Mathematics-in-Industry NZ 2017
Challenge Report:

Equation-Free Summaries

Co-Directors: Dr Luke Fullard
Dr Richard Brown, Massey University

Deputy Directors: Prof Graeme Wake, Massey University
Lisa Hall, Fonterra

Administrators: Seumas McCroskery, KiwiNet Innovation Network

Plenary Speaker: Dr Melanie Roberts, IBM Australia

Guest Speakers: Prof Jan Thomas, Vice-Chancellor, Massey University

Organising Committee:
Dr Tammy Lynch
Christine Burr
Dr Cami Sawyer
Dr David Simpson
Dr Chris Tuffley
Assoc Prof Brice van-Brunt
Prof Igor Boglaev
Prof Robert McLachlan
Dr Rose Davies

Foreword

Mathematics is more relevant today as it has ever been. Educators looking for ways to inspire the youth of today in the importance of maths should look no further than the Mathematics in Industry NZ study week. Now in its third year, it is being held at Massey University, Palmerston North on June 26th-30th. The study week concept has been going now for over half a century around the world.

This national event was established to add value to our community and our industry as well as provide academic opportunities for many of us. We warmly acknowledge support from all our sponsors, but especially KiwiNet: a consortium established to foster industry links with experts such as those in the mathematics community. KiwiNet continues to provide the administrative structure to make this event happen.

We have six exciting challenges put forward to the mathematical group from six dynamic and important companies: Fonterra, Transpower, Horizons Regional Council, Sanford, Zespri, and Fisher & Paykel Appliances, it is a pleasing mix of those that have taken part in similar events and those new to the study group concept.

We were very pleased to welcome many participants from around New Zealand and further afield. One such guest, Dr Melanie Roberts, an applied mathematician working at IBM Research Australia. It was a delight to have her here, delivering an exceptional plenary talk, contributions both formal and informally throughout the week.

It was a great honor to also welcome Professor Jan Thomas, the newly appointed Vice Chancellor of Massey University, who graciously accepted our invitation to open MINZ 2017.

Co-Directors:

Dr Luke Fullard, and Dr Richard Brown,

2017

Challenges- for more information see http://www.minz.org.nz

Challenge 1:
Fonterra

Optimising the flavour profile and longevity of milk powders

Full Challenge Details

Challenge 2:
Fisher & Paykel Appliances

Modelling the mechanical action of a front loading washing machine

Full Challenge Details

Challenge 3:
Horizons Regional Council

How could we best optimise our Regions freshwater monitoring networks

Full Challenge Details

Challenge 4:
Zespri

Predicting timing of kiwifruit harvest

Full Challenge Details

Challenge 5:
Sanford Ltd

Comparing and contrasting shear forces and hydrodynamics of Industrial Larval Mussel tank design and operation

Full Challenge Details

Challenge 6:
Transpower

Transmission Line Conductors - Big Data Cleansing, Probability of Failure Derivation and Asset Health Relationship

Full Challenge Details

Challenge from Fonterra:

Optimising the flavour profile and longevity of milkpowders

Industry Representatives:
Lisa Hall, Roger Kissling and Grant Abernethy

Challenge Moderators:
Tammy Lynch, Massey University, Palmerston North, New Zealand
Steve Taylor, University of Auckland, Auckland, New Zealand

Student Moderator:
Valerie Chopovda, Massey University, Albany, New Zealand

The flavour of milk powder can change over time, especially when exposed to varied conditions during shipment around the world. Oxidation of lipids in the powder contributes to flavour changes. Oxidation produces volatile organic compounds (VOCs), and it is believed that the presence of some of these VOCs in the powder is an indicator of the quality of milk flavour.

Fonterra provided a dataset of concentrations of these VOCs for samples of whole milk pow- der stored for two months post manufacture. The measurements were taken with a Selected Ion Flow Tube Mass Spectrometer (SIFT-MS). The same milk powder was also subjected to a pass/marginal/fail sensory test conducted by a trained sensory panel.

Our goals for this project were to determine which VOCs contribute to a failed sensory test and to model the concentrations of these VOCs over the shelf life of the powder. If successful, this research could allow Fonterra to supplement or replace the sensory tests with SIFT-MS measurements. Further, the mathematical model could be used to quantify the sensory profile over time and predict whether or not powder would fail the sensory test during the required shelf life.

Our group used statistical methods to analyse the data supplied by Fonterra. Our initial analysis used a multinomial model that achieved 90 to 95% accuracy. This analysis was done with a full multinomial regression Artificial Neural Network model which identified a list of twenty VOCs that contributed most to the model. This is a promising result, but it is limited by the fact that most of Fonterra's powders do not fail a sensory test, and thus the data did not contain enough failures for us to reliably predict failures or quantify the VOC profile of failed samples.

We then used a general linear model based on the twenty variables that we had identified and ran a step AIC process to reduce the list of twenty down to six. "AIC" is the Akaike information criterion, an estimator of the relative quality of statistical models for a given set of data. It is used to reduce the number of variables needed to model the data.

The step AIC process takes various variable interactions into account. We found that the step AIC process did not converge with all interactions, but including some interactions improved the fit. Notably all significant interactions include ethyl decanoate. It is currently uncertain whether there is scientific logic behind this.

The mathematical models formulated by the group took the form of systems of ordinary differential equations. We formulated two different models, the first being a system of eight differential equations for the peroxidation of a general lipid group. The model is based on the fact that all lipid peroxidations follow the same general mechanism, as shown in Figure 1.

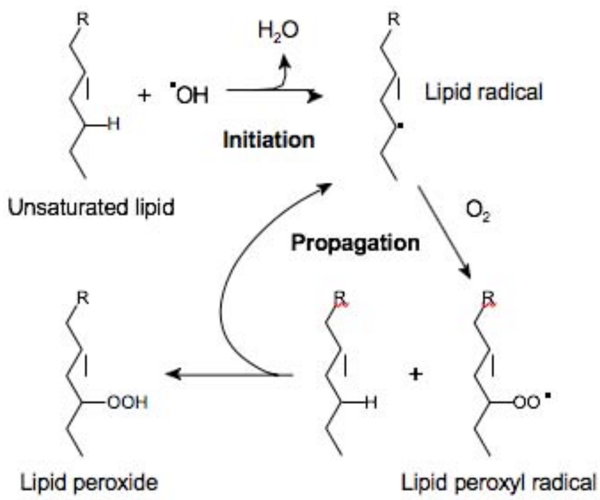


Figure 1: Mechanism of lipid peroxidation. Diagram by Tim Vickers, after Young IS, McEneny J (2001). "Lipoprotein oxidation and atherosclerosis". Biochem Soc Trans 29 (Pt 2): 358–62. PMID 11356183. Vectorized by Fvasconcellos. [Public domain], via Wikimedia Commons.

This model allows calculation of the progress of the lipid peroxidation reactions based on a knowledge of the concentrations of the reactants at an initial time. It also depends on several reaction rate constants that need to be determined for the particular lipid involved.

