Fourth Annual Red River Valley Statistical Conference

North Dakota State University
Department of Statistics

Wednesday, April 30, 2014

Fourth Annual Red River Valley Statistical Conference

Session 1: Chair: Gang Shen Location: Dunbar 152

12:00 pm *Prediction of 2014 World Cup soccer winner: Two statistical methods*, Mohamed Sylla
12:15 pm *Predicting Outcomes of NBA Basketball Games*, Scot Jones
12:30 pm *A predictive Model of Division I Women's College Basketball*, Wenting Wang
12:45 pm *March Sanity: Predicting the March Madness Tournament Using Least Squares Regression*, Bryan Alan Rask

1:00 pm
Refreshment Break Location: Morrill 217
Poster Session

Session 2: Keynote Speakers Chairs: Tatjana Miljkovic & Seung Won Hyun Location: Dunbar 152

2:00 pm *Identifying genes that are differentially expressed in both of two independent experiments*, Megan Orr
2:30pm *Test Equality of Curves with Homoscedastic or Heteroscedastic Errors*, Yarong Yang

Session 3A: Chair: Megan Orr Location: Morrill 105

3:05 pm *Multiple comparisons in experimental designs with large numbers of treatments: an assessment of the performance of the false discovery rate and Dunnett's test*, Kayéromi Gomez
3:20 pm *Identification of differentially expressed genes and gene sets taking into consideration the distribution of effect sizes in microarray analysis*, Ekua Bentil

Session 3B: Chair: Seung Won Hyun Location: Dunbar 152

3:05 pm *T-Optimal Designs for Model Discrimination in Probit Models*, Yue Ming
3:20 pm *Robust c-optimal design for estimating the EDp*, Anqing Zhang

3:35 pm
Refreshment Break Location: Morrill 217

Session 4A: Chair: Megan Orr Location: Morrill 105

4:00 pm *An Analysis of Salary for Major League Baseball Players*, Michael Hoffman
4:15 pm *Cluster Analysis Comparison for Ellipsoidal Data*, Shane Loeffler
4:30 pm *Boundary Estimation*, Yingfei Mu
4:45 pm *Difference in Proportions for Men and Women Running Marathons*, Jennifer Johnson

Session 4B: Chair: Yarong Yang Location: Dunbar 152

4:00 pm *Proposed Nonparametric Tests for the Simple Tree Alternative in a Mixed Design*, Susan Olet
4:15 pm *Two Approaches to the Isotonic Change-Point Problem: Nonparametric and Minimax*, Karl D'Silva
4:30 pm *Comparison of Classification Probabilities among Logistic Regression, Neural Network and Support Vector Machines in the Presence Of Missing Data*, Sudhi Upadhyaya
4:45 pm *Nonparametric Tests for the Non – Decreasing and Umbrella Alternatives in the Incomplete Block and Completely Randomized Mixed Design*, Alfred Ndungu

ABSTRACTS FOR SPEAKERS

(in alphabetical order by first author's last name)

Author: Ekua Bentil

Title: Identification of differentially expressed genes and gene sets taking into consideration the distribution of effect sizes in microarray analysis

Microarray technology allows for gene expression levels to be compared across treatments for thousands of genes simultaneously. Many statistical methods exist for identifying differentially expressed genes and gene sets while controlling for multiple hypothesis testing error. Most methods do not take into account the distribution of effect sizes. However, statistical methods that take into account the asymmetry in the effect sizes (when this asymmetry is present in a data set) may improve identification of differentially expressed genes. In this study, an improved method for computing q-values when the distribution of the effect sizes is asymmetric is presented. This technique is extended to microarray experiments with more than two treatment groups when the treatments can be ranked. The results suggest that the improved method for computing q-values improves upon the traditional method for computing q-values in the identification of differentially expressed genes and gene sets while controlling the false discovery rate.

