and position of tractor versus quantization level for the field in Figure 18. Note: Level 15 represents no quantization.

a larger number of beacons generally improves the accuracy of our findings, it is not significantly different after about 5 beacons, as demonstrated in Figure 21.

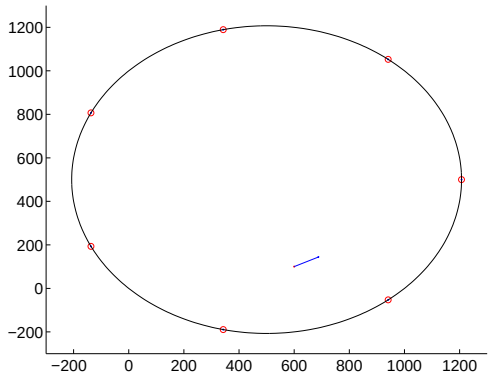

Figure 20: A circular arrangement of beacons.

Experiment 3: Number of steps. Each step in our calibration consists of one reading of the fractional portions of the wavelength. At least two steps are needed to determine direction and position. We use the field in Figure 18 and take increasing numbers of steps. Figure 22 shows that in general, larger numbers of steps do not help accuracy, and indeed hurt in some cases. We do not understand these results, however, notice that the scale on these graphs is extremely small.

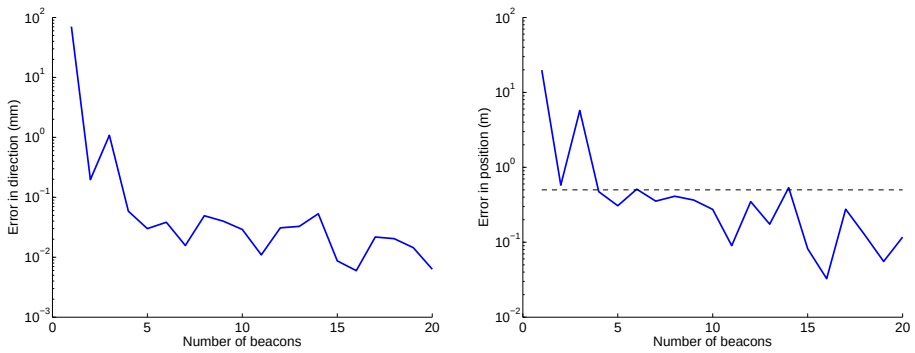

Figure 21: Error in direction and position of tractor versus number of beacons used.

We recommend increasing the quantization levels. These produce enormous amounts of error compared even to measurement errors.

5.3.5 Further Work

As previously mentioned, the bounded constraint nonlinear solver that we use tends to find local minima. It is possible that an unbounded solver, or a solver that uses various heuristics, may be able to find better solutions, thus reducing the need for our meshing algorithm.

In addition, our meshing algorithm could be improved. Much work has been done using multilevel meshes, and there are many techniques that could be applied in our case to improve location of global minima. Our current implementation is fairly basic and can occasionally be "tricked."

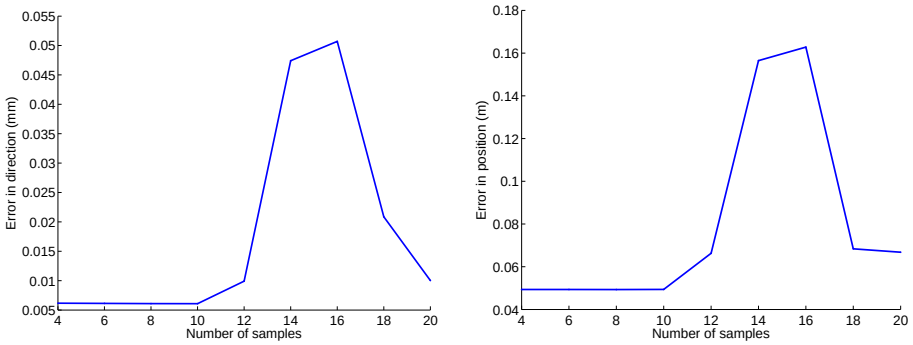

Figure 22: Error in direction and position of tractor versus number of steps taken.

5.4 Assessing an Off-the-Job Safety/Wellness Strategy at Mosaic Potash

Solution Team: Laura Cowen, Ortho Flint, and Ryan Tifenbach.

5.4.1 Problem Statement

As of 2004, Saskatchewan had an injury rate that was twice the national average; accordingly, injuries are now considered an epidemic in the province. The yearly cost of these preventable injuries to the economy is estimated at one billion dollars province-wide. It was found that most injuries occur off-the-job.

In 2004 and 2005, Mosaic Potash Colonsay (one of the 4 Mosaic mine sites in Saskatchewan) kept track of the number of days lost due to workplace injuries each year and averaged 10 days lost per year. However, OJIs (off-the-job injuries) resulted in 50 times more average-work-days lost. Similar results were observed for days lost due to illness. These findings drove Mosaic Potash to develop an Off-the-Job Safety/Wellness Strategy for their employees, families, communities, and Saskatchewan as a whole.

After these initiatives were instituted, the number of days lost due to OJIs was compared for the first quarters of 2006/07 and 2007/08. A 14% reduction in days lost was found between these two years.

Mosaic Potash would like to determine if their safety/wellness programs are having an impact on reducing the number of days lost per year to OJIs and to illness.

5.4.2 Before After Control Impact (BACI) Studies

In this report, we will provide suggestions as to what data is important to collect in future studies, provide a "first draft" of a form employees could fill out to describe their injuries when processing a claim or returning to work, discuss potential experimental designs for assessing impact, and discuss models to determine what factors are affecting the number of days lost. Further, we will emphasize the importance of collaboration with other parties interested in OJIs.

The need for a control group

One of the problems with the finding of a 14% reduction in days lost is that we don't know if that reduction is due to the safety programs implemented or was there a general 14% reduction during that time in Saskatchewan. This statistic needs to be compared with a similar statistic for people who did not receive safety programs for the same time period (a control group). Two things are crucial for the mcontrol group — the same time period, and the lack of safety programs. For instance, we cannot simply use the data from say 2004/05, when off-the-job safety programs were not in place, as a control, as we do not know if there was an external event that was associated with off-the job injures that occurred in 2004/05 but did not occur in 2007/08. For example, perhaps 2004/05 had an icy winter where many OJIs were due to falls. Thus, time is said to be confounded with the impact of the safety program.

Data

There is a need for collecting the number of days lost due to OJIs per year in all of the mines as is currently done. However, we also need to collect the number of days lost due to OJIs per year for a control group. This control group could be another potash mining company such that the employee demographics were similar to Mosaic Potash, a similar industry group with similar demographics who use the same medical insurance company (this would make data collection by the insurance company easier), or if neither of these is possible, population data for Saskatchewan. It may also be possible to use Mosaic Potash employees as their own controls; this would require a rotating treatment design which is not discussed further in this report.

A 2-year study design

For a simple 2-year study with a BACI design you would need data as in Table 1. Here we define y_{ijk} as the number of days lost due to injury divided by the number of people at mine site k for treatment group i and year j. The difference between the two years can then be used to determine if there is a treatment effect. For example, d_{11} would be the difference between the number of days lost due to injury for year 1 versus year 2. These differences are treated as two independent samples (from a control and a treatment group) and a 2-sample t-test or confidence interval can be constructed to determine if the treatment groups differ significantly.

Table 7: Data table for a 2-year BACI study where the treatment group has 4 sites (Esterhazy, Colonsay, Belle Plaine and Hersey) and the control group has 3 sites (A, B, and C).

		Year		Difference
Group	Site	1	2	$y_{ijk} - y_{ij+1k}$
Treatment	Esterhazy	y_{111}	y_{121}	d_{11}
	Colonsay	y_{112}	y_{122}	d_{12}
	Belle Plaine	y_{113}	y_{123}	d_{13}
	Hersey	y_{114}	y_{124}	d_{14}
Control	A	y_{011}	y_{021}	d_{01}
	B	y_{012}	y_{022}	d_{02}
	C	y_{013}	y_{023}	d_{03}

Multi-year designs

Before-after control-impact (BACI) designs appear in statistical ecology research. They were first introduced as a tool in analysing the effects of a disturbance or disaster on wildlife. In what might be called "traditional" statistics, an analysis of an experiment requires that the experiment be repeated (or, at the very least, be repeatable). But, what about experiments that cannot or should not be repeated? The Exxon Valdez disaster is an example. There is no way to conduct *multiple* experiments concerning this or similar events. None of the classic statistical models are entirely appropriate for such a situation, and so the BACI-type design was introduced.

How exactly can a design intended for such purposes apply to analysis of the effectiveness of safety initiatives? In a sense, Mosaic Potash and Safe Saskatchewan are involved in an experiment that can only be conducted

once; once the initiatives are implemented, there is no way to "reset" the population dynamics and start over. The beginning of these initiatives is a one-time event that cannot be recreated feasibly (for budgetary and time-influenced reasons).

Terminology

We refer to the event under investigation as the "treatment"; the treatment may be an ongoing process. We are counting numbers of observations of some event, object, or individual, that may or may not have been affected by the treatment. A BACI-type analysis asks the question, "has the treatment altered the number of observations?" In our case, have the safety initiatives altered the number of work days lost to off-the-job injuries?

The design looks at observation counts taken at different times. We refer to these times as year 1, year 2, etc., and assume that the treatment occurred between two known years.

Required Data

In order to proceed with this type of analysis, we need a certain sort of collection of data. The observations are recorded at a number of sites, at a series of different times. Each site is either part of the experimental control or it is treated (for example see Table 2).

Table 8: Example data from a 3 year study where the impact occurs between years 1 and 2.

		Year		
Group	Site	1	2	3
Treatment	1	35	13	12
	2	16	13	3
	3	10	28	6
Control	1	32	7	24
	2	3	1	15

Limitations

A BACI-design is only appropriate if data from before and after the experiment, and from both a control and an experimental group is available.

A BACI analysis *cannot* be used to infer causation or correlation; it is *part* of an argument for such a relationship.

5.4.3 Details

Let μ_{ijk} be the number of observations under treatment i, at time j and location k (the control is treatment 0 and the experimental group is treatment 1). For example, $\mu_{1,3,5}$ is the number of observations at treated site 5 during year 3, $\mu_{0,2,1}$ is the number of observations at control site 1 during year 2.

Let y_{ijk} be the number of observations recorded under treatment i at site k during year j divided by some scaling factor that denotes the relative size of site k. In this study, for example, we will take y_{ijk} to be the total number of lost days (to off-the-job injury) divided by total number of employees at a specific mine during a specific year.

The BACI design includes three distinct models for the random variable y_{ijk}, we present the most basic: the untransformed additive model.

The additive model

The observation counts are assumed to be

$$y_{ijk} = \mu + \tau_i + \alpha_j + (\tau\alpha)_{ij} + w_{ik} + \epsilon_{ijk}$$

where

- μ is the overall population mean;
- τ_i is the effect of treatment i;
- α_j is the effect of year j;
- $(\tau\alpha)_{ij}$ is the interaction effect between treatment i and time j;
- w_{ik} is the site effect, these are assumed to be independent identically distributed normal random variables with mean 0 and variance σ_w^2; and
- ϵ_{ijk} is the experimental error, these are assumed to be independent normal random variables with mean 0.

With these assumptions and the given data set, a variety of hypotheses can be tested. For example, we can test the hypothesis that the treatment has no effect on the year immediately following it as compared to the year immediately previous. We let $\hat{\delta}_{ik} = y_{ij_2k} - y_{ij_1k}$, where j_1 and j_2 are the years immediately before and after the beginning of the treatment. We would then perform a two-sample pooled t-test on these differences to obtain a confidence interval for this hypothesis; this is known as the BACI-contrast.

In some cases, this model can lead to unreliable results. For example, if a large number of relatively small counts occurs (especially counts equal to 0), the assumption that each ϵ_{ijk} is normal can cause difficulties. In these cases, more advanced models such as the log-transformed multiplicative model or the generalized linear mixed model are more appropriate. For an in-depth discussion of the BACI-design.

5.4.4 Data Collection Survey

To determine if safety programs are addressing a particular area of safety, we suggest monitoring where these OJI's occur . For example, SmartRisk states that 5% of all injuries in Saskatchewan are due to poisoning and 36% are due to falls. If OJIs for Mosaic Potash employees were similar, then perhaps they would implement a safety program that focussed on falling. This will result in a spreadsheet with column headings similar to, for example, Table 3.