The second model formulated by the group is based on the assumption that in milk powder these reactions will rarely continue to completion during the shelf life of the powder. This allows us to simplify the model. In particular, the lipid concentration will not change much during the shelf life. Our simplified model consisted of just

The second model formulated by the group is based on the assumption that in milk powder these reactions will rarely continue to completion during the shelf life of the powder. This allows us to simplify the model. In particular, the lipid concentration will not change much during the shelf life. Our simplified model consisted of just two ordinary differential equations involving only two rate constants. These rate constants will depend on that particular lipid being considered. We computed solutions of the model for the peroxidation of hexanal and found that our model matched experimental data well.

In summary, our group focused on two sub-projects. The first was to match SIFT-MS data for concentrations of VOCs to sensory test data. The second was to model the evolution of these VOCs over time. Combining the results of these two projects has the potential to allow Fonterra to predict the sensory acceptance of milk powders over time. Completion of this work will require more data so that the VOC concentration profile can be more reliably matched to sensory test results. In particular, more data for powder that leads to failed sensory tests is needed. Such data was lacking in our work, for the simple fact the majority of Fonterra's milk powder passes sensory tests. Such an analysis of new data will also tell us which lipid peroxidation reactions are important, allowing us to determine the relevant rate constants.

Challenge from Fisher & Paykel Ltd.

Modelling the mechanical action of a front loading washing machine

Industry Representatives:

Kirsty Davies and Jennifer Trittschuh

Challenge Moderators:

Melanie E. Roberts, IBM Research - Australia

Celia Kueh, Massey University

Student Moderator:

Emma Greenbank, Victoria University of Wellington

Background

Fisher & Paykel Appliances is a major New Zealand appliance manufacturing company who, amongst other products, develops front-loading washing machines. Their research and development teams are committed to improving washing machine performance, focussing on the consumer experience in addition to standardised testing. Machine performance is a combination of the wash performance, water and power consumption, and ease of use for the customer. From a consumer perspective, the key criteria for wash performance are the degree to which clothes are cleaned, the soil removal, and the degree of wear and tear on clothing due to washing action.

The clothes washing process removes dirt and grease-like products through a combination of chemical, thermal and mechanical actions. To a degree, these processes can compensate for each other, for example additional deter- gent can be used to compensate for cooler washing temperatures or a shorter wash cycle. The challenge posed to Mathematics-in-Industry New Zealand was to investigate mathematical models for the wash performance of front- loading washing machines due to mechanical action. The mechanical actions for soil removal are garment-to-garment rubbing, within-garment rubbing, and garment-to-washing machine drum-skin rubbing. These actions are influenced by how clothes move around within the washing machine, and thus Fisher & Paykel can modify a number of parameters of front-loading machines and wash cycles to change the wash performance. Parameters able to be altered include: drum dimensions, drum skin profile and drum ends, vane placement and geometry, door protrusion, drum speed, rotation schedule, wash time, and water volume. Other factors that influence wash performance due to mechanical action, within the consumer's control, include the volume of clothes washed, mix and type of clothing washed (e.g. towels vs. delicates), how the machine is loaded, and the wash program. By fixing the water temperature, detergent used and concentration, and loading factors (load type, loading method etc.), Fisher & Paykel are able to investigate the effect of varying parameters on the wash performance.

Wash performance testing is performed under controlled conditions; wash temperature, detergent concentration, load composition, load size, soil type and wash program are held fixed to investigate the impact of varying specific machine parameters. Wash performance is measured using two criteria: soil removal (SR) and Gentleness of Action (GA). To measure wash performance, swatches of special fabrics are attached to garments to provide estimates of soil removal and gentleness of action. The SR and GA is scored using quantitative measures on a scale of 0 to 1. For SR, scores closer to 1 are desirable

as this corresponds to a high rate of soil removal, whereas for GA low scores are preferred, as a score of 0 corresponds to no wear on the garment. The evenness of the wash is also a performance measure, with a low standard deviation in the SR and GA scores across swatches in a load being desirable. For the workshop Fisher & Paykel Appliances made available results from wash performance testing with a prototype machine to inform the investigations. The team focussed on three approaches to understand the impact of machine parameters on the wash performance: data analysis, physical modelling, and image analytics. These approaches are summarised in the below sections.

Data Analysis and Physical Modelling

Fisher & Paykel provided summaries of in-house testing, together with some original data sets, relating wash performance (SR and GA scores) to a number of investigated parameters. Furthermore, they shared their intuition, based on their studies, about the expected relationships between various parameters and the wash performance. One of the datasets provided by Fisher & Paykel Appliances showed experimental results for SR and GA due to variations in the drum speed for a standard test rig. This experiment was duplicated for two fixed parameters, the time duration of the test cycle and the number of drum rotations within the cycle. A machine cycle is the time that a drum rotates in one direction before switching to the other direction. Analysis of this data indicated that:

- SR and GA were not correlated
- landing location and departure location within the drum are correlated to the drum speed
- SR varies with the drum speed, and hence landing angle, however there is high variability within the test results at each speed

- GA varies with the drum speed.

Further investigation of this data failed to elicit a relationship that could be used to build a mathematical model of the wash performance. However, knowledge of washing mechanisms provides insight into the physical features of a wash that will increase the performance. The action of dropping the clothing to the base of the drum is known to facilitate soil removal. In a front-loading washing machine, vanes fixed to the side of the machine pick up clothing as the drum rotates. At low spin speeds, when the centrifugal force is less than the gravitational force, the clothing is carried around the side of the drum before falling to the base of the drum to be picked up by a subsequent vane. The point on the drum where the clothing leaves the side of the drum is referred to as the departure angle, while the point of the drum where the clothing lands is referred to as the landing angle. Approximating an item of clothing as a point mass on the perimeter of the drum, a mathematical model for the movement of clothing within the drum was developed, identifying the departure angle, landing angle, and landing speed as a function of the drum speed and drum radius. A balance of centrifugal and gravitational weight forces acting on the mass rotating in the drum, discounting friction, gives the relationship between departure angle and drum speed and radius. Assuming free-fall, the trajectory after departure of the mass and subsequent landing position in the drum was determined. The velocity of the mass on landing was also defined, with the horizontal velocity component being the same as on departure, while the vertical component was the same as on departure minus the velocity difference due to gravitational acceleration during the free-fall time. This model identified a quadratic relationship between the drum speed and departure angle and a quadratic relationship with sixth order correction for the landing angle. A sensitivity analysis of the motion to varying drum speed and drum radius was conducted by setting the drum speed to 40 rpm, 45 rpm and 50 rpm and drum radius to 0.2 m, 0.248 m and 0.3 m. To include the possibility of a vane deviating the angle of the clothing mass departure velocity away from the tangential direction, trajectories were generated with modified angles of the departure velocity.