Author: Karl D'Silva

Title: Two Approaches to the Isotonic Change-Point Problem: Nonparametric and Minimax

A change in model parameters over time often characterizes major events. Situations in which this may arise include observing increasing temperatures, intense rainfall, and the valuation of a stock. The question is whether these observations are simply the result of natural variation, or rather are indicative of an underlying monotonic trend. This is known as the isotonic change-point problem. Two approaches to this problem are considered: Firstly, for correlated data with short-range dependence, we prove that a particular U-statistic based on a modified version of the Jonckheere-Terpstra test statistic is asymptotically equivalent to a more complex U-statistic discussed by Shen and Xu (2013); one that has been shown to outperform other existing tests in a variety of situations. Secondly, we shall justify and utilize the minimax criterion in order to identify the optimal test statistic within a specified class. We shall see that, as motivated by the projection method, the aforementioned class is the class of contrasts. It shall be proven that the set of coefficients originally proposed by Abelson and Tukey (1963), and utilized by Brillinger (1989) in the isotonic change-point setting, are in fact minimax in the independent data case. For correlated data with short-range dependence, we shall demonstrate a sufficient condition for minimaxity to hold.

Author: Kayéromi Gomez

Title: Multiple comparisons in experimental designs with large numbers of treatments: an assessment of the performance of the false discovery rate and Dunnett's test

Many statistical analyses require the comparison of a large number of treatments after initially rejecting the null hypothesis that treatments are the same following an ANOVA. In other cases, the researcher wants to make a specific recommendation as to which set of treatments are the best. Multiple comparisons procedures have evolved around the control of family-wise and comparison-wise error rates and many post-hoc testing methods have been developed to address the issue. With the recent surge in data sets to compare hundreds of treatments simultaneously in disciplines such as genomics, there is a need for statisticians to develop new multiple comparisons testing procedures. The publication in 1995 of the False Discovery Rate (FDR) method opened the door to a world of new techniques that have since enabled researchers to find solutions to a multitude of treatment comparison problems (Benjamini & Hochberg, 2000).

The current study aims at comparing the somewhat newly-published method FDR with a preferred method of many statisticians, Dunnett's test. A simulation is performed using the characteristics of an agronomic data set with 300 different genotypes of dry beans and the FDR and Dunnett's test are compared with respect to their control of type I errors and their powers.

Author: Michael Hoffman

Title: An Analysis of Salary for Major League Baseball Players

This thesis examines the salary of Major League Baseball (MLB) players and whether players are paid based on their on-the-field performance. Each salary was examined on both the yearly production and the overall career production of the player. Several different production statistics were collected for the 2010-2012 MLB seasons. A random sample of players was selected from each season and separate models were created for position players and pitchers. Significant production statistics that were helpful in predicting salary were selected for each different model. These models were deemed to be good models having a predictive r-squared value of at least 0.70 for each of the different models. After the regression models were found, the models were tested for accuracy by predicting the salaries of a random sample of players from the 2013 MLB season.

Author: Jennifer Johnson

Title: Difference in Proportions for Men and Women Running Marathons

Marathons have become a popular activity for both men and women across the Midwest. I want to investigate whether the popular method of age-grading results is gender based. The USATF uses age-grading as a method to compare results across age and gender lines. I compared the proportion of men to the proportion of women who finished in the top 33% after age-grading in each of four Midwest marathons that took place in 2013. These marathons are the Fargo Marathon (Fargo, ND), Grandma's Marathon (Duluth, MN), Columbus Marathon (Columbus, OH), and Medtronic Twin Cities Marathon (Minneapolis, MN). The null hypothesis for these tests is that the proportion of men is equal to the proportion of women with an alternative hypothesis of the proportion of men is greater than the proportion of women in the marathon.

Author: Scot Jones

Title: Predicting Outcomes of NBA Basketball Games

A stratified random sample of 180 NBA basketball games was taken over a three-year period, between 2008 and 2011. Models were developed to predict point spread and to estimate the probability of a specific team winning based on various in-game statistics. The year the games were played did not matter. The models were verified using exact in-game statistics for a random sample taken during the 2011-2012 season, and were found to have an accuracy of 91%. Three methods were used in an effort to estimate in-game statistics of a future game so that the models could be used to predict a winner in a game played by Team A and Team B. Models using these methods had accuracies of approximately 64%.