Table 9: Example of data that would be collected on a survey to be used in a generalized linear model

Number of days lost per year	Injury Type
10	Farm
1000	Falls
250	Poison
100	Other

We have developed a draft survey form (Appendix A) that Mosaic Potash could use to collect this data on OJIs. By no means do we see this as a final version, as many organizations would need to have input into this form. If collaborating with a university, a final version would have to go through an ethics review along with having the approval of union officials, Safe Saskatchewan and other concerned organizations.

Table 10: Example of a possible contingency table resulting from data collected in the Off-the-Job Injury survey.

	Injury Type				
Group	Domestic	Farm	Sport	Motor Vehicle	Other
Mosaic Potash					
Other Potash company					

5.4.5 Generalized Linear Models

Linear models attempt to model a response (number of days lost per year (n)) to covariates (such as age, sex, injury type, safety program type, etc.) in a linear fashion. For example, if we just had *sex* (male=1 or female=0) as a covariate, we could develop a simple linear regression model (assuming n is normally distributed) along the lines of

$n = \beta_0 + \beta_1 sex + error$

where β_1 would be interpreted as the change in number of person-days lost for males versus females.

Contingency Tables

Taking data from the form described above, we could create a contingency table which categorizes the number of person days lost by two or more covariates. For example, if we were interested in categorizing the number of days lost per year by injury type and group we could create the following table (Table 4).

Again, a linear model could be developed, but we might have different assumptions as to how the number of days lost is statistically distributed (Poisson or Normal). A general model would be where we write the model in terms of the expected number of days lost

$E(n) = \beta_0 + \beta_1 group + \beta_2 domestic + \beta_3 farm + \beta_4 sport + \beta_5 vehicle$

where all covariates are indicator variables (taking on values of 0 or 1). For example group=1 would indicate Mosaic Potash, and group=0 would indicate not Mosaic Potash. Farm=1 would indicate a farm related injury and farm=0 would indicate an injury other than farm. If we had five types of injuries in the table, these can be modeled by four indicator variables. When all indicator variables for injury type are equal to zero, this indicates

the last category (other=not domestic, not farm, not sport, and not motor vehicle).

When n is assumed to be Poisson distributed, some of the β parameters have an interesting interpretation: the log odds or the log odds ratio of events.

Odds

The odds of an event is the probability of success over the probability of failure of that event. For example, if p is the probability of a farming injury and $1-p$ is the probability of not a farming injury, then $\frac{p}{1-p}$ is the odds of a farming injury. So for example if $p = 0.8$ and $1-p = 0.2$ then the odds of a farming injury would be $\frac{0.8}{0.2} = 4$. This is often stated as "the odds of a farming injury is 4-to-1". Thus odds are not probabilities of an event, rather they are relative probabilities.

Odds ratio

An odds ratio is a method of determining if the odds of an event is the same in two groups. It can be expressed as

$OR = \frac{p/1-p}{q/1-q}$

where p is for example, the probability of a farming injury, for say, males, and q is the equivalent for females. An odds ratio of 1 implies a farming injury would be equally likely for both males and females.

If we had the model:

$ln(E(n)) = \beta_0 + \beta_1 group + \beta_2 farm + \beta_3 group * farm$

where group and farm are the indicator variables defined above. Then we can break this model into 4 cases as shown:

Case	Associated part of model
1) Mosaic and farming injury	$ln(E(n)) = \beta_0 + \beta_1 + \beta_2 + \beta_3$
2) Mosaic and non-farming injury	$ln(E(n)) = \beta_0 + \beta_1$
3) Other company and farming injury	$ln(E(n)) = \beta_0 + \beta_2$
4) Other company and non-farming injury	$ln(E(n)) = \beta_0$

Then looking at the difference between two of these cases give the log odds. For example, comparing case 3 with case 4 gives $(\beta_0 - \beta_2) - (\beta_0) = \beta_2$. Thus β_2 is the log odds of a farming injury versus a non-farming injury for the other company. Similarly, $(\beta_0 + \beta_1 + \beta_2 + \beta_3) - (\beta_0 + \beta_1) = \beta_2 + \beta_3$ would be the log odds of a farming injury versus a non-farming injury for Mosaic.

To compare Mosaic with he other company we look at the difference between these two log odds to form the log odds ratio: $(\beta_2 + \beta_3) - (\beta_2) = \beta_3$.

This is the log odds ratio of a farming versus a non-farming injury for Mosaic versus another company.

5.4.6 Collaboration

To perform this study well, a great effort must be made to collaborate between various stakeholders. Good data is collected when all parties are informed well about the study and in particular what action Mosaic will take in response to the results. Union representatives, the insurance company, potential control companies, and Safe Saskatchewan are some of the parties who will need to agree on the final survey design, logistics, etc. Finally, it is our recommendation that a statistician be among those involved from the beginning to help bring about a well designed study with the potential to collect good data.

5.4.7 Conclusion

With Saskatchewan's injury rates being among the highest in the country, a study initiated by Mosaic Potash seems timely. Careful consideration must go into obtaining a control group so that a BACI-type design may be implemented. To determine which areas of injury the safety/wellness programs should target, a survey is recommended who's data can later be analyzed through a general linear model. Finally, we stress that a study of this magnitude, involving many people, would require the collaboration of all interested parties.

5.4.8 Appendix

Appendix A: Example of an off-the-job injury assessment survey

The purpose of this study is to determine the hazards that may occur for employees away from work. There is no need to provide your name or address anywhere on this questionnaire. This study is sponsored by the Department of X at the University of X.

Please describe the injury.

Was the injury due to a fall?

To the best of your memory, give the date of the injury.

Day Month Year

Did the injury occur...
- before your work day at Mosaic?
- after your work day at Mosaic?
- on your weekend?
- during vacation or time off?

Where did the injury occur?
- a farm
- any public area including streets and roadways
- urban home (population 5,000 or greater)
- rural home
- other

Did the injury occur while or during...
- farming or any other agricultural activity except gardening?
- gardening or cutting grass?
- home renovations or improvements or cleaning?
- day to day household activities?
- a non-recreational motor vehicle accident?
- a recreational motor vehicle accident involving an ATV?
- a recreational motor vehicle accident involving a snowmobile?
- a recreational motor vehicle accident involving neither of the above?
- bicycling?
- swimming?
- sports or recreation other than bicycling or swimming?

- a house fire?
- on any public transit (including planes, trains or buses)?
- a natural disaster?
- other

In your opinion, when the injury occurred, were you mostly ...
- tired?
- stressed?
- neither?

If you answered neither above, was the injury due mostly to...
- being rushed?
- not paying attention?
- distractions?
- other

In your opinion, was the injury due to a lack of attention (rather than being tired, stressed, rushed)?

Yes No

Was a hospital stay required?
Yes No

If yes, how long?
- 1-3 days
- 4-7 days
- more than a week

5.5 How to Optimize Combination Chemotherapy in Cancer Treatment?

The following problem was posed by Professor Jack Tuszynski from the Cross Cancer Institute in Edmonton, Canada. Find ways to optimize combination chemotherapy in cancer treatment. There were three distinct subgroups working on this problem. We include just one approach in this abbreviated version of the report. All approaches will be included in the final report.

Introduction

The current medical treatments for cancer include surgery, radiation therapy, gene therapy and chemotherapy. Combinations of these therapies have shown to be effective. Primarily we restricted our attention to chemotherapy and specifically addressed the general question: 'how can one evaluate and best improve the outcome for a patient by changing the application frequency of a particular drug or by using a combination of drugs'.
The outcome involves two components:

benefits The benefits can be measured in terms of the patient survival time, or the time to metastases, or may be measured in terms of the tumour size or death rate of tumour cells.

side effects Of course chemotherapy drugs are 'poisons' which destroy both tumour and normal cells and/or effect the total physiology of the body, so balanced against the benefits are the side effects which may be either of short term duration and a matter of inconvenience, or of long term duration or indeed life threatening.

Mitosis and Drugs

Individual cells (normal and tumour) go through a cell cycle, the end product of which is a daughter cell. The cell cycle is typically 24 hours in duration (but varies greatly depending on cell type) and has four development stages: G1 ('gap 1', lasting typically 10 hrs), S ('DNA synthesis', tasting typically 5-6 hrs), G2 ('gap 2' , 4 hrs), M ('mitosis', 2 hrs), see Figure 24.

Mitosis is itself usually divided into different stages, see Figure 23. Tumour cells generally have a shorter cell cycle and, whereas normal cells die after a number of cycles, cancel cells continue to reproduce. This results in uncontrolled growth and proliferation with accompanying biochemical and environmental change (hypoxia, acid environment etc.). Conventional drugs

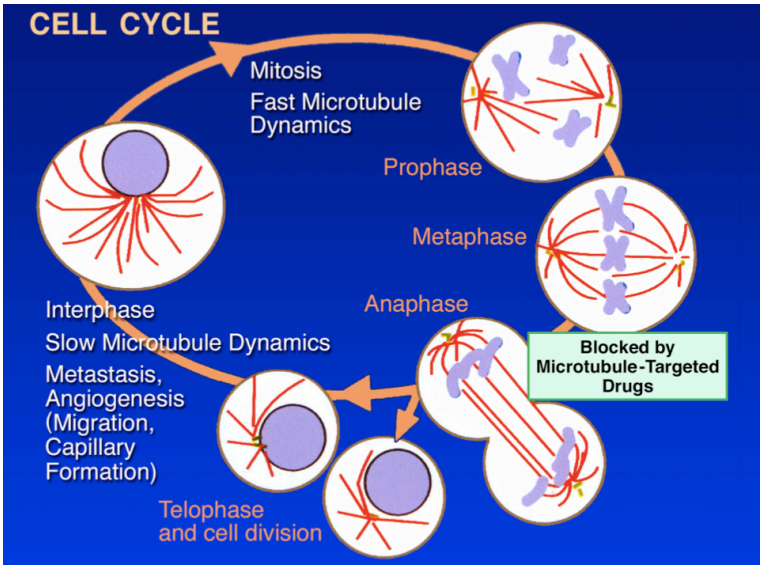

Figure 23: The cell cycle

bind onto particular key molecules produced during the cell cycle. The cell is particularly vulnerable during the 'most active' development stages of DNA synthesis and cell division, so most drugs in present day use target either one or other of these cycle stages, see Figure 24 (DNA alkylating agents, DNA processing agents, microtubules). Other drugs act on the cellular environment and so effect the cell at all developmental stages; anti-angiogenesis drugs fall into this category.

There are over 100 cancer drugs and a very large number of possible drug combinations and possible application schemes. At the present time treatments are mostly based on ad-hoc empirical 'formulas'. The Cross Cancer Institute would like to develop a sounder theoretical basis for treatments.

Typical Drug Combinations

(Notice Ringa, Petko Kitanov)

We will first briefly describe a typical drug combination/s used for breast cancer. This specific example will both familiarize the reader with the treatment process and provide a foil for some of the work to follow. Data concerning tumour cell survival rates/application levels for these drugs are used in models presented later.