Comparison of the mechanistic model with the experimental data showed that while the general shape of the function was representative of that observed, the function values were inaccurate. Regression methods were used to fit a quadratic relationship to the data for both the landing and departure angles, with very high agreement. This relationship was subsequently used to identify a relationship between vane position and drum speed in order to minimise the time clothes roll along the drum before being lifted by the next vane. This relationship indicates a potential improvement to wash performance may be achieved by investigating alternative vane arrangements. In parallel to the mathematical modelling work, a numerical model of the clothing motion in the drum was developed using the Discrete Element Method (DEM). The DEM equations were numerically integrated in

LIGGGHTS open source DEM particle simulation software. The drum geometry was generated in SolidWorks CAD software to give a Stereolithography STL file format mesh of the drum imported into the DEM particle simulation software. Multiple items of clothing were approximated as soft inelastic spheres in the drum and their trajectories around the washing ma- chine were numerically modelled. Three cases were modelled, where the drum was filled with 13, 25 and 39 spheres to approximate different sizes of washing load. The trajectories of the clothes depend on the load size. The numerical model shows that the mean clothing departure angle increases with load size, as the leading article of clothing is pushed further up the drum by the following articles of clothing. Velocity fields of the clothing spheres in the drum were generated which showed regions of varying velocity within the drum. The velocity maps can be used to determine whether there are points where the clothes have a low velocity or are trapped causing a poor wash performance. This can be compared with the velocity fields produced by the image analysis or with velocity fields produced by other methods, such as radioactive tracking.

Image Analytics

With the expectation that few, if any, of the study group would have experience with the washing action of a front-loading machine, Fisher & Paykel Appliances provided five videos shot using a handheld mobile phone camera so that study group participants could see first-hand how clothes moved within the machine. Given the challenges associated with direct measurement of washing action (Positron Emission Particle Tracking, which requires a radioactive particle, was used in one study), image analytics techniques can provide a proxy from which valuable information can be derived. A soft- ware prototype was developed that analyses the washing videos to identify the velocity field evident through the clear door of the machine.

The velocity field is calculated by tracking the movement of features within the image between frames in order to calculate the velocity at various locations visible through the door. These velocities were then averaged over a 60 second period to identify a representative velocity field. Video quality introduced a number of challenges for this analysis, in particular the camera's light reflecting off the washing machine door, and the movement of the camera during the recording. A tripod to stabilise the camera, indirect lighting to avoid reflection, and a higher frame rate, would alleviate many of the challenges for the image analytics. This analysis showed variations in the velocity fields between different wash parameters, for example load size, and we therefore expect that such analysis would assist Fisher & Paykel Appliances in their prototyping.

Conclusions and Recommendations

The Mathematics-in-Industry New Zealand workshop identified a number of wash performance factors that will assist Fisher & Paykel Appliances in their development of front-loading washing machines. Although it was expected that soil removal (SR) and gentleness of action (GA) would be negatively correlated, analysis of the data failed to illicit such a relationship. This suggests that Fisher & Paykel Appliances can identify mechanisms to in- crease the SR, as desired, without unduly compromising the GA. Informed by the physical model for the departure and landing angles of clothing, a data driven model for these parameters was developed. Together, these two models provide information on how variation in the drum radius and speed will influence the movement of the load within the drum. This information can be used to increase SR as the landing speed, which is related to the departure and landing angles, is a factor in SR. Moreover, these models have suggested that variation in vane placement may be beneficial. The image analytics and numerical analysis have provided insights into the variation in load velocity within the drum, identifying regions of higher and lower velocity. It is hypothesised that this leads to variation within the observed SR and GA, although further investigation is required. Improvements in video quality would enable image analytics to be more generally employed. Further investigation into the impact of load size is also recommended, as initial investigations, via the numerical analysis, suggests that load size directly impacts the departure angle and hence SR through the mechanical action.

One challenge in identifying a mathematical model of the mechanical action of a washing machine for this project was limited understanding of how different factors influence the SR and GA. It is recommended that Fisher & Paykel Appliances standardise the reporting of experiments to ensure that results can be more widely used.

Challenge from Horizons

How do we best optimise our Regions freshwater monitoring networks

Industry Representatives:

Stacey Binsted, Abby Matthews and Manas Chakraborty

Challenge Moderators:

Jamas Enright, Statistics New Zealand

Stephen Marsland, Massey University

Student Moderator:

Alex White, Massey University

Introduction

In common with the other regional councils, Horizons (the regional council for the Manawatu-Wanganui region) routinely monitor the state of rivers, lakes, and streams across the region. The condition of our fresh water has been of great public interest over recent months, particularly since the New Zealand Government released their 'Clean Water' report in February 2017 (see http://www.mfe.govt.nz/publications/ fresh-water/clean-water-90-of-rivers-and-lakes-swimmable-2040). This has the aims of making 90% of all rivers and lakes being 'swimmable' by 2040. However, the definition of 'swimmable' was also changed in that report.

Our challenge from Horizons was to consider how to produce a spatially representation and efficient network of water quality sampling stations across the region to meet both national and regional objectives. The region is fairly typical of New Zealand and includes high-country, towns and cities, flood plain, and estuaries, with a wide variety of land use. Inevitably, the network had to be based upon their current monitoring stations.

There are four principal measurements taken at a monitoring station, (i) *E. coli*, (ii) phosphorus, (iii) nitrogen and (iv) turbidity (sediment). We decided to focus on the first of these, as it is the one that forms the basis of the swimmability targets. We also identified a variety of available datasets, of which the most important were the Land Air Water Aotearoa website (https://www.lawa.org.nz/explore-data/) for which the Manawatu-Wanganui data is provided by Horizons from their network of monitoring sites, the Land Information New Zealand data (http://www.linz.govt.nz/data) and the New Zealand Land Resource Inventory (https://catalogue.data.govt.nz/dataset/nzlri-land-use-capability).

Our work split into three parts, which are described in the next sections.

The current state of the rivers

We began by looking at the current state of the rivers in the region by using the available data of *E. coli*, which was typically one sample per month over the last ten years. Based on this data, we asked how many of them passed the new definition of swimmable (a maximum of 540 *E. coli* per 100 mls of water at least 80% of the time). We also considered the other two measures proposed, which are that the median count must be less than 130, and that 90% or measures must be less than 1200. Figure 1 below shows the state of the measurement sites in our region as a plot of mean against standard deviation for the 10 years' worth of data. The lines show the various water quality measures; the brown one is a level of 540 80% of the time, while the blue line is the median. Some particularly good (left) and bad (right) sites are labelled, as are some that are close to the boundary. It seems that 55% of the sites in the Horizons region currently fail the swimmability standards .

In order to test our assumptions, we also checked that the *E. coli* distribution is log-normal (it is), and asked whether or not 10 years' worth of data was sufficient to see a trend at an individual site (it isn't, quite). We did manage to find one positive, though: across the Horizons region, the trend is a decrease in the *E. coli* count of approximately 2.6% a year, which was statistically significant at the 95% level. If this continued, it would mean that 73% of sites would have counts below 540 80% of the time by 2040.