Another stratified random sample of 144 NBA basketball games was taken in the last two weeks of March 2013 for the playoff contending teams. Based on this sample, a new model was developed to estimate the point spread of a basketball game. The following in-game statistics were significant in the model: FGS; 3PS; FTS; ORS; ASTS; TOS; FTAS. Seasonal averages for these in-game statistics will be found for each team in the playoffs and will be used in the model developed to predict the winner of each game for the 2014 NBA Championship.

Author: Shane Loeffler

Title: Cluster Analysis Comparison for Ellipsoidal Data

This is a comparison of clustering algorithms to find the clustering algorithm that performs the best. The clustering algorithms used: Partitioning Around Medoids (PAM), K-means, Mclust, Hierclust with Ward's linkage, Single linkage, Complete linkage, Average linkage, McQuitty's method, Gower's method, and Centroid method.

A mixture of Gaussians simulated with pre-specified number of dimensions (2, 5, and 10), number of clusters (4, 8, and 16), and average overlap (0.001, 0.005, 0.01, 0.05, 0.1, and 0.2) between the clusters. Data will then be simulated from the mixtures with the average number of points in each cluster (10, 25,100).

From each combination of dimensions, clusters, average overlap, and points in each cluster, the adjusted Rand index will be calculated from each clustering algorithm. This will be repeated 1000 times; the clustering algorithm with the highest average adjusted rand will be considered the best for this particular case. A t test will be used to do pairwise comparisons between the clustering algorithms to show if there is any significant difference from the best clustering algorithm for that particular case.

Author: Yue Ming

Title: T-Optimal Designs for Model Discrimination in Probit Models

When dose-response functions have a downturn, one interesting feature to study is the significance of the downturn. The interesting feature can be studied using model discrimination between two rival models (model describing dose-response functions with a downturn versus model describing only increasing part of the response functions). In this article, we study *T*-optimal designs that can best discriminate between these two rival models. Three different sets of model parameter values are considered to demonstrate various shapes of dose-response functions. Under the different sets of the parameter values, the *T*-optimal designs are obtained, and their performances are compared to two other known designs for the model discrimination (*Ds*-optimal design and Uniform design) through Monte Carlo Simulation.

Author: Yingfei Mu

Title: Boundary Estimation

A dataset consists of observations taken at the nodes of a grid. An unknown boundary partitions the grid into two regions. All the observations coming from a particular region share a common distribution, but the distributions, if they are different, differ only in their means. We will first construct a test for the existence of a change-curve; then, we will develop a test-based method to generate a change-curve that delineates the regions. We will also study the asymptotic properties of the new algorithm. We expect the boundary estimate to have a desirable convergence rate to the true boundary and some asymptotic minimaxity to support its superiority over all the other competing boundary detection methods as well.

Author: Alfred Ndungu

Title: Nonparametric Tests for the Non-Decreasing and Umbrella Alternatives in the Incomplete Block and Completely Randomized Mixed Design

This research study proposes a solution to deal with missing observations which is a common problem in real world datasets. A nonparametric approach is used because of its ease of use relative to the parametric approach that beleaguer the user with firm assumptions. The study assumes data is in an Incomplete Block (IBD) and Completely Randomized (CRD) Mixed Design. The scope of this research was limited to three, four and five treatments. Mersenne - Twister (2014) simulations were used to vary the design and to estimate the test statistic powers.

Two test statistics are proposed if the user expects a non – decreasing order of differences in treatment means. They are both applicable in the cited mixed design. The tests combine Alvo and Cabilio (1995) and Jonckheere – Terpstra ((Jonckheere (1954), Terpstra (1952)) in two ways: standardizing the sum of the standardized statistics and standardizing the sum of the unstandardized statistics. Results showed that the former is better.

Three tests are proposed for the umbrella alternative. The first, Mungai's test, is only applicable in an IBD. The other two tests combine Mungai's and Mack – Wolfe (1981) using the same methods described in the previous paragraph. The same conclusion holds except when the size of the IBD's sample was equal to or greater than a quarter that of the CRD.