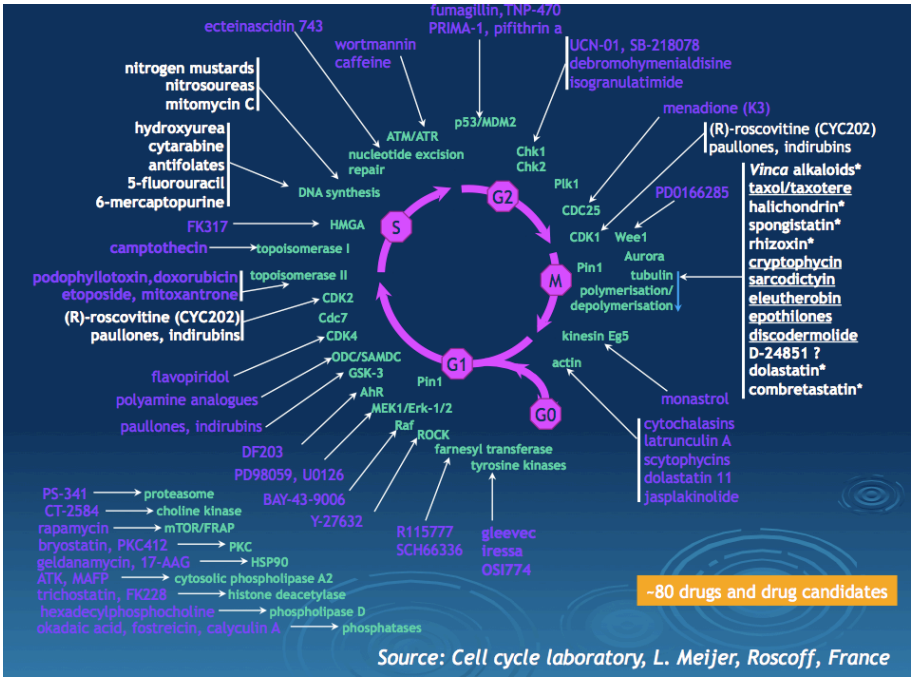

Figure 24: Drug action

AC chemotherapy:

The drugs administered for this treatment regime are:

- *Doxorubicin*, which acts at stage S of the cell cycle and has a half life (footnote on half life?) of 55 hours, and
- *Cyclophosphamide* which acts at stage S of the cell cycle, and has a half life of 3 to 12 hours.

Both drugs are given at the same time every 3 weeks. Typically the drug reaches the tumour site (the cell DNA) in 4 hours and is active for about 48 hrs.

Some Side Effects

Side effects for both include lowered resistance to infections, bruising and bleeding, anemia, hair loss, irritation of the bladder, diarrhea. When used in combination immune system problems arise.

There are other treatments (eg TAC). Comparisons in outcomes are important.

In an attempt to address various features of the problem different groups were formed and their works are presented in the following sections. The first group produced a general probabilistic model for determining the outcome of various treatment regimes using parameters to be determined empirically.

5.5.1 A General Procedure for Evaluating Drug Treatment Regimes

Luz Angélica Caudillo Mata	CIMAT
Harish Krishnamurthy	Northeastern University, Boston
Sadia Mwangangi	University of Regina
George Price	University of Regina
Hugo Rodríguez Ordóñez	University of Regina

Cancer is responsible for the majority of non–accidental deaths in the modern world and has been a major cost to health systems worldwide. Enormous amounts of research time and resources have been invested in the fight against this disease. Although still in early stages, mathematical modeling is playing an increasingly important role in this fight.

Methods currently used to treat cancer include surgery, radiotherapy and drug therapy. Drug therapy exists in the form of chemotherapy or anti–angiogenesis. Chemotherapy interferes with the natural life cycle of cells eventually causing their death while anti-angiogenesis inhibits the growth of blood vessels near the tumor, depleting the nutrient supply while enhancing the flow of other therapy drugs.

The various stages of the cancer cell life cycle are described. The cell[2] cycle is then generalized to M stages. A model to determine the time dependence of cell populations at each stage of the generalized cell cycle is given. In section 5.5.3 the model given in the previous section is treated as a discrete dynamical system and its time evolution derived. In section 5.6 the model is applied to the case where two chemotherapy drugs are used. In this case it is assumed that the drugs are administered at the same frequency but the time between when the drugs are given is varied. In section 5.7 a method to estimate parameters for the model is described. In section 5.7.1 future work and conclusions are summarized.

5.5.2 The Model

The cell cycle has four basic stages:

1. A *first gap* stage or $\mathbf{G_1}$ where a cell grows in size.
2. The *synthesis* stage or **S** during which DNA is replicated.
3. A *second gap* stage or $\mathbf{G_2}$ where the cell continues to grow.
4. The *mitosis* or **M** stage, in which a cell splits into two genetically identical *child* cells. The *child* cells then proceed to $\mathbf{G_1}$ and the cycle repeats.

Each stage of the cell cycle can be divided into several substages. For this reason, in the model that follows, the cell cycle is assumed to have M stages. At any time t, the cell cycle is considered to be a series of "buckets", S_k, each one symbolising a stage in the cycle and containing the portion of all cells that are currently at that stage (see figure 26).

Let $N_k(t)$ be the average number of cells at phase k and time t and

$$N(t) = \sum_{k=1}^{M} N_k(t)$$

the average total number of cells. One goal is to choose a treatment plan that forces $N(t)$ to be a small as possible while maximizing the patients quality of life during the treatment interval.

As cells undergo their cycle, some will migrate from a given stage S_k to the next stage after some time. Let $\lambda_k(t)$ be the average rate at which cells

[2]Throughout the remainder of this section cell will always mean cancer cell.

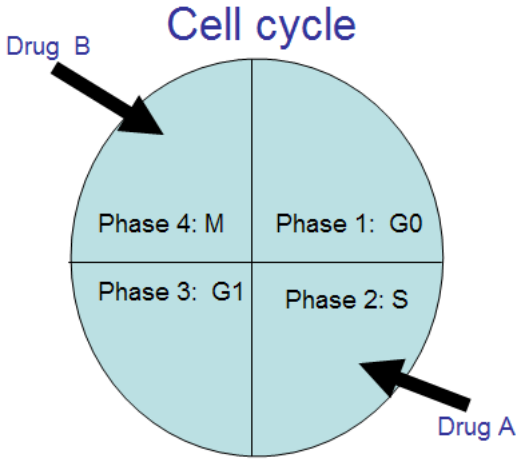

Figure 25: The cell cycle

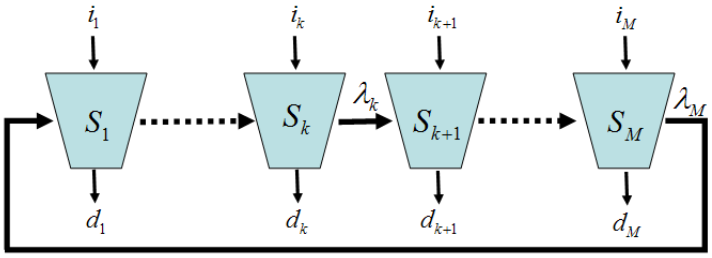

Figure 26: A simplified model for the cell cycle

transition from stage S_k to stage $S_{(k+1) \mod M}$ at time t. In addition, let $d_k(t)$ be the average rate at which cells die in stage S_k at time t and $i_k(t)$ be the average rate at which cells divide in stage S_k at time t (it is expected that only one i_k is not zero).

The rate of change of average cell population in a given stage, S_k, equals the average rate that cells flow into S_k minus the average rate that cells flow out of S_k. This can be stated mathematically as:

$$\frac{dN_k}{dt}(t) = \lambda_{k-1}(t) + i_k(t) - \lambda_k(t) - d_k(t) \tag{19}$$

Equation (19) is a statement of conservation of average cancer cell population at stage S_k. In the above equation it is assumed that cells can transition from a given stage to the next but cannot skip stages.

5.5.3 Time Evolution of the Model

Equation (19) gives the time dependence of the average number of cells in stage S_k. Since $\lambda_{k-1}(t)$, $\lambda_k(t)$, $i_k(t)$ and $d_k(t)$ refer to average rates that cells flow between various stages it is natural to expect that these rates can be defined in terms of *transition probabilities*. Let $P_k(t,\theta)$ be the transition probability density. That is the probability that a cell transitions from stage k to $k+1$ in time interval $[t, t+\beta]$ is be given by:

$$F_k(t,\beta) = \int_0^\beta P_k(t,\theta)d\theta \tag{20}$$

It is assumed that:

- At any time t, $P_k(t,\theta)$ is a valid probability density function.
- $P_k(t,\theta) = 0$ for $\theta < 0$
- $P_k(t,\theta)$ is continuous for $\theta \in [0,\infty)$

$F_k(t,\beta)$ in equation (20) will be referred to as the transition probability distribution. By the Fundamental Theorem of Calculus, $F_k(t,\beta)$, is differentiable on $(0,\infty)$ with derivative:

$$P_k(t,\beta) = \frac{\partial F_k}{\partial \beta}(t,\beta) \tag{21}$$

If β is sufficiently small, then by the Mean Value Theorem

$$\int_t^{t+\beta} \lambda_k(s)\,ds = \lambda_k(\tilde{t})\beta = N_k(\tilde{t})F_k(\tilde{t},\beta) = N_k(\tilde{t})[F_k(\tilde{t},\beta) - F_k(\tilde{t},0)]$$

for some $\tilde{t} \in [t, t+\beta]$, which can be assumed equal to t for small enough β. Therefore,

$$\lim_{\beta\to 0} N_k(t)\left[\frac{F_k(t,\beta) - F_k(t,0)}{\beta}\right] = \lambda_k(t)$$

and finally

$$\lambda_k(t) = N_k(t)\frac{\partial F_k}{\partial \beta}(t,0) \tag{22}$$

The above derivation of $\lambda_k(t)$ can be repeated for $i_k(t)$ and $d_k(t)$. Let $P_k^i(t)$ and $P_k^d(t)$ be the transition probability densities associated to $i_k(t)$ and $d_k(t)$, respectively. In addition, let $F_k^i(t)$ and $F_k^d(t)$ be the transition probability distributions associated to $i_k(t)$ and $d_k(t)$, respectively. We then have the following equations:

$$i_k(t) = N_k(t)\frac{\partial F_k^i}{\partial \beta}(t,0) \tag{23}$$

$$d_k(t) = N_k(t)\frac{\partial F_k^d}{\partial \beta}(t,0) \tag{24}$$

Using equation (21) and the equivalent equations for $i_k(t)$ and $d_k(t)$ the expressions (22), (23) and (24) can be rewritten as:

$$\lambda_k(t) = N_k(t)P_k(t,0) \tag{25}$$

$$i_k(t) = N_k(t)P_k^i(t,0) \tag{26}$$

$$d_k(t) = N_k(t)P_k^d(t,0) \tag{27}$$

Substitution of (25), (26) and (27) into (19) yields

$$\frac{dN_k}{dt}(t) = N_{k-1}(t)P_{k-1}(t,0) + N_k(t)P_k^i(t,0) - N_k(t)P_k(t,0) - N_k(t)P_k^d(t,0) \tag{28}$$

In most situations equation (28) will need to be solved numerically. Meaning that one wants to calculate $N_k(t)$ in time steps of β. Given $N_k(t)$ then $N_k(t+\beta)$ is found by integration of equation (28).

$$N_k(t+\beta) = N_k(t) + \int_t^{t+\beta} \left[N_{k-1}(s)P_{k-1}(s,0) + N_k(s)P_k^i(s,0) - N_k(s)P_k(s,0) - N_k(s)P_k^d(s,0)\right] ds \tag{29}$$

Where $N_k(t)$ is the initial average population of cells at time t. Using the Mean Value Theorem and if β is sufficiently small then equation (29) becomes

$$N_k(t+\beta) = N_k(t) + \beta\left[N_{k-1}(t)P_{k-1}(t,0) + N_k(t)P_k^i(t,0) - N_k(t)P_k(t,0) - N_k(t)P_k^d(t,0)\right] \tag{30}$$