Some members of the group also looked into how fast *E. coli* diffuses, dilutes, and degrades in the rivers – a rough estimate based on linear regression using data from the scientific literature suggested that it takes approximately three days to travel through the river system to the sea, and degrades over approximately the same timeframe.

Correlation with land use

The Land Resource data provided us with an opportunity to see how well land characteristics and land use correlated with *E. coli* pollution. The data contained information on every reach in Manawatu-Wanganui, where a reach is a stretch of river of approximately one kilometre. Along with each reach was information on the climate (how warm/wet), the source of flow (e.g., mountain or wetland), geology (the types of rock), land cover (e.g., forest or pastoral), network position (how big the river is) and valley landform

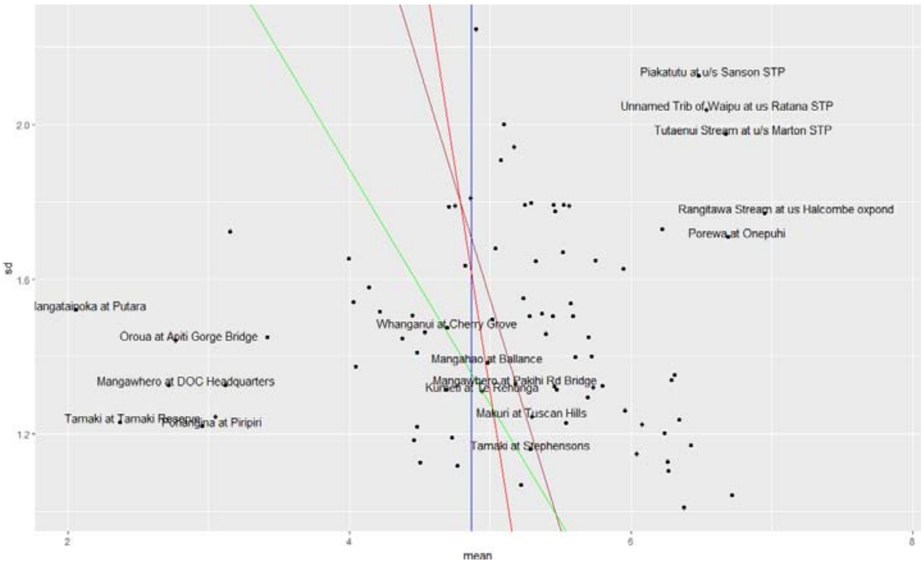


Figure 1: Plot of the mean and standard deviation of log(*E. coli*) for the monitoring sites in the Horizons region. The lines show the various water quality standards. Points to the right are failing those standards.

(slope of the area). We also added in the major monitoring areas (outlined in black). From this data, we were able to model the *E. coli* amount based on the land use and predict an amount for all medium and large reaches, and determined that 60% of the reaches satisfy the swimmibility requirement (note that a river can have many reaches, so this doesn't directly compare to the 55% of failing rivers above).

We also clustered the land data and used it to compare the current network of monitoring stations with random samples that covered all types of land and land use, as can be seen in Figure 2.

Optimisation criteria for monitoring stations

Given that pollution in rivers flows downstream we decided to model pollution flow as a transport problem, with point sources (pollution from factors and sewage outlets) and diffuse sources of pollution (run-off from fields) added in to a convection and diffusion model. We built these into a model of the Manawatu river and catchment area. We then used this model to consider two different optimisation criteria for the locations of sampling sites, starting *tabula rasa*.

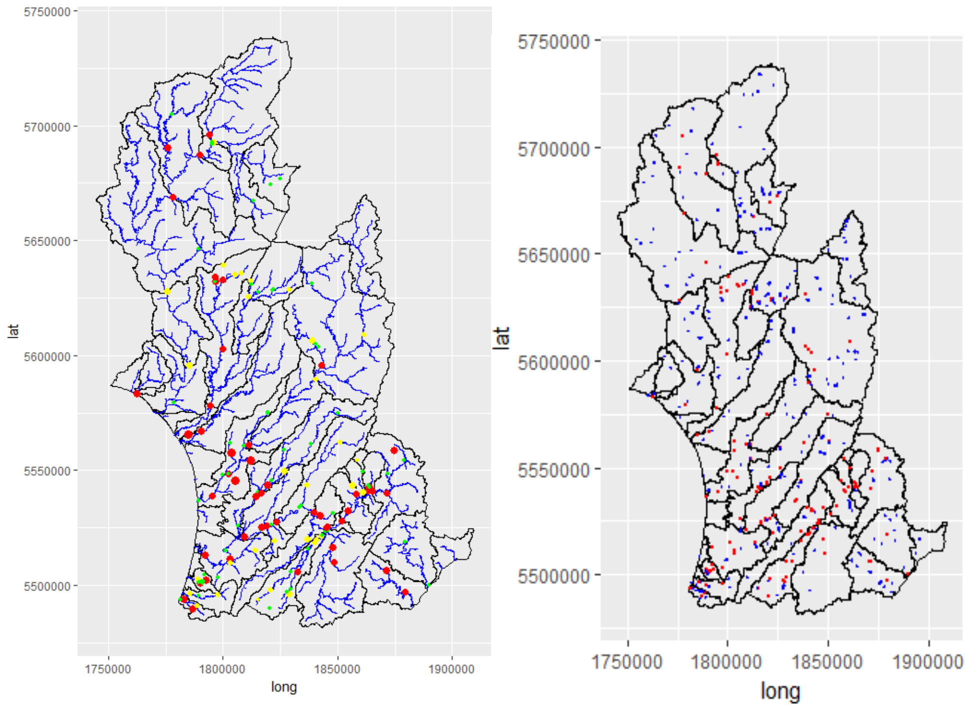


Figure 2: Map of the Horizons region showing left: the current monitoring sites and right: a set of randomly sampled sites that cover the various types of land and land use. On the left plot, red sites are not swimmable, yellow sites are, and green sites do not have information available.

The two criteria we considered were to use information theory and place sites where they provided the most information (i.e., reduced the entropy) the most, and to place sites where they produced the highest expected improvement in a Gaussian Process model. We had hoped to compare these two methods, but unfortunately time was against us. Either of these criteria could also be used to suggest new sites based on the current network of sites.

New monitoring stations cost money to both install and then visit. At the moment, they are often used as a compliance tool, being placed either where there are potential polluters (as a combination of an upstream and downstream site), near areas of important (such as drinking water outtakes), or to detect sources of unexpected pollution.

For our models to be useful we would need to merge it with the land use data model, and also incorporate real-world constraints such as accessibility of the sites. However, we believe that this could be a useful model for Horizons and other regional councils in the future.

Acknowledgements

We thank Horizons (particularly Stacey Binsted, Abby Matthews and Manas Chakraborty) for providing such an interesting question and useful advice during the week.