Authors: Susan Olet & Rhonda Magel

Title: Proposed Nonparameteric Tests for the Simple Tree Alternative in a Mixed Design

Six different test statistics consisting of modified versions of the Page's test and the Fligner-Wolfe test are proposed for testing the simple tree alternative in a mixed design consisting of an RCBD portion and a CRD portion. The test statistics all have asymptotic normal distributions under the null hypothesis of no differences in treatments. A simulation study is conducted to compare the six test statistics under a variety of conditions including changing the number of treatments, varying the underlying distributions, increasing the variance difference between the RCBD and CRD portions, changing the proportions of the RCBD portion to the CRD portion, and changing the treatment effects. The tests are compared on the basis of estimated powers. Recommendations as to which of the test statistics are better to use are given and depend mainly on the proportion of the RCBD portion to the CRD portion.

Authors: Megan Orr, Peng Liu, and Dan Nettleton

Title: Identifying genes that are differentially expressed in both of two independent experiments

Identifying genes that are differentially expressed (DE) in two independent experiments generally involves two steps. In the first step, gene expressions from each experiment are analyzed separately to produce a list of genes that are declared DE in each experiment while controlling false discovery rate at some desired level alpha. Then, genes common to both lists are declared to be DE in both experiments. We call this approach the “intersection method". Little, if any, research has been done to evaluate how well this method controls the false discovery rate (FDR) or ranks genes based on significance. In addition to exploring these questions, we also propose a new method for estimating FDR. These two methods, as well as another method developed with a different goal in mind, are compared through two simulation studies, one involving independent normal data and one involving real gene expression data. These simulation studies demonstrate the advantages of the proposed method. We conclude the paper by providing an analysis of data from two experiments involving gene expressions in maize leaves.

Author: Bryan Alan Rask

Title: March Sanity: Predicting the March Madness Tournament Using Least Squares Regression

Every year in March, 68 of the best NCAA men's basketball teams compete in a single-elimination tournament to determine a champion in what has been deemed March Madness. Along with the tournament, millions of fans fill out their brackets to determine the winner. This year Warren Buffet, along with Yahoo Sports, offered one billion dollars for a perfect bracket. Although that is near impossible, they also offered one hundred thousand dollars to each of the top 20 brackets out of the millions that were entered. In order to better predict the tournament results, I used least squares regression with the independent variables being per game averages of certain team statistics throughout the regular season. The dependent variable was the point spread between the two teams. I separated the tournament into three rounds: first, second, and third-championship and developed a model for each accordingly. All three of the models determined for the tournament resulted in an R^2 of below 40% and overall, it predicted 59% of the games correctly. Although these models were not as accurate as hoped, there could possibly be more research in the future to build better and more efficient models in helping to predict tournament outcomes.

Author: Mohamed Sylla

Title: Prediction of 2014 World Cup soccer winner: Two statistical methods

Soccer game is considered the most popular sport on earth and applying statistical models to analyze Soccer data has been of a keen interest to modern researchers. Statistical modeling of soccer data also provides guidance and assistance to stakeholders across board. The goal of this paper is to establish a consistent statistical approach to help in the prediction of the upcoming Brazil 2014 World Cup winner. Since the Point spread per game is not only used to predict the outcome of a game but also to forecast the score of a game, a Logistic and Least Squares regression are performed to create a win probability model and to scrutinize the Point Spread estimate per game approach. Discriminant Analysis was also used to determine which variables significantly influence individual game wins. Fisher classification procedure allows for interpretability while providing a robust approach to classifying the 32 contestants of the 2014 World Cup using the previous data from 2006 and 2010. Analyzing the past data in depth from the 2006 and 2010 qualifiers allowed for a sufficient and accurate prediction of games outcome stage by stage until the final two contestants are identified.