Using (30) one can define the following discrete dynamical system:

$$\begin{bmatrix} N_1(t+\beta) \\ N_2(t+\beta) \\ \vdots \\ N_M(t+\beta) \end{bmatrix} = \begin{bmatrix} 1+C_1(t)\beta & 0 & \dots & P_M(t,0)\beta \\ P_1(t,0)\beta & 1+C_2(t)\beta & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & P_{m-1}(t,0)\beta & 1+C_M(t) \end{bmatrix} \begin{bmatrix} N_1(t) \\ N_2(t) \\ \vdots \\ N_M(t) \end{bmatrix} \tag{31}$$

where $C_k(t) = P_k^i(t,0) - P_k(t,0) - P_k^d(t,0)$, $k = 1, \dots M$. Notice further that the system can be rewritten in the somewhat more convenient form

$$\begin{bmatrix} N_1(t+\beta) \\ N_2(t+\beta) \\ \vdots \\ N_m(t+\beta) \end{bmatrix} = \left\{ \begin{bmatrix} 1+T_0(t)\beta & 0 & \dots & P_M(t,0)\beta \\ P_1(t,0)\beta & 1+T_1(t)\beta & \dots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & P_{m-1}(t,0)\beta & 1+T_M(t)\beta \end{bmatrix} - \right.$$
$$\left. - \begin{bmatrix} P_0^d(t,0)\beta & 0 & \dots & 0 \\ 0 & P_1^d(t,0)\beta & \dots & 0 \\ & & \ddots & \\ 0 & \dots & 0 & P_M^d(t,0)\beta \end{bmatrix} \right\} \begin{bmatrix} N_1(t) \\ N_2(t) \\ \vdots \\ N_M(t) \end{bmatrix} \quad (32)$$

where $T_k(t) = P_k^i(t,0) - P_k(t,0)$, $k = 1, \dots m$. The first matrix encodes the evolution of the cell population due to intrinsic (natural) causes whereas the second one encodes the influence of the drugs in the cell population. Either of equations (31) or (32) can be applied recursively in order to determine the vector

$$\mathbf{N}(t) = [N_1(t), N_2(t), \dots, N_M(t)]$$

describing the time evolution of average cell populations of all stage in the cell cycle. Of primary importance is $N(t)$. From (30) have that

$$N(t+\beta) = \sum_{k=1}^{M} \left\{ N_k(t) + \beta N_k(t) \left[P_k^i(t,0) - P_k^d(t,0) \right] \right\}$$

or

$$N(t+\beta) = N(t) + \beta \sum_{k=1}^{M} N_k(t)[P_k^i(t,0) - P_k^d(t,0)]$$

In summary, the time evolution of the model explained in section 5.5.2 is given in equation (32). The primary effect of drugs is to modify all $P_k^d(t,0), k \in D$, where $D \subset \{1, 2, \cdots, M\}$. D will depend on the type of drugs used during treatment. In order to calculate $\mathbf{N}(t)$ the parameters, $P_k^i(t,0)$, $P_k(t,0)$, and $P_k^d(t,0)$ must be estimated and β chosen. Equivalently $F_k^i(t,\beta)$, $F_k(t,\beta)$, and $F_k^d(t,\beta)$ must be estimated.

5.6 Application of the Model

In what follows it is assumed that the cell cycle has 4 stages and two drugs, A and B are administered (see figure 25). Drug A is administered at time

$t = 0$ and its effect is assumed to take place α_A hours later. Drug B is administered at time τ and its effect is assumed to take place α_B hours later, in other words, τ is the time between the initial administration of the two drugs. It is assumed that both drugs are taken with the same periodicity T. This is summarized in Figure 27, where D_A and D_B are the corresponding doses of the drugs. Furthermore, it assumed that drug A acts on stage 2

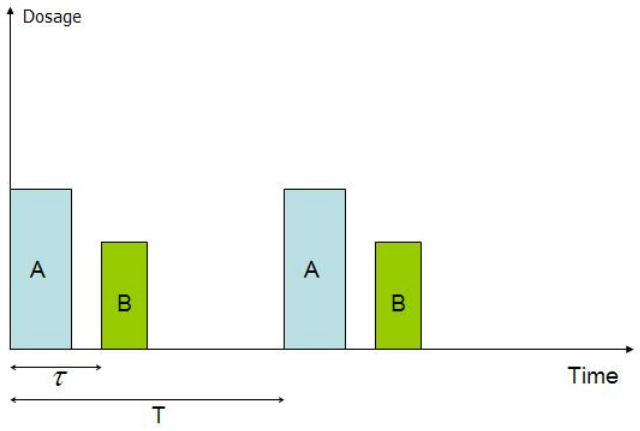

Figure 27: Dosage scenario for combination of two drugs

and drug B acts on stage 4. Figure 28 depicts the survival probabilities of cells under the action of each of the drugs and for different values of τ.

Cell death due to natural causes will be the objective of further study and for the time being it is assumed that the probability of a cell surviving through time β within any stage of its life cycle is a constant 0.95. It is well known that the concentration of either drug over time can be modeled by exponential decay. Assuming a linear relationship between the concentration of the drug and the probability that it kills a cell, the survival probability under the effect the two drugs combined is given by

$$F_k^s(t,\beta) = \begin{cases} 0.95 - K_A D_A e^{-\delta_A(t-\alpha_A)} & k = \sigma_A \\ 0.95 - K_B D_B e^{-\delta_B(t-\alpha_B-\tau)} & k = \sigma_B \\ 0.95 & \text{elsewhere} \end{cases}$$

where δ_j is the decay rate and K_j is the proportionality constant for the linear relationship between killing probability and concentration for drug $j = A, B$, and σ_j is the phase in which drug j acts. This survival probability

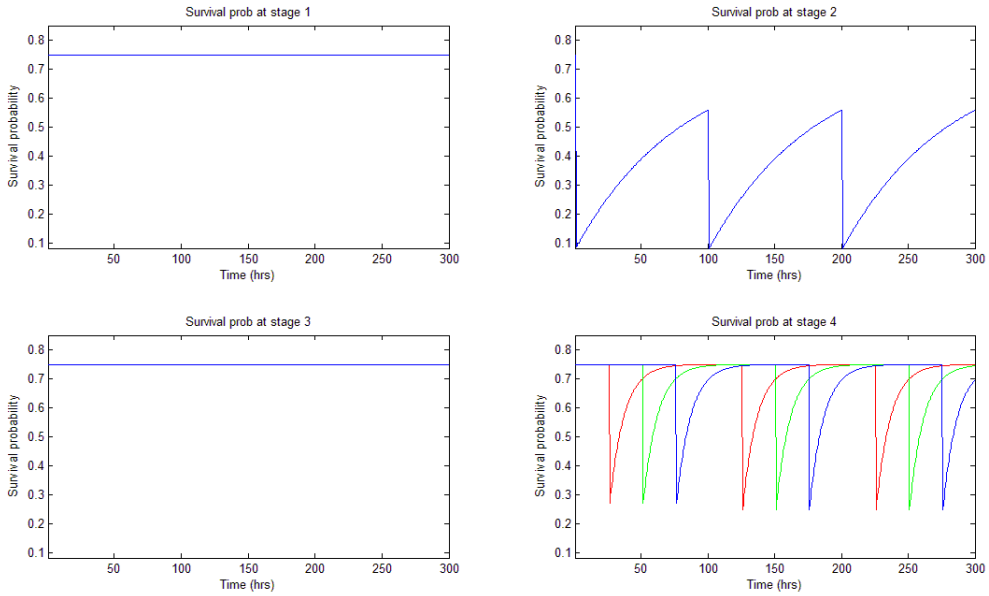

Figure 28: $T = 100$, red$\tau = 25$, green$\tau = 50$, blue$\tau = 75$

Table 11: Known data for a combination therapy

Drug	Doxorubicin	Cyclophosphamide
Stage of action, σ_D	**G1**	**M**
Dosage in mg, D_j	2000	500
Half–life in hours	55	3 to 12
Decay rate, δ_j	.0126	.0924[3]

is related to the cell death probability by $F_k^d(t,\beta) = 1 - F_k^s(t,\beta)$. For the time being, the interaction between the two drugs is overlooked. A combination of two chemotherapy drugs was considered at random. Known data for these drugs appears in Table 11. These parameters along with the artificial ones for the two drugs are shown in Table 12 and Table 13.

Using these parameters the time evolution of average cell populations in each stage of the cell cycle were determined using MATLAB. In Figure

[3]The value of the decay rate was computed using an average half–life of 7.5 hours.

Table 12: Artificial data for a combination therapy

Drug	Doxorubicin	Cyclophosphamide
Proportionality constants (slopes) between concentration and effectiveness (kill rate), K_j	1/3000	1/1000
Delay between drug application and its effect in hours, α_j	0.5	0.5

Table 13: Artificial data for the cell cycle

Stage	$\mathbf{G_1}$	**S**	$\mathbf{G_2}$	**M**
Transition probabilities P_k	1/4	1/4	1/4	1/4
Replication probabilities P_k^i	0	0	0	1/4
Initial number of cells per stage $N(0)$	50	50	50	50

29 the average cell populations in each stage are plotted for $T = 100$ hours and the initial departures of $\tau = 25$, 50 and 75 hours. In addition, the total number of cells, $N(t)$, are plotted in Figure 30 for the three administrations. Examination of the Figures suggests that the time of separation between the initial administration of the two drugs does play a role in the overall cell mortality. Although this example might suggest that the closer the two drugs are administered, the more effective the combination treatment will be, it should be kept in mind that the present model does not yet consider drug interactions.

5.7 Estimation of parameters

Post introduction of the drug, many events happen over time. In a particular phase, some of the main events include:

- Annihilation: The cancer cells are annihilated at each time.
- Migration: The cells in that phase transform into next phase and some cells enter into the phase from the previous phase.

Since the number of cells are large (order of 10^5) and probability of a cell either dividing or getting killed is small, the process can be characterized as

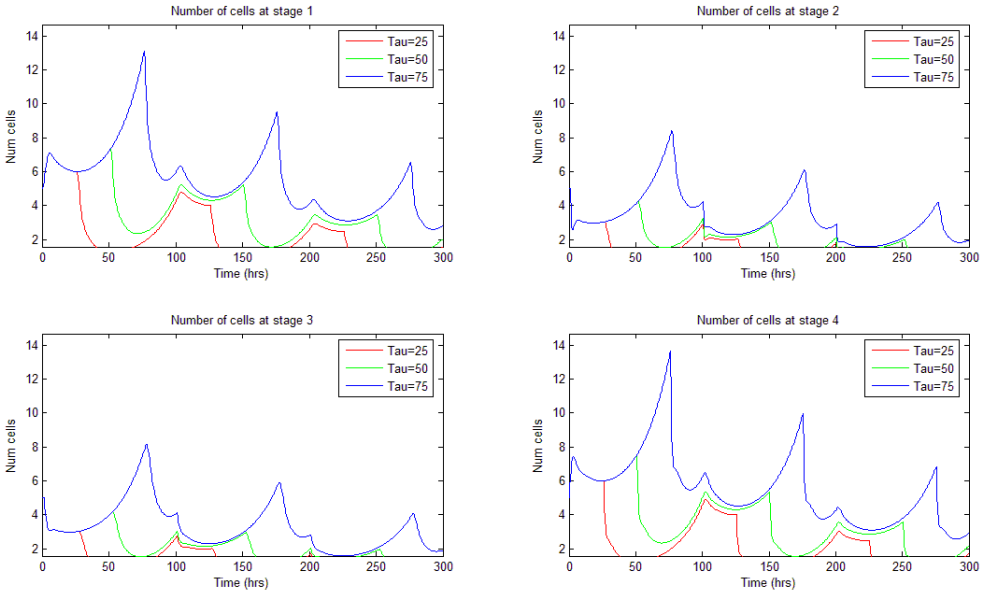

Figure 29: $T = 100$, red$\tau = 25$, green$\tau = 50$, blue$\tau = 75$

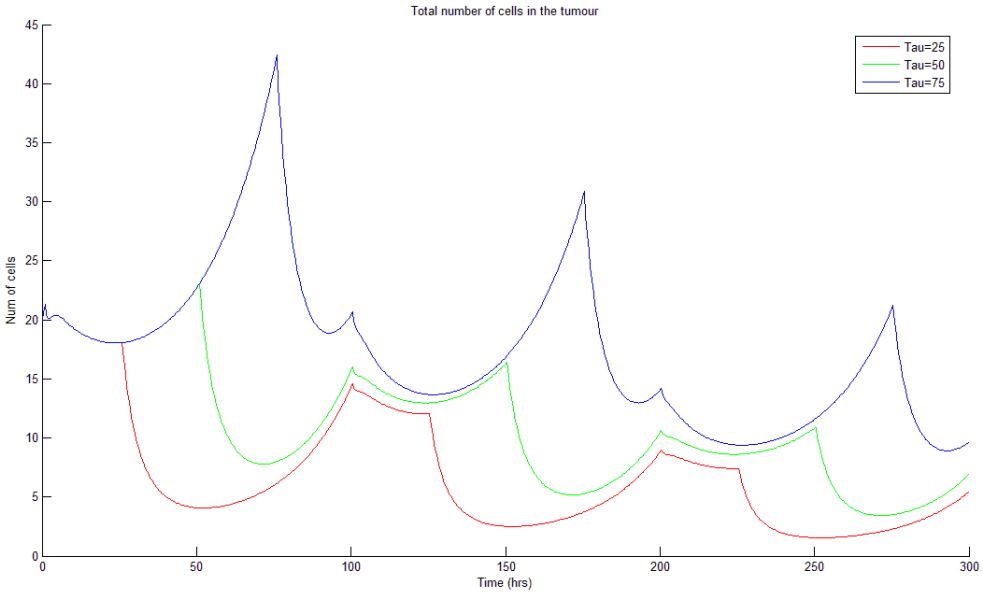

Figure 30: $T = 100$, red$\tau = 25$, green$\tau = 50$, blue$\tau = 75$

a Poisson process. Hence, at a given time t, the annihilation and migration events can be characterized:

The probability that k_1 cells are killed at time t is the random variable A,

$$p(A = k_1) = \frac{\lambda_1^{k_1} \times \exp^{-\lambda_1}}{k_1!} \tag{33}$$

The probability that k_2 cells transform into next phase at time t is the random variable T,

$$p(T = k_2) = \frac{\lambda_2^{k_2} \times \exp^{-\lambda_2}}{k_2!} \tag{34}$$

The probability that k_3 cells infuse into current phase at time t is the random variable I,

$$p(I = k_2) = \frac{\lambda_2^{k_2} \times \exp^{-\lambda_2}}{k_2!} \tag{35}$$

The time taken for a cell that has just entered to migrate into next phase is large compared to time for an existing cell to transform into next phase and hence these events can be assumed to be independent. The drugs are known to destroy a select portion of the tumour such as portion of the outer layer. Hence, each of these processes can be assumed to be independent.