Challenge from Zespri

Predicting the timing of the kiwifruit harvest

Industry Representatives:

Frank Bollen, Mark Edgecombe, Margot Cotter and Sam Gillings

Academic Moderators:

John Maindonald, StatsResearch Associates

Graeme Wake, Massey University

Alistair Hall, Plant and Food Research

Student Moderator:

Rory Ellis, University of Canterbury

Challenge participants:

Catherine McKenzie (AgResearch), Sunny George Gwanpua (Massey), Mo Li (Massey), Hyuangyo Hong (NIMS South Korea), Yuancheng Wang (Massey), Marnus Stoltz (Otago).

Outline

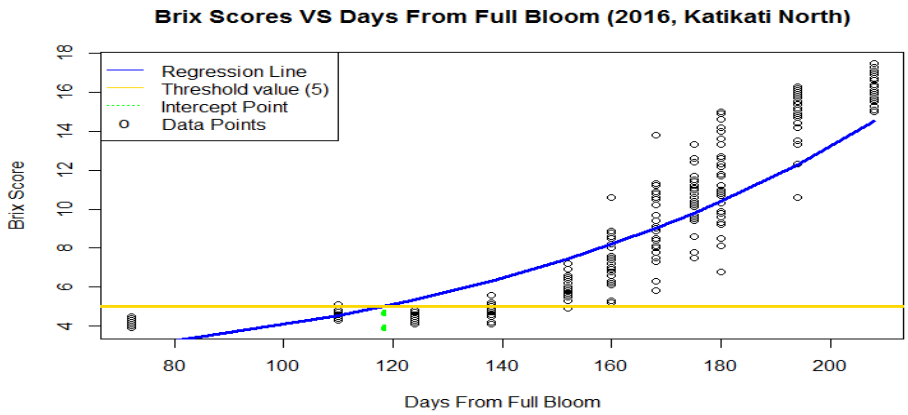
- Description of the Problem
- Exploratory Data/Weather Analysis
- Minimum Threshold Analysis
- Systems Modelling of Fruit Development
- A Priori Determination of Model Parameters
- Concluding Remarks

The Original Problem

- Make sense of available datasets to try and produce valuable modelling and prediction techniques.
- Determine effects of external factors to kiwifruit production**.**

Available Data

- Fruit-specific level data and sample level data (90 observations per grower). An illustrative sample is below.
- Week 9 Monitoring Data, which contained measurements of kiwifruit characteristics at week 9 in the growth cycle.
- The data provided was from 2013 to 2017 inclusive.
- Despite availability of both green and golden kiwifruit data, time constraints and types of data only allowed for consideration of golden kiwifruits.
- Weather data from various weather stations in the area.



(a) **Minimum Threshold Analysis**

- Data Understanding
- Determining data consistency and availability.
- Lead to a reduction in quality data.
- Consideration of quality measures.
- When are these quality measures met?
- Links between quality and time.
- Could not consider all three quality measures.

(b) **Analysis**

- Considered two areas in particular, Katikati and Opotiki.
- For relevance, focussed on 2016 and 2017 data.
- Tested multiple models to fit data, using numerous types of regression to fit both Brix and Dry Matter.
- Fitted these against amount of time passed since fruit started growing.
- Attempted to find the minimum time elapsed to satisfy Brix and Dry Matter Conditions.

(c) **Future Work and Considerations on the Data**

- Determine if these models help optimise Zespri prediction models.

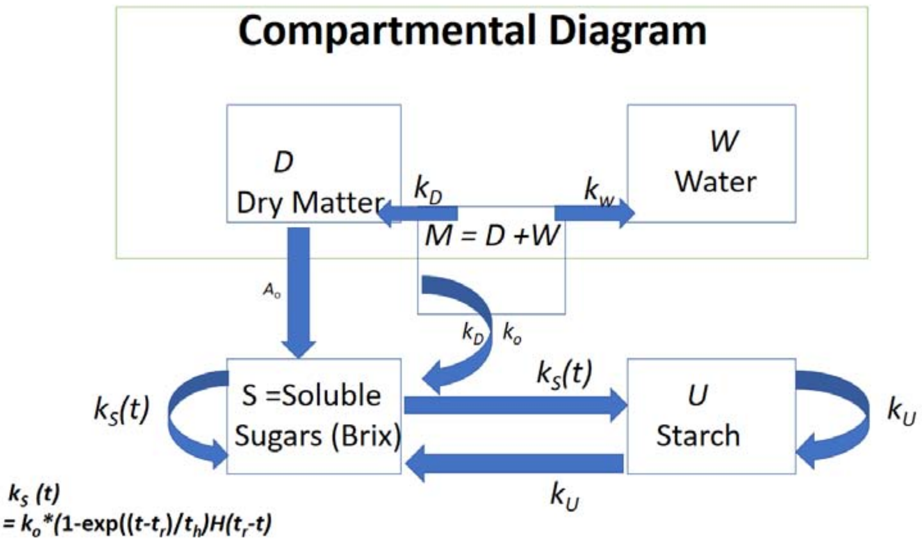
- Examine more consistent data collection methods.
- Apply methodology to other kiwifruit species.
- Consider implementation of more rigorous modelling

4. Systems Modelling of Fruit Development

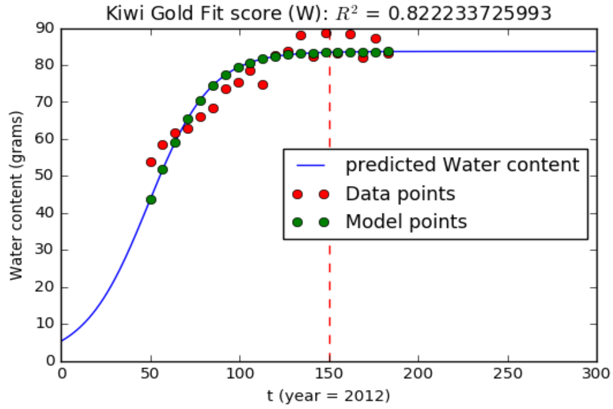
(a) A short summary and aim of the systems model

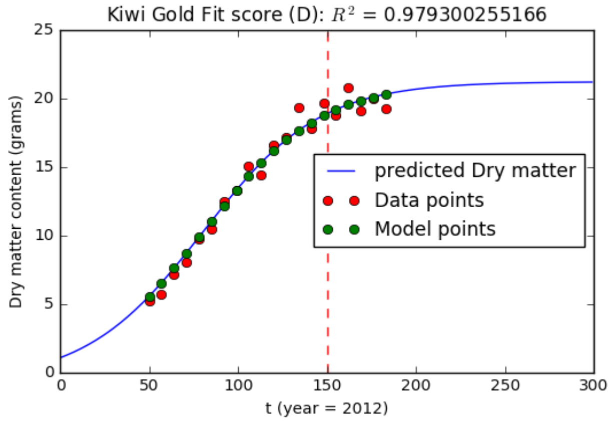
- The proportion of dry matter in kiwifruit is an important indicator of fruit quality, because it is related to the sweetness of the ripe fruit.
- Also, New Zealand growers are paid a premium for fruit that contain a high proportion of dry matter.
- The aim of this model is to provide a decision tool for predicting the harvest time of Kiwi fruit that takes the above mentioned factors into account.