In order to find the joint probability of the events, one must maximize the likelihood of the above processes

$$\begin{aligned} L &= \{p[I = k_2] \times p[A = k_1] \times p[D = k_3]\} \\ &= \prod_{i=1}^{4} \frac{\lambda_i^{-k_i} \exp^{-\lambda_i}}{k_i!} \\ Log(L) &= \sum_{i=1}^{4} \{(-k_i log(\lambda_i) - \lambda_i) - log(k_i!)\} \end{aligned}$$

for which experimental data could be put into this process to determine the mentioned parameters.

5.7.1 Conclusion

This final section contains some thoughts on potential future directions for the model described in this work. The authors concede that this is a quite general model and that further instantiation must take place in order to apply it to a real situation, but at the same time, they believe that the generality described herein might potentially allow for a wider variety of applications.

In the application of the model, given in section 5.6, it was shown that changing the initial time between when the two drugs are administered does effect the cancer cell survival rates. A first task will be to evaluate the accuracy of the present model using real parameters, that is, determining all the parameters of the model for drugs that are actually used in cancer therapy.

Certainly, the probabilistic process involved in modeling the migration probabilities discussed in Section 4 is general enough to estimate these parameters for any type of cells; moreover, even for healthy cells. Estimating these parameters from real data to fit a specific cancer type is certainly a challenge.

Once the real parameters are established for a specific type of cancer and a particular treatment against this disease, it would be in place to evaluate the impact of side effects on the patient and to quantify the interaction of the drugs depending on the departure between treatments τ. This model would then be able to describe the expectancy of improvement in patients and serve as a guideline for more effective clinical trials of combination therapies. Further, by varying dosage times as well as initial departures might eventually sharpen the effectivity expectation and suggest even better clinical trials.

Over longer treatment intervals, drug resistance would need to be factored into the model. The authors believe that this can be done by keeping the relatively simplistic approach of modeling through a discrete dynamical system. Where over time the original cell cycle would become a coupled system of cell cycle. The additional cell cycles would correspond to cells that have developed resistance to given drugs. Instead of having the one matrix given in equation (31) there would now be multiple matrices.

In the future we might imagine that a patient is diagnosed with cancer and a series of tests are performed to determine the type and stage of cancer as well the overall health of the individual. These test results would be used as inputs to a treatment algorithm that could be used to determine optimal therapy options[4]. With regards to chemotherapy the algorithm would potentially choose the best combination of drugs and drug schedules to maximize the cancer cell kill rate over a given treatment interval while at the same time maximizing the patients overall quality of life. As the patient is being treated additional test may be performed that could be used as feedback for the algorithm.

[4]Potentially, several models of cancer growth may be used by the algorithm.

Acknowledgements

The authors would like to thank the organizers of the Industrial Problem Solving Workshop for providing them with the opportunity to try their hands in real life situations and contribute to the prosperity of industry. Our gratitude goes also to Dr. Jack Tuszynski for introducing us to this fascinating topic via a very appealing presentation and to Dr. Comron Nouri for his support during the mentioned event. We find ourselves short of words to express our deepest and sincere thanks to Dr. Neville Fowkes, our team mentor, for his generous guidance that stimulated us to work harder and better.

5.7.2 A Model of Drug Resistance in Small Tumours

Dominic Neslon
Shaughnessy Hawkins
Heidi Muller
Mary-Jane Richardson
Justin Schwark

Our focus was on the way tumours develop resistance to chemotherapy drugs, and how treatment plans can be chosen to minimize this effect, specifically those in which two different drugs are used in combination. The way we modeled drug resistance was to assume that each tumour is relatively heterogeneous and so will contain some cells that are killed by a given chemotherapy drug and some cells that are not. With two drugs (say A and B) being administered, we end up with cells that are resistant to neither drug (usually the vast majority), cells that are resistance to drug A, and cell that are resistant to drug B. For simplicity, since we assume that each resistant population is quite small compared to the total population, we ignored the chance of some cells being resistant to both drugs. The model could be easily extended later to include this case if need be. We assume also that the tumours are small enough that it still grows exponentially, and also that exponential death occurs in the presence of chemotherpay drugs which a population is not resistant to. Thus, our model is

$$\frac{d}{dt}N_0(t) = kN_0(t) - f(A)N_0(t) - g(B)N_0(t) \tag{36}$$

$$\frac{d}{dt}N_A(t) = kN_A(t) - g(B)N_A(t) \tag{37}$$

$$\frac{d}{dt}N_B(t) = kN_B(t) - f(A)N_B(t) \tag{38}$$

where N_0, N_A, and N_B represent cells within the tumour resistance to no drugs, drug A, and drug B respectively, k represents the growth rate of the tumour in the absense of drugs, $f(A)$ represents the kill rate of drug A as a function of concentration, and $g(B)$ is the respective kill rate for drug B. Both these functions are time-dependent because the concentration of drug in the body varies in time due to drug administration and natural decay as it is eliminated or used up.

For our preliminary work, we take our kill rates to be

$$f(A) = p_A \frac{2}{\pi} \arctan(A(t)) \tag{39}$$

$$g(A) = p_B \frac{2}{\pi} \arctan(B(t)) \tag{40}$$

where $A(t)$ and $B(t)$ are the concentrations of drugs A and B in the body, and we can adjust the relative strengths of the drugs by choosing appropriate values for p_A and p_B. These values have not yet been fitted with real clinical data, but this could easily be done in the future. When we ran the model, we found that the timing of drug application could make a significant difference in the size of the tumour. An example is shown in Figure 25.

This figure compares two different drug regimens, with the number of cells in the tumour plotted against time (measured in hours). One regimen gives both drugs at the same time over intervals of 36 hours, and another where drug A is given over intervals of 36 hours and drug B is given over intervals of 24 hours, offset 12 hours from the beginning of taking drug A. The results so far show that taking as much drug as early as possible is the best strategy, but since our model does not yet include side effects of the drugs, this result might not remain optimal.

Future work on this model could incorporate realistic growth rates for tumours large enough to have halted their exponential growth, the addition of side effects to include some measure of patient quality of life, and an examination of the effect of mutation within the tumour on the development of drug resistance.

5.7.3 Deterministic Mechanistic Model

Stephen Hudson	University of Western Ontario
Parisa Hudson	University of Western Ontario
Shannon Collinson	
Jiyung	

Modelling has become an important tool for simulating tumour growth

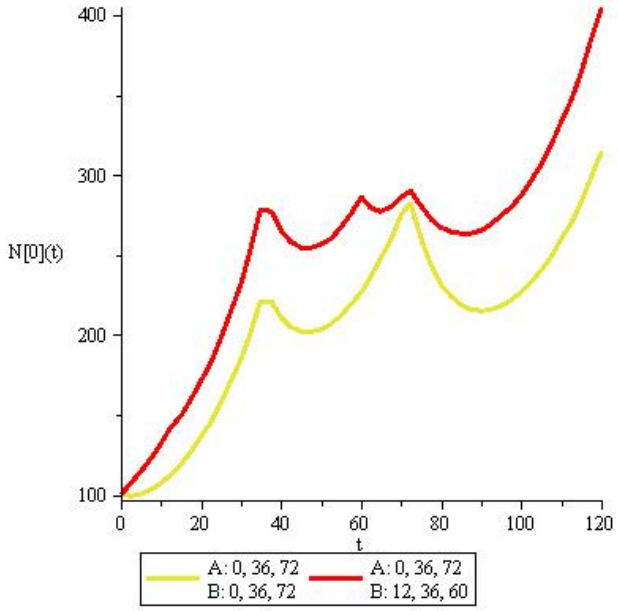

Figure 31: Response to chemotherapy of a non-resistant cell population within a tumour

and the application of treatments. For example coupled reaction-diffusion equations have been used for optimizing the combination of anti-angiogenic therapy and chemotherapy in a generic tumour [**?**], for optimizing radiotherapy and chemotherapy in brain tumours [**?**], and for optimizing the sequencing of surgery and chemotherapy for ovarian cancer [**?**].

Diffusion Model

The model used by Powathil et. al consists of a reaction diffusion equation for the concentration of tumour cells in the form

$$\frac{\partial c(\vec{x},t)}{\partial t} = \cdot(D(\vec{x})c(\vec{x},t)) + f(c) - R(c,t) - G(c,t) \tag{41}$$

where $c(\vec{x},t)$ is the density of tumour cells, $D(\vec{x})$ is the diffusion coefficient, $f(c)$ is a tumour cell growth term, $R(c,t)$ describes the effects of radiotherapy, and $G(c,t)$ describes the effects of chemotherapy [**?**]. For simple exponential growth, $f(c) = \rho c$ could be used, but Powathil et. al use $f(c) = \rho c(1 - c/c_{\text{lim}})$ where c_{lim} is called the carrying capacity, because the non-zero volume of tumour cells limits the maximum concentration. Even with the limiting concentration, this model will show simple exponential growth over a long period of time.

Dose Scheduling Investigation

Using the diffusion model described above with no radiotherapy, we kept the total dose of chemotherapy drugs constant while changing the dose schedules. In each case, drug was administered on 30 days, each day for 6 hours. For each of those days, 1 6-hour dose, 2 3-hour doses, 3 2-hour doses, 6-1 hour doses, or no dose was given.

Fig. 32 shows tumour size versus time when drug is administered for 30 days in a row. Fig. 33 shows tumour size versus time when drug is administered in 2 15-day segments with a 7-day break in between. Fig. 34 shows tumour size versus time when drug is administered in 3 10-day segments with 7-day breaks in between.

The effectiveness of each treatment schedule can be determined from the additional time it takes the tumour to reach a specific size compared to an untreated tumour. The figures show that for 30 days of drug administration, the tumour growth is impeded by approximately 60 days. The difference between any two dose schedules is about 10 days or less.

Fig. 35 shows the best treatment schedules from Figs. 32, 33, and 34. It is clear that changing the dose schedule while holding the total amount of drug constant has little effect on the total number of days of tumour

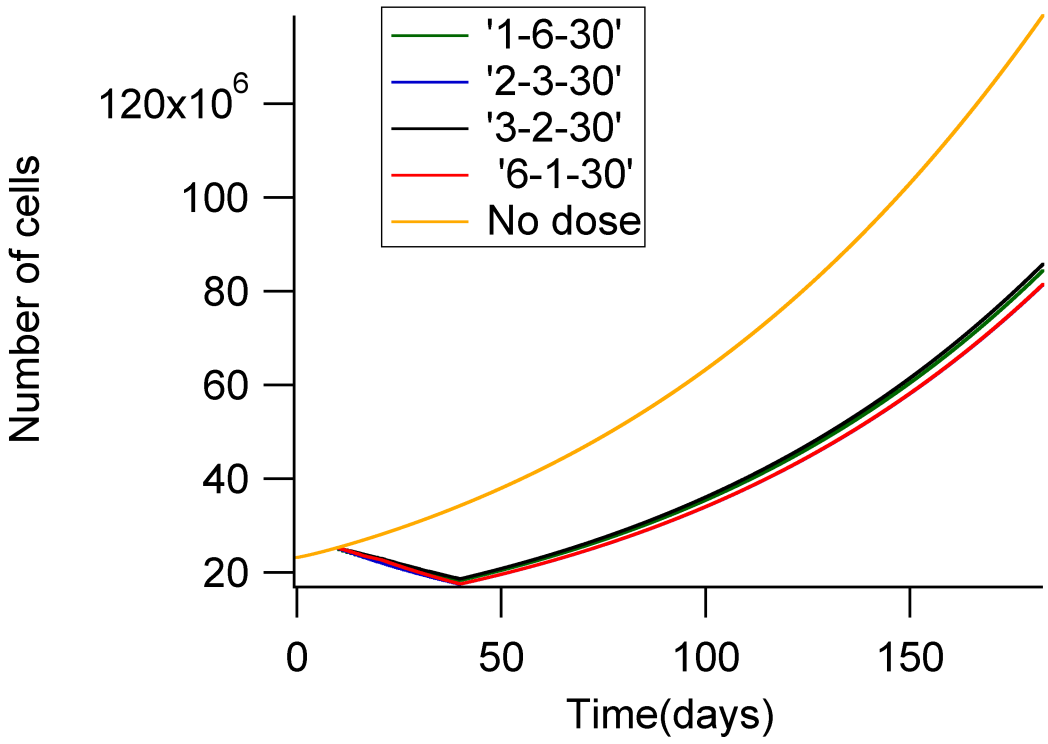

Figure 32: Tumour size versus time in days is shown for the following dose schedules: 1 6-hour administration per day (green), 2 3-hour administrations per day (blue), 3 2-hour administrations per day (black), 6 1-hour administrations (red), and no drug administered (yellow). The drug administration was repeated 30 days in a row.

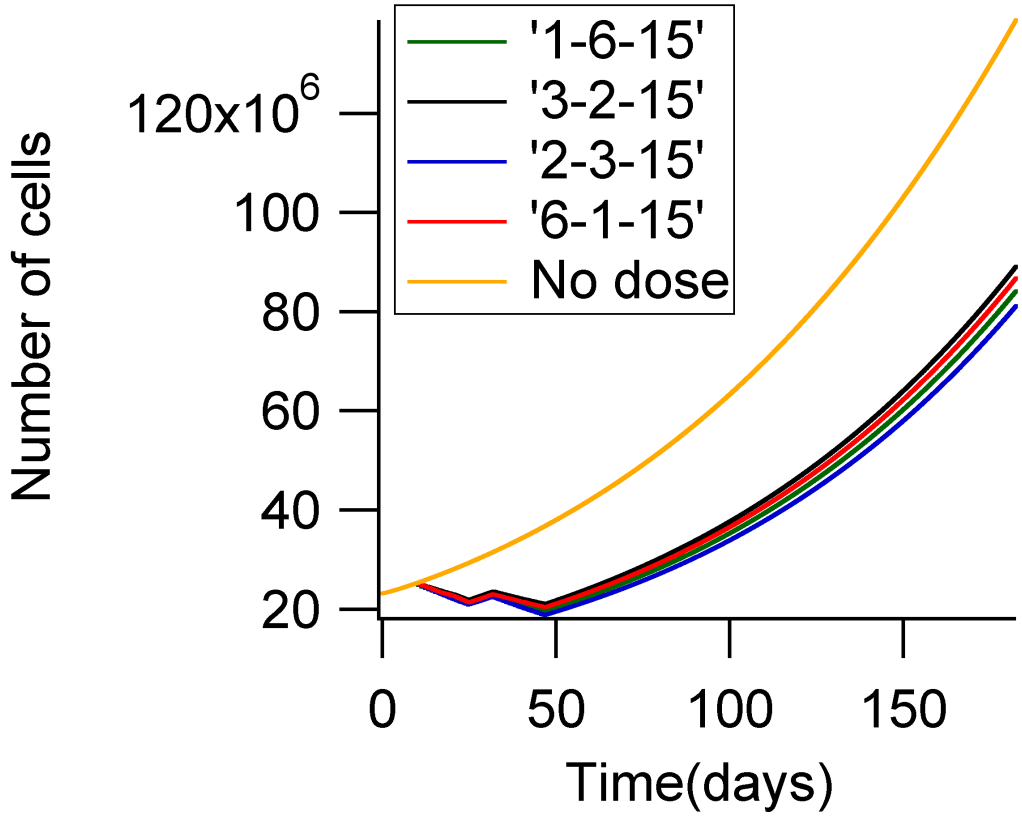

Figure 33: Tumour size versus time in days is shown for the following dose schedules: 1 6-hour administration per day (green), 2 3-hour administrations per day (blue), 3 2-hour administrations per day (black), 6 1-hour administrations (red), and no drug administered (yellow). The drug was administered for 30 days total, but treatment was divided into 2 15-day segments with a 7-day break in between.

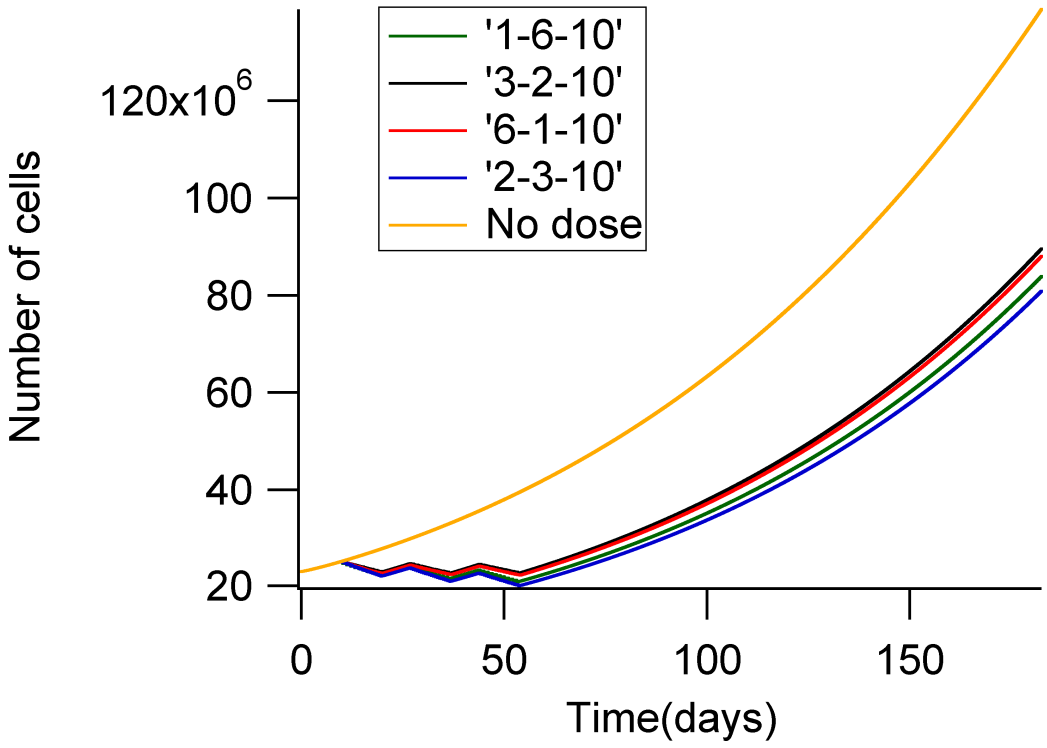

Figure 34: Tumour size versus time in days is shown for the following dose schedules: 1 6-hour administration per day (green), 2 3-hour administrations per day (blue), 3 2-hour administrations per day (black), 6 1-hour administrations (red), and no drug administered (yellow). The drug was administered for 30 days total, but treatment was divided into 3 10-day segments with 7-day breaks in between.

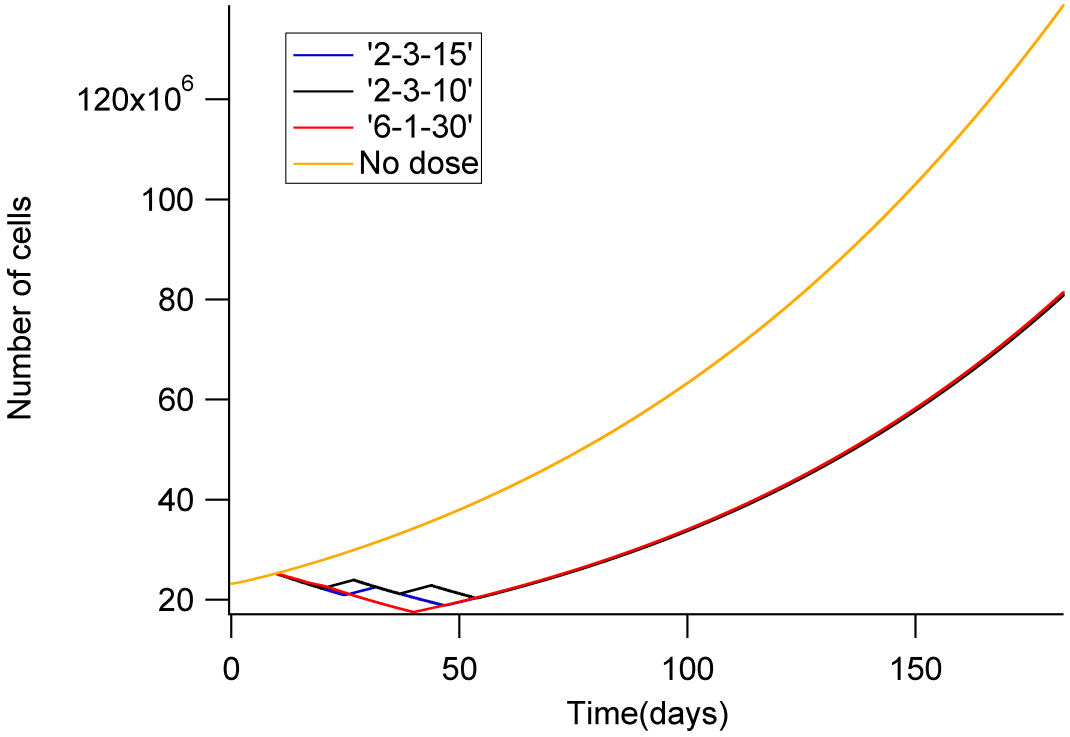

Figure 35: The best treatment schedules from Figs. 32, 33, and 34. The yellow curve shows no treatment for comparison.

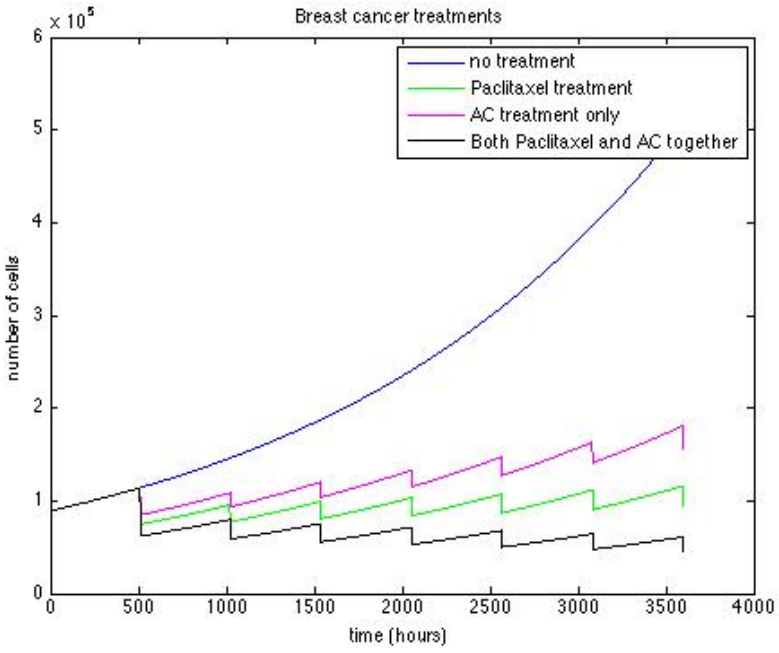

Figure 36: Tumour size versus time for growth without treatment, with Doxorubicin (AC) alone, with Paclitaxel alone, and with a combination of Doxorubicin and Paclitaxel.

growth delay. However, it may be possible that some dose schedules will allow a patient to tolerate larger total doses, resulting in more significant delay of tumour growth. In the limited time available, we were not able to investigate this hypothesis.