(i) Schematic of the model



(b) Outcome of the model against the data.



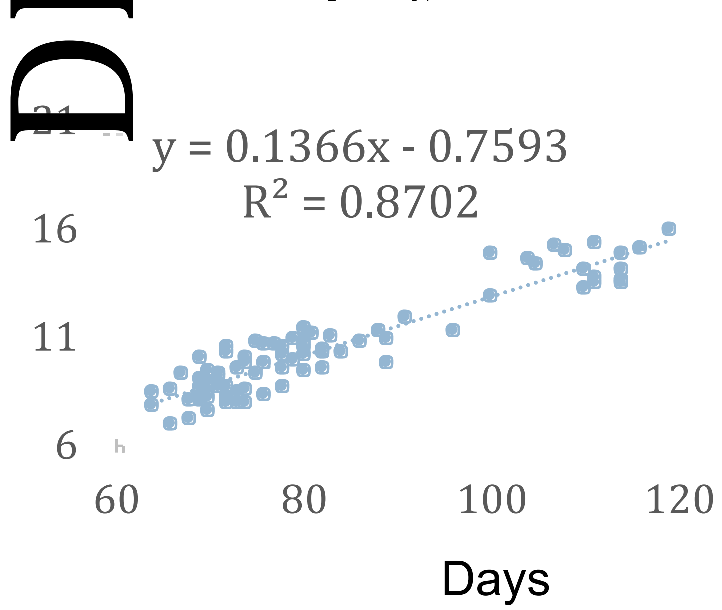


The parameters were determined by fitting the model to some typical data, and so give a good fit (see R^2 values). Optimal harvesting times can then be estimated by using this model in real time, for example the dotted line.

(c) Future work

- Getting some more data and putting it through the model.
- Identifying environmental forcing factors that influence growth and including it in the model
- Making growth rates time dependent to improve prediction.
- Getting actual soluble sugar harvest threshold for Gold Kiwis.

***A priori* determination of batch-specific parameters will allow batch-specific predictions (illustrative example only).**



6. Concluding Remarks

Zespri comments include

- Some of our attendees really enjoyed it and others felt we didn't get much out of it. The latter reflects our frustration at our not being sufficiently organised.
- We have such large datasets that getting a decent problem is quite difficult.
- We need to do more pre-meeting preparation.
- It was a pity the group split into several smaller groups, with the data analysis remaining mostly separate from the three people in the systems modelling team.

The moderating team acknowledges that further work could be undertaken on both the data analysis and the systems modelling fronts by a broad group working more closely together. We are grateful for the opportunity of studying this problem.

Challenge from Sanford Ltd

Comparing and contrasting shear forces and hydrodynamics of Industrial Larval Mussel tank design and operation

Industry Representatives:

Sarah Cumming (SpatNZ), Andrew Stanley (Sanford)

Challenge Moderators:

Rose Davies, School of Aviation, Massey University, New Zealand

Richard Clarke, Department of Engineering Science, University of Auckland, New Zealand

Student Moderator:

Stephen Waite, Auckland Bioengineering Institute, University of Auckland, New Zealand

MINZ Team Moderator:

Liam Bignell, Gonzalo Martinez, Harriet Miller-Brown, David Burton, Clive Davies, Robert McKibbin, John Cater

Aquaculture is one of the world's fastest growing primary industries and demand for aquaculture products is expected to continue growing as the world's population grows and wild-catch levels remain relatively static. Globally, aquaculture will soon produce more seafood than wild fisheries. The New Zealand aquaculture industry, although relatively small on a global scale has positioned itself at the high-end of the market, exporting premium seafood products around the world. Sanford Limited is New Zealand's largest producer and exporter of aqua-cultured products, with Greenshell mussels representing the largest by volume and value. Greenshell™ mussels (*Perna canaliculus*) are unique to New Zealand and are one of New Zealand's most iconic seafood offerings. Mussel aquaculture is one of the world's most efficient forms of food production and is considered a highly sustainable method of producing high protein foods.

The Government's Aquaculture strategy and five-year action plan supports sustainable growth of the aquaculture industry – balancing our economic, social, cultural and ecological values.

Historically, most green-lipped mussels in New Zealand are farmed in the same way. Spat (juvenile mussels) are collected from Ninety Mile Beach and elsewhere in New Zealand, where they wash up in their billions attached to clumps of seaweed. After arriving at a mussel farm, spat are transferred to nursery ropes and grown on the ropes in seawater until about 6 months of age. At this point, they are removed and reseeded onto longlines (stretches of rope up to several kilometres long) that are suspended between buoys.

Mussels are grown for a further 9–12 months before they are harvested. Mussel barges, which harvest the mussels, are mechanised and contain equipment for removing mussels from lines, then declumping, washing, sorting and packing.

Until now, New Zealand's mussel growers have relied on catching wild spat (baby mussels) around our coastline. Supply is unpredictable, yield levels are extremely low and the genetic profile of the mussels are uncontrolled. Through partnership with the New Zealand Government in the form of a primary growth partnership, Sanford have developed a facility capable of selectively breeding Greenshell™ mussels and producing spat on a regular and controlled basis so our growers have the spat they need. SPATNZ (Shellfish Production and Technology New Zealand) operates this hatchery and research facility and its aim is to produce innovations to advance New Zealand's mussel aquaculture industry and deliver benefits for New Zealand's economy.

One of the greatest challenges in the SPATNZ project is to produce batches of larvae year-round. Mussels are naturally seasonal, and the technical challenges involved make it all the more difficult to consistency rear the highly sensitive larvae. They have observed differences in survival of larval rearing in various tank designs. The large commercial tank is shaped like a conical cylinder, while the smaller (non-commercial) tank is more bullet-like in shape. A further significant different between the two tanks is that water\nutrients are introduced at the top of the larger tank, and at depth in the smaller one.

Figure: Sketch of the larger commercial tanks (left) and smaller non-commercial tank (right)

At certain times of year the hatchery can experience total loss of larvae in the commercial scale tank while the smaller scale (non-commercial) tanks continue to perform very well, so it has a big impact on annual production. The reasons for the variation is unknown and one hypothesis is that the shear forces or hydrodynamics generated by aeration differ between these tanks, and interact with microbial communities and mussel larvae in ways that determine the success or failure of the batch.