5.8 Combining Drugs for Breast Cancer Treatment

In this section, we investigate the combination of two drugs that are commonly used for treating breast cancer. Fig. 36 shows tumour size versus time for growth without treatment, with Doxorubicin alone, with Paclitaxel alone, and with a combination of Doxorubicin and Paclitaxel. The figure shows that Doxorubicin and Paclitaxel administered together gives the greatest reduction in tumour size.

Given that combining the two drugs is most effective, we wanted to determine the best method of combining the drugs. We found that greatest

reduction in tumour size occurs when the two drugs are given at the same time.

Incorporating Angiogenesis

Through diffusion of oxygen alone, a tumour will grow to a size of 1–2 in diameter. To grow further, the tumour needs to induce a process called angiogenesis, which is the growth of new blood vessels towards the tumour from existing blood vessels. When blood vessels begin to feed oxygen and nutrients to a tumour, it will undergo rapid growth and quickly progress towards metastasis.

Clearly, a treatment that hinders angiogenesis will slow tumour growth. To incorporate angiogenesis, Kohandel et. al coupled Eq. 41 with an equation that evolves into a distribution of islands that represent blood vessels:

$$\frac{\partial m(\vec{x},t)}{\partial t} = D_2\nabla^2 m(\vec{x},t) + m(\vec{x},t)\left[\alpha + \beta m(\vec{x},t) + \gamma m(\vec{x},t)^2\right] \qquad (42)$$

and an equation that introduces oxygen and nutrients from the blood vessels and describes their diffusion:

$$\frac{\partial K}{\partial t} = D_3\nabla^2 K(\vec{x},t) + \delta m \exp\left[-\left(\frac{m}{m_{\text{lim}}}\right)^2\right] - \lambda c(\vec{x},t) - \eta K(\vec{x},t) \qquad (43)$$

where the terms represent, from first to last, diffusion, oxygen introduced by the vessels, consumption by tumour cells.

After metastasis occurs, the probability that a patient will survive cancer drops drastically. For this reason, early detection and treatment of tumours is essential for patient survival. Unfortunately, diagnosis is often made when tumours reach a size of 1 in diameter or larger, well after angiogenesis has occurred. However, the availability and resolution of various imaging modalities is improving, and this will lead to earlier diagnosis. For this reason, accurate modelling of tumour growth before, during, and after angiogenesis is important.

Simulation of Early Tumour Growth

The diffusion equation for angiogenesis used by Kohandel et. al (Eq. 42) is independent of the tumour concentration and evolves into vessels regardless of the state of the tumour. We chose a different approach where angiogenesis is triggered when the tumour becomes large enough that diffusion of oxygen becomes insufficient to feed the tumour. The concentrations

used are c for tumour cells, K for oxygen and neutrients, m for blood vessels, c_{taf} for tumour aniogenic factors, and c_{ec} for endothelial cells and blood vessels.

Our diffusion equation for tumour cells is

$$\frac{\partial c}{\partial t} = D_1 \nabla^2 c + A(K)c\left(\frac{K}{K_{lim}} - 1\right) \tag{44}$$

where K_{lim} is the minimum oxygen concentration needed for survival. When the oxygen concentration is above the limit tumour cells divide, and when the oxygen concentration is below the limit tumour cells die.

Our diffusion equation for oxygen and neutrients is

$$\frac{\partial K}{\partial t} = D_2 \nabla^2 K - \lambda c \ - \eta K + A_{ec} c_{ec} \tag{45}$$

where $-\lambda c$ describes consumption of oxygen by tumour cells, $-\eta K$ describes decay of oxygen, and $A_{ec}c_{ec}$ describes oxygen delivered by blood vessels created by endothelial cells.

Similarly, we have a diffusion equation for tumour angiogenic growth factors where the factor is created when tumour cells die and a diffusion equation for endothelial cells where cells will begin to migrate from an existing blood vessel at the boundary when the concentration of the growth factor surpasses a threshold value.

Fig. 37 shows tumour size versus time for the model described above. At early times, oxygen is plentiful and diffusion of oxygen from surrounding tissue is adequate to allow exponential tumour growth. Before a time of 500 days, the tumour becomes large enough that cells near the centre become oxygen starved. The tumour growth becomes linear as cells near the outside continue to receive oxygen, while cells in the centre do not. At this time, a tumour angiogenic factor (TAF) is created and begins to defuse outwards. The TAF induces endothelial cells to migrate towards the tumour and to create new blood vessels. At a time of approximately 1500 days, the new blood vessels provide enough oxygen that the tumour begins to grow exponentially again. Fig. **??** shows snapshots of the tumour at three different times: 150 days (early exponential growth before angiogenesis), 750 days (retarded linear growth, angiogenesis begins), and 1750 days (exponential growth after angiogenesis).

Fig. 38 shows tumour size versus time when chemotherapy is applied to our model. When chemotherapy was applied before angiogenesis, we found that no benefit was achieved in the long term. The chemotherapy does not reduce the dimensions of the tumour, but reduces the tumour cell

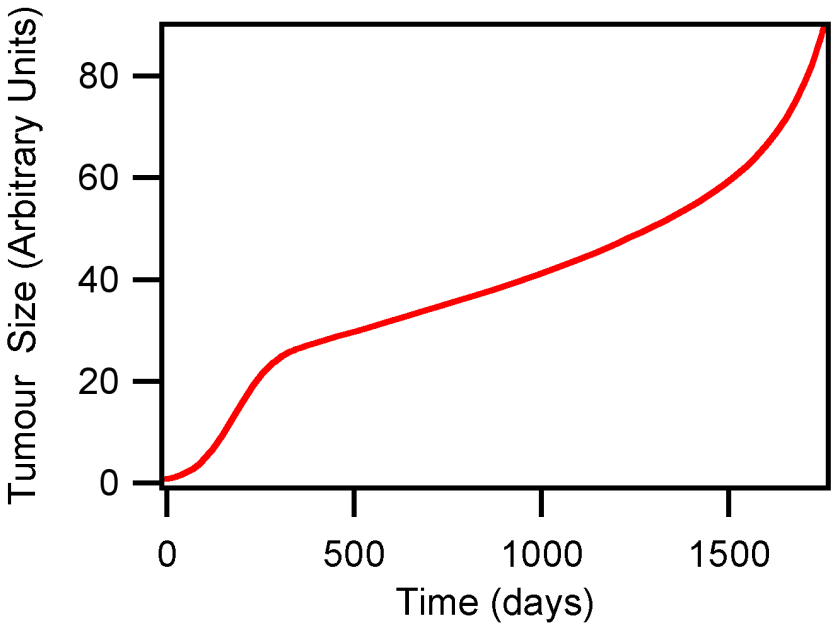

Figure 37: Tumour size versus time for our model.

concentration. As the diffusion process is much slower than the exponential growth, the treated tumour quickly catches up to the untreated tumour while there is an abundance of oxygen.

The model only qualitatively gives more realistic growth and could be improved in several ways. By tuning parameters, more realistic tumour sizes and growth rates could be achieved. Rather than having tumour cells only divide or die, tumour cells could be allowed to survive at low oxygen concentrations but be inactive. The migration of enothelial cells and the assumption that the blood vessel concentration is proportional to the endothelial cell concentration is simplistic. More accurate models for angiogenesis exist, and with some work, they could be incorporated into this model. On online review of angiogenesis modelling is available [**?**].

5.8.1 An Optimal Strategy for Combating Cancer

Janice Cotcher	University of Regina	cotcherj@uregina.ca
Matt Hennessy	Ontario Institute of Technology	matthew.henessy@mycampus.uoit.ca

Introduction

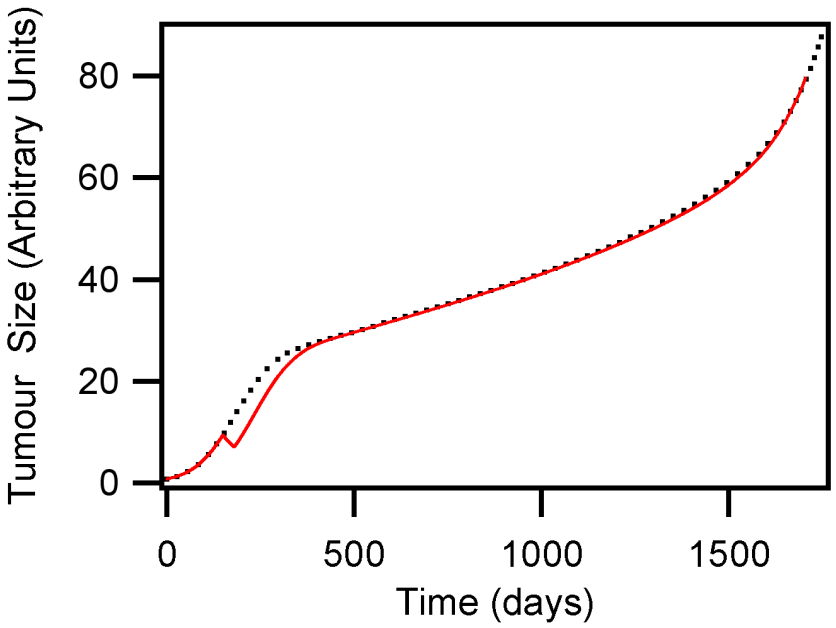

Figure 38: Tumour growth without chemotherapy (dotted line) and with chemotherapy (solid red line).

New hope for tumorous cancers is found in anti-angiogenesis therapies. Unlike chemotherapy, anti-angiogenesis factors already exist within the body and inhibit normal cells. Since these factors do not target the quickly adapting cancerous cells, there is hope there will not be resistance to angiogenic inhibitors.

After a period of avascular growth, a tumour must form its own blood vessels in a process called angiogenesis to develop beyond 2mm. The tumour produces vascular endothelial growth factors (VEGF) which activate cell receptors present in nearby pre-existing veins, thus starting the process of angiogenesis. The activated endothelial cells release an enzyme that degrades the outer membrane of the blood vessel allowing the endothelial cells to escape. The liberated cells undergo mitosis in the surrounding matrix to form solid sprouts extending from nearby pre-existing veins to the tumour. The sprouts hook on to the tumour and then form loops that develop into bloods vessels.

Anti-angiogenesis therapies aim to bind to VEGF so it cannot activate endothelial cells, reducing the number of immature and leaky vessels. The remaining mature tumour vasculature remodels, resulting in a more normalized network. Anti-angiogenesis therapy on its own does deprive the tumour of some nutrients but the greatest benefit may be improved delivery of cytotoxic drugs, like chemotherapy, through the remaining blood vessels to the tumour. Current research is trying to determine the optimal dosage and delivery schedule.

The Model

We model tumour growth in the presence of anti-angiogenic and chemotherapy treatments. The tumour growth can be modelled logistically with a carrying capacity that is related to the number of endothelial cells since they indirectly provide nutrients. Other ways to model the tumour growth include Gompertzian growth as done by Ledzewicz and Schättler, but the essential growth dynamics of these models are identical to ours.

Traditionally, modelling the growth of the endothelial cells includes a term for the natural death of the cells, plus additional terms for stimulation, inhibition, and the death of the cells caused by the anti-angiogenic treatment. These terms usually depend upon both the tumour size and the current population of endothelial cells. However, we use the approach taken by Ledzewicz and Schättler to eliminate the tumour size from the equations that govern the endothelial growth. This is allows for major simplifications because it decouples the endothelial cell growth from the tumour growth.

To model the effects of chemotherapy, we introduce the most basic terms

into the governing equations for cell growth. In particular, we assume the dynamics of the anti-angiogenic and chemotherapy drugs are the same in the equation for the endothelial cells. However, since it is the vasculature that is ultimately responsible for the transport of the chemotherapy drug to the tumour, the effectiveness of the chemotherapy on the tumour is going to depend on the tumour size, the number of endothelial cells, and the amount of drug that is in the body. To model this, we introduce a cubic term into the equation for tumour growth that is trilinear in these three factors.