Sanford were therefore interested in comparing and contrasting the shear forces and hydrodynamic regime in these two tank designs, and hypotheses about why the difference in tank performance is seasonal. We considered a number of different factors which could be leading to the seasonal loss of spats:

i) Environmental Factors

a. Temperature

Sea water is collected at high tide and stored in a tank outside (possibly for several days). It is then drawn into the hatchery, being filtered, UV sterilised and heated before being pumped into the tanks. For the smaller tanks, the water goes to a 15L header tank first, where it is mixed with the algal feed, before being pumped into the tanks. For the larger tanks, the algae is injected just upstream of the inlet into the tank. Seawater salinity is typically 35 ppt (parts per thousand). Measurements were taken at the hatchery in both tanks, at different depths. No significant differences were observed, however, and so this factor was dismissed.

b. Oxygen

Air is drawn from outside, go through a particular and carbon filter, and then blown into the pipes. The flow rate is controlled with a valve. An experiment was conducted to measure the dissolved oxygen content in the two types of tanks, at a nominal operating temperature of 18^{o}C. These measurements indicated that oxygen levels in the larger tank were higher. One possible connection between oxygen levels and seasonality could be due to the fact that water enters the facility with an assumed oxygen saturation of 100% year round, but at different temperatures in summer and winter. After storage in large closed tanks, the seawater is heated (and the algae added) to a consistent temperature before introduction into the larval breeding tanks. This means that in the water will have a higher oxygen content in the Winter.

A possible explanation for the lower measured values in the smaller vessels is that the water in the header tank that feeds the bullet tanks has more time to desaturate, which is born out by oxygen measurement in this header tank (at operating temperature 18^{o}C). However, the opinion of the larval physiologist at the hatchery is that the oxygen levels would need to be significantly higher than seen in either tank in other to be detrimental to the larval health. This factor was therefore not pursued further, however, the response of larvae to dissolved oxygen could be further investigated.

c. Atmospheric Pollutants

Even though the incoming sea water is UV sterilised, there remains the potential for some chemical contaminant to be present at certain times of the year. One hypothesis to explain the seasonality rests on the possibility that there may be seasonal contaminants, from agricultural activities perhaps. The hatchery is located close to neighbouring farmland.

If stock are grazed on pasture or fodder crops during winter, there is a possibility of exposing bare ground, and this will lead to soils becoming pugged and the ground becoming saturated in water, urine and nutrients (Environment Southland 2014). As the nutrient uptake of plants is lowest during winter, nutrient leaching — especially of nitrate (N) — are higher during winter months (Monaghan 2012). These excess nutrients could leach into groundwater or move across the land into waterways. It was therefore suggested that the hatchery periodically perform chemical analyses of the waters entering the tanks, to gauge the significant of any such effect.

d. Bubble Aeration

The water in both types of tanks is aerated through injection of air bubbles at the tank bases.

It was noted that the volume flowrate of air per tank volume is significantly different between the tanks, with values of 0.002 Hz and 0.0004 Hz. This means that the relative volume flowrate of air in the larger tank is approximately 5 times smaller than the smaller tank. It was previously demonstrated that a primary effect of the bubble flow is to drive mixing in the vessels, and the smaller tank has a larger aspect ratio (depth to radius) with air released near the base, so it might be expected that more bubble-driven mixing is occurring in the smaller tank.

To test whether the dissolved oxygen levels in the tank are affected by the bubble flowrate, an increased air flow (rate unknown, but expected to be 4-5 times greater) was added to a larger tank for a period of 1 hour at the Hatchery in Nelson, with the dissolved oxygen content measured before and after as shown below:

The effect of the increased flowrate was to decrease the dissolved oxygen in the tank, and therefore the saturation of the dissolved gas. There are two possible mechanisms proposed to explain this reduction. The first is that the increased size and number of bubbles provides more surface area for gas exchange. However, published literature suggests that the bubbles would need to be much smaller than the ones generated in the tanks to be effective oxygen sinks (https://en.wikipedia.org/wiki/Water_aeration).

The second possibility is that a high flowrate generates a larger surface disturbance at the top of the tank, which in turn creates more surface area for gas transfer. This is consistent with CFD predictions.

e. Acoustic Noise Level

Previous measurements in the 160L tanks have shown that the ambient noise levels are approximately 131dB (with 1 umPa reference pressure), this is consistent with previous results for water-filled hatching facilities (Lillis et al., 2013). There is also evidence of other captive marine larvae responding to the acoustic perturbations (de Soto et al, 2013), and therefore the sound field should be considered. The density and salinity of seawater at 20 degrees Celsius was used to estimate the acoustic impedance inside the tanks which was calculated to be 1.6MRayl, which is a typical value for water. The maximum particle velocity at 131dB is calculated to be approximately 3 um/s (pressure amplitude 5Pa), which is much smaller than the characteristic velocity scale of the larvae.

The geometry for the smaller tank is considerably different, so the resonant frequencies were estimated for both tanks. As both have a relatively small frustum, it was assumed that the dominant acoustic behaviour corresponds to the resonant response of a closed cylinder, using the overall depth as quarter the maximum wavelength. This results in first resonant frequencies of 609 Hz and 326 Hz for the large and small tanks respectively. End effects were neglected as the medium changes (water to air) at the top surface. Unfortunately, sounds levels in the small tank have not been recorded, so it is not possible to make a firm conclusion, but this is worth further experimental investigation. There is no evidence of seasonality in the acoustic environment.

ii) Tank Hydrodynamics

Computational Fluid Dynamic (CFD) simulations, performed using OpenFoam on the Auckland High Performance Computing cluster, were used to determine whether the hydrodynamic regime was significantly different in the two different tanks. In particular, the hydrodynamic shear stresses. It was observed that for the original tank designs, the smaller tank experiences greater hydrodynamic wall stresses. It was hypothesises that this might counter formation of biofilms on these surfaces, and hence greater overall improved water health. Using CFD, we explored whether moving the nutrient supply to the bottom of the larger tank, as is done in the smaller tank, would similarly increase the hydrodynamic shear stresses.

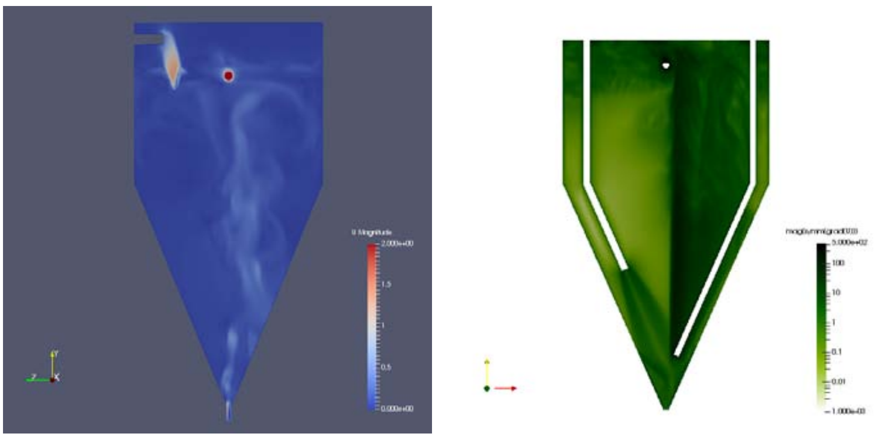


Figure: (Left) Flow velocity of original large tank. (Right) New design of large tank, with water and air supplied at the bottom

These simulations did show that the different tank design does substantially alter the levels of hydrodynamic shear stresses within the tank. A submerged water inlet with air inlet at the bottom leads to a plume of air shear along the centreline of the tank associated with the rising bubbles and a much greater region of elevated shear in the bulk fluid within the tank.

Challenge from Transpower

Transmission line conductors – big data cleansing, probability of failure derivation and asset health relationship

Industry Representative:
Irvin Chew, Transpower, Wellington

Challenge Moderators:
Mark McGuinness, Victoria University, Wellington
Napoleon Reyes, Massey University, Auckland

Student Moderator:
Sarah Pirikahu, Massey University, Palmerston North

MINZ team:
Joon Heo, Roelant Hofmans, Gerrard Liddell, Matthew Sawyer, Alan Williams, James Williams, Matt Wilkins, Xun Xiao

Electricity is distributed around New Zealand by thousands of kilometres of transmission line conductors. As grid owners, Transpower are vigilant in performing regular inspections and maintenance on these conductors, and the joints which connect them, to ensure the safe flow of electricity around the country. Each instance where work has been undertaken is recorded by Transpower. This historical work order data is rich in information, but is made up of many columns (data fields) which contain an abundance of unstructured free text. The first challenge asked of the MINZ group was to take this large dataset, which consisted of over 120,000 work orders, and provide a method of converting the free text summaries into a structured format. Once a cleaned dataset was available, Transpower were interested in determining the probability of defect/failure and investigating the relationship between types of defects and conductor failures.

The data provided consisted of four fields, work order description, long description, log summary and work log long description, which needed to be analysed in order to assess whether an inspection, fault or replacement had occurred. Lexical analysis showed a great variety of expression, and also the presence of many words that did not have immediate dictionary meanings. Through this exploration process, approximately 60,000 records could be identified as being related to vegetation. Clustering of the work orders based on the words that occur in these four text fields was performed using the technique called Latent Dirichlet Allocation (LDA) [1]. LDA is based on a Bayesian model which allows us to determine the probability of a work order arising in a certain topic (cluster), based on the words associated with it. By creating an interactive widget in Python, we were able to visualize the clustering of data (see Fig.1). We observed that there was a clear divide in topics which could then be identified as work orders related to vegetation, those which are projects that have been undertaken by Transpower, and finally those related to replacements or maintenance of joints.

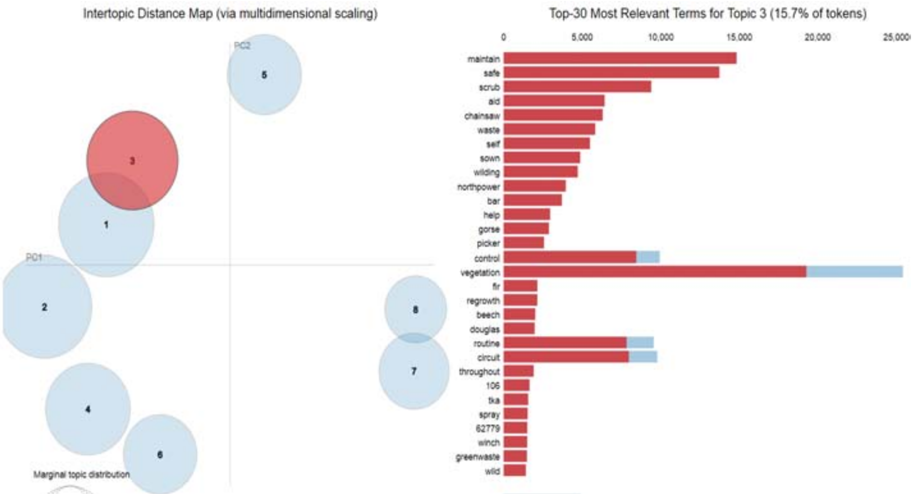


Fig. 1: Using LDA for discovering topics hidden in a collection of work orders
Historically there have been very few total conductor failures, due to the regular maintenance of the lines. This makes linking the type of defect to conductor failure quite difficult. Using the relatively clean dataset from the first part of this problem, we were able to derive the probability of a defect occurring. Two alternative approaches were suggested in order to derive the probability of defect, namely a statistical approach and a machine learning approach.

Exploring the probability of defect

Statistical modelling of the time to defect via a Cox proportional hazard model allowed us to incorporate multiple covariates of interest, for example the location of the conductor, the material it was made out of, its length and voltage. This model provided us with the hazard rate for any combination of covariates, which could then be used to determine the probability of defect. In the experiments, it was observed that Conductor Group ACSR AC has a large hazard rate (5.002). This suggests that it will have a low probability of survival (that is a high probability of defect), whereas Conductor GroupCopper has a low hazard rate (0.090), suggesting that there a higher probability of survival. On the other hand, with respect to corrosion code, it was least expected that an area of high corrosion would result in a low hazard rate (0.967). After examining the data set further to shed some light on this result, we found some inconsistencies. Some records have install dates > defect dates, and some with missing attributes. Transpower also confirmed that the more durable conductors were installed in areas of high corrosion.

Consequently, the group designed a neural network-based defect prediction system but limiting its applicability only on a span; which is a section of a circuit, going from one town to another. This necessitated the construction of a new data set that possesses all the relevant attributes of spans for prediction purposes. Programs were written to perform cross-referencing between the data sets, and also to merge all details corresponding to individual spans. Existing models from the Tensorflow computational software library was utilised in our preliminary testing and design, but the team realised that the prediction problem calls for a customised auto-regressive recurrent neural network with long short-term memory (LSTM) cells [2]. We envisage that different configurations of neural networks should be trained and tested (i.e. different sizes of input vectors, output vectors, number of hidden nodes, number of hidden layers, learning rates, topology, etc.) in order to find the best performing one. Furthermore, we think that the prediction system could significantly benefit from a time series data that is measured in smaller, regular time steps.

In summary, Transpower's challenge has certainly brought out the participants' best in order to tackle it, as it called for algorithms for text mining, data cleansing and time-series prediction. The Latent Dirichlet Allocation model was found to be instrumental in text mining the data set of mostly free text summaries, while the Cox proportional hazard model shed some light on the conductors' probability of defect. Lastly a customised auto-regressive RNN with LSTM cells was designed for predicting the emergence of future defects within a span. Unfortunately, due to time constraints, the group was not able to complete the implementation and testing of the proposed neural network prediction system.

References

1. Blei, D.M., Ng, A.Y., Jordan, M.I., Lafferty, J.: Latent dirichlet allocation. Journal of Machine Learning Research 3 (2003) 2003

Gers, F.A., Schmidhuber, J., Cummins, F.A.: Learning to forget: Continual prediction with LSTM. Neural Computation 12(10) (2000) 2451-2471

Sponsors

ANZIAM

NZMS

KiwiNet
Kiwi Innovation Network

Centre for
Maths in Industry

MASSEY
UNIVERSITY
TE KUNENGA KI PŪREHUROA

UNIVERSITY OF NEW ZEALAND