The time evolution of the drugs is assumed to obey a first order linear ODE with one term representing regular exponential decay and the other term representing an externally administered dosage. Although our model of drug interaction is rather rudimentary, is it possible to extend the model without drastically changing the solution of the optimal control problem. This will be elaborated on below.

Mathematically, the described model can be written as

$$\begin{aligned} \dot{p} &= \xi p(q-p) - k_1 qpc, \\ \dot{q} &= -\mu q + bq^{2/3} - dq^{4/3} - Gaq - k_2 qc, \\ \dot{c} &= -k_3 c + u_c(t), \\ \dot{a} &= -k_4 a + u_a(t), \end{aligned} \tag{46}$$

where the overdot represents differentiation with respect to time and p is the tumour size, q is the number of endothelial cells, and c and a are the amount of drug related to chemotherapy and anti-angiogenesis, respectively. We impose the initial conditions $p(0) = p_0 > 0$, $q(0) = q_0 > 0$, $c(0) = c_0$, and $a(0) = a_0$. The two functions $u_c(t)$ and $u_a(t)$ are the administered dosages that are to be optimized. A summary of each term can be found in Table 14. Parameter values can be found in Table 15.

Optimal Control

We seek the optimal dosage schedule for administering anti-angiogenic and chemotherapeutic drugs while being mindful of the consequences for the healthy cells of the patient. This task is indeed possible with the use of control theory.

To be precise, we aim to find the optimal dosages $u_a(t)$ and $u_c(t)$ in order to minimize the tumour growth at some time T in the future. However, we also try to minimize the amount of drug that is required in order the maximize the quality of life for the patient. This problem can be mathematically

Table 14: Summary of each term in the model. Analogous descriptions can be made for the equation governing the angiogenesis treatment.

Tumour growth	
$\xi p(q-p)$	Logistic growth with variable carrying capacity
$-k_1 qpc$	Reduction of tumour due to chemotherapy
Endothelial cell growth	
$-\mu q$	Natural death
$bq^{2/3}$	Stimulation
$-dq^{4/3}$	Inhibition
$-Gaq$	Death due to angiogenesis treatment
$-k_2 qc$	Death due to chemotherapy
Chemotherapy treatment	
$-k_3 c$	Natural decay of drug through body
$u_c(t)$	Externally administered dosage, a control variable

Table 15: Numerical parameter values for the governing equations.

Parameter	Value
ξ	0.084
μ	0.02
b	5.85
d	0.00873
G	0.15
k_1	0.001
k_2	0.001
k_3	0.01
k_4	0.01

modelled by finding the optimal dosages that minimize the functional

$$I[u_a(t), u_c(t)] = \frac{1}{2}\, p^2(T) + \int_0^T \left[u_a^2(t) + u_c^2(t)\right]\, dt, \tag{47}$$

where the first term represents the tumour size at time T and the second term represents the amount of drug used during the treatment. There are many ways that could describe similar quantities but we chose a quadratic functional because of its convexity properties and its susceptibility to analytical analysis.

Thus far our problem is not completely realistic – there is no constraint on how large the dosages can be. If the concentration of a drug is too high there could be serious physical harm done to the patient. Therefore, we impose constraints on the magnitude of the dosages. In particular, we restrict the dosages to the domain

$$\mathcal{D} = \{(u_a, u_c) \in \mathbb{R}^2 : 0 \le u_a(t) \le U_a,\ 0 \le u_c(t) \le U_c,\ \forall t \in \mathbb{R}^+\}.$$

Thus, the constrained optimal control problem that must be solved is

$$\min_{u \in \mathcal{D}} I[u(t)],$$

where $u(t) = [u_a(t), u_c(t)]$ denotes the control variables (dosages). This is a standard optimal control problems that is treated in many texts. The optimal control variables can be found by forming the Hamiltonian function and minimizing it subject to the above constraints ($u \in \mathcal{D}$). If we write the governing equation for the tumour growth as $\dot{p} = f_p(p, q, c)$ and the equation for endothelial cell growth as $\dot{q} = f_q(q, a, c)$, then the Hamiltonian for this system is given by

$$H = u_a^2 + u_c^2 + \lambda_p f_p(p, q, c) + \lambda_q f_q(q, a, c) + \lambda_c(u_c - k_3 c) + \lambda_a(u_a - k_4 a), \tag{48}$$

where the λ_i are *adjoint* variables that satisfy the ODEs

$$\dot{\lambda}_i = -\frac{\partial H}{\partial i}, \quad i = p, q, c, a. \tag{49}$$

Boundary conditions for the adjoint variables can be found by applying Pontryagin's principle. We have $\lambda_p(T) = p(T)$ and $\lambda_i(T) = 0$ for $i = q, c, a$. The boundary condition for λ_p is different because of the way $p(t)$ appears in the functional (47).

As previously mentioned, the optimal controls can be found by minimizing the Hamiltonian subject to the constraints. However, this minimization problem is identical to

$$\min_{u\in\mathcal{D}} \left(\lambda_a u_a + \lambda_a u_c + u_a^2 + u_c^2\right) \tag{50}$$

To solve this constrained nonlinear programming problem and find the optimal controls, we must look at three distinct cases:

1. $\lambda_a > 0$, $\lambda_c > 0$. The minimum of (50) is trivial, for $u_a = u_c = 0$ are the minimizers and thus the optimal control.

2. $\lambda_a > 0$, $\lambda_c < 0$. Here $u_a = 0$ and $u_c = \min\{(-\lambda_c/2)^+, U_c\}$, where

$$x^+ = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

   Notice that the case $\lambda_c > 0$, $\lambda_a < 0$ is analogous to this case.

3. $\lambda_a < 0$, $\lambda_c < 0$, The optimal controls are given by $u_a = \min\{(-\lambda_a/2)^+, U_a\}$ and $u_c = \min\{(-\lambda_c/2)^+, U_c\}$.

The solutions for each of the three cases can, in fact, be combined into a single solution. Thus optimal controls for our model are then given by

$$u_a = \min\{(-\lambda_a/2)^+, U_a\}, \quad u_c = \min\{(-\lambda_c/2)^+, U_c\}. \tag{51}$$

We now make a crucial observation that relates to other possible models. *As long as the governing equations and the objective functional have the same dependence on the control variables as in* (46) *and* (47)*, the optimal controls given above in* (51) *will be "model independent"*. That is, by incorporating the control parameters into linear ODEs as done in (46), we can modify the equations for endothelial and tumour growth without having to redo the optimization. The control parameters do not directly enter these equations so the objective function in (50) does not change. This observation implies that our crude model can be easily extended to accommodate more realistic drug interactions. It should be made clear that the actual $u_i(t)$ will change because they depend on the $\lambda_i(t)$, which *are* model dependent, since $\lambda_i = -\partial H/\partial i$.

Unfortunately, since the governing equations are nonlinear, the above equations (51) are a close approximation to an analytical solution for the optimal controls. Since the form of optimal controls are known, we can

insert them into (46) and numerically solve the boundary value problem. Written explicitly, the boundary value problem that must be solved is

$$\begin{aligned}
\dot{p} &= \xi p(q-p) - k_1 qpc, \\
\dot{q} &= -\mu q + bq^{2/3} - dq^{4/3} - Gaq - k_2 qc, \\
\dot{c} &= -k_3 c + \min\{(-\lambda_c/2)^+, U_c\}, \\
\dot{a} &= -k_4 a + \min\{(-\lambda_a/2)^+, U_a\}, \\
\dot{\lambda}_p &= \lambda_p(2\xi p - \xi q + k_1 qc), \\
\dot{\lambda}_q &= \lambda_p(k_1 pc - \xi p) + \lambda_q \left(\mu - \frac{2}{3}\, bq^{-1/3} + \frac{4}{3}\, dq^{1/3} + Ga + k_2 c\right), \\
\dot{\lambda}_c &= \lambda_p k_1 pq + \lambda_q k_2 q + \lambda_c k_3, \\
\dot{\lambda}_a &= \lambda_q Gq + \lambda_a k_4.
\end{aligned} \tag{52}$$

with endpoint conditions

$$\begin{aligned}
&p(0) = p_0,\ q(0) = q_0,\ c(0) = c_0,\ a(0) = a_0, \\
&\lambda_p(T) = p(T),\ \lambda_q(T) = 0,\ \lambda_c(T) = 0,\ \lambda_a(T) = 0.
\end{aligned} \tag{53}$$

We attempted to solve this boundary value problem using a shooting method. The main idea behind this method is that the boundary value problem is transformed into an initial value problem by "shooting" (performing a time integration) for the correct endpoint conditions using iteratively computed initial conditions. This method is known to be unstable, and in fact, we had limited success when it was applied to this problem. When trying to find the optimal solution over a long time interval (365 days), the time integrations diverge and the shooting method fails completely. Thus, the obtained results are limited to short time intervals of seven days as demonstrated in Figure 39.

To relieve some of the numerical difficulties, it may be be possible to rescale the equations to minimize the stiffness of the system. In addition, higher order numerical differential equation solvers could be implemented with an adaptive stepsize. Also, a more stable numerical method could be implemented for the solution of the boundary value problem. Perhaps the most trivial method to try would be a finite difference method, but this could fail because it assumes solutions have a certain degree of regularity (recall that the equations that govern the drugs are only piecewise differentiable). Instead of trying finite differences, multiple shooting could be attempted. This involves breaking up the time interval and performing shooting over each subinterval. However, multiple shooting can have the same shortcomings as regular shooting.

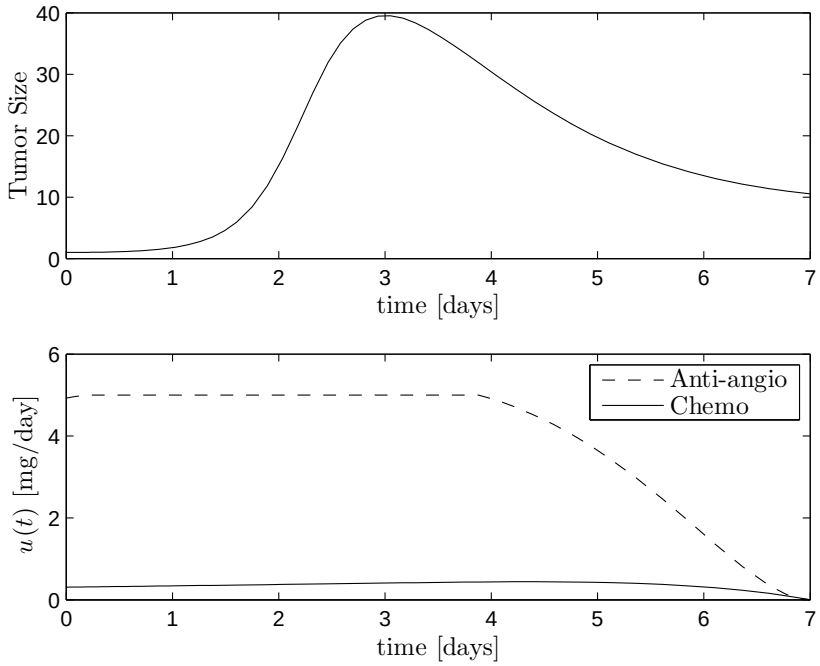

Figure 39: Example of the tumour growth (top) with the optimal treatment (bottom). The calculations were done using the initial conditions $p(0) = q(0) = 1$ and $c(0) = a(0) = 0$. The upper bound on dosages were $U_a = 5$ mg/day and $U_c = 1$ mg/day.

Discussion and Conclusions

We have formulated a rough mathematical model of cancerous cell development in an environment with anti-angiogeneic and chemotherapeutic treatments. Using optimal control theory we determined the *best theoretical* dosage schedules that will minimize the tumour size while using a minimal amount of drugs. Future work not only includes implementing improved numerical methods, but also developing a more realistic model of drug interaction. This could be done by looking at experimental data or developing a microscopic model of how the drug interacts with cells.

Another approach would be to compare results with different objective functionals. For example, coefficients could be introduced in front of the terms in (47) to describe how one treatment could be more helpful than the other. Also, the drugs could be dropped from the functional altogether and a new equation that models the patient health could be introduced. New constraints could be added that state the health cannot fall below a certain level. Regardless of the model, it is our hope that optimal control theory will provide new treatments for cancer patients that will improve their quality of life while still effectively reducing the tumour size.