Author: Sudhi Upadhyaya, Curt Doetkott, Dr. Rhonda Magel, Dr. Tatjana Miljkovic & Dr. Kambiz Farahmand

Title: Comparison of Classification Probabilities among Logistic Regression, Neural Network and Support Vector Machines in the Presence Of Missing Data

Missing data have challenged researchers since the beginnings of research and recently analyses of incomplete data has been the object of many studies. Several statistical models such as Logistic Regression, Neural Network and Support Vector Machines use these datasets with missing values as an input while making inferences regarding the population. When the quality of the results obtained using these models mainly depends on the quality of the data set used, the presence of missing data can severely skew the results and reduce the efficiency of the model. The objective of our study was to identify a robust model among LR,NN, SVM in the presence of missing data. The study was conducted by simulating 10000 observations using the characteristics of a test dataset based on Monte Carlo simulation and missing data was introduced randomly at 10% level. Single mode imputation technique was used to impute missing values. A simple random sample of 120 , 240 and 500 observations were chosen and these three models were executed for each of these scenario. Results showed that the performance of SVM (89.1%, 84.43%, 79.48%) was far superior compared to LR (61.5%, 62.14%, 62.9%) or NN (59.39%, 60.65%, 60. 34%) models. However, the classification efficiency of SVM was gradually decreasing as sample size increased.

Author: Wenting Wang

Title: A predictive Model of Division I Women's College Basketball

Women’s basketball game is becoming more and more popular, spreading from the east coast of the United States to the west coast, in large part among women’s colleges. The National Collegiate Athletic Association (NCAA) Women’s Division I Basketball Tournament is one of the most famous ones. It is also known as March Madness or the Big Dance. The objectives of this study is: Develop least squares regression models to predict for Round 1, Round 2 and Rounds 3- 6, to predict winners of basketball games in each of those rounds for the NCAA women's basketball tournament;

Statistic data was collected for two seasons of the NCAA Women’s basketball tournament. This included the 2011 and 2012 tournaments. The least squares regression models were used with the data from 2013 and 2014 tournaments to verify the accuracy of the models.

Author: Yarong Yang

Title: Test Equality of Curves with Homoscedastic or Heteroscedastic Errors

Testing equality of two curves occurs often in functional data analysis. We develop procedures for testing if two curves measured with either homoscedastic or heteroscedastic errors are equal. The method is applicable to a quite general class of curves that can be either specified up to some unknown parameters, or are only assumed to be smooth. It does not have to have repeated measurements to obtain the variances at each of the explanatory values, as other related procedures do. Instead, it calculates the overall variances by pooling all the data together. The null distribution of the test statistic is derived and an approximation formula to calculate the P-value is developed when the heteroscedastic variances are either known or unknown. Simulation experiments are conducted to show that this procedure works well in the finite sample situation. Comparisons with other test procedures are made based on simulated data sets. Applications to our motivating data example from an environmental study will be illustrated. An R package is compiled for ease of general applications.

Author: Anqing Zhang

Title: Robust c-optimal design for estimating the EDp

Optimal design provides the most efficient design to study dose response functions. It is often observed to adopt the four-parameter logistic model to describe the dose-response relationship in many dose finding trials. Under the four parameter logistic model, optimal design to estimate the ED_p accurately is presented. The ED_p represents the dose achieve 100p% of the maximum treatment effect. C-optimal design works the best to estimate the ED_p , but the value of p must be predetermined in order to obtain the c-optimal design. Here we investigate the efficiency of c-optimal design to estimate the ED_p for different values of p and present robust c-optimal design that works well for the changes in the value of p. Five different values of p are considered in this study: ED_{10}, ED_{30}, ED_{50}, ED_{70}, ED_{90}. The performance of the robust c-optimal design is obtained and compared to the c-optimal designs and traditional uniform designs.

ABSTRACTS FOR POSTERS

(in alphabetical order by first author's last name)

Authors: Taryn Chase, Peter Martin, & Tatjana Miljkovic

Title: Loss Triangles

The development of loss triangles is done by insurance companies in order to predict the losses in future years. Most insurance companies use the deterministic method to develop these losses. Using statistical software such as R or WinBugs, these losses can be calculated probabilistically. This project will develop a loss triangle for claim data from a real insurance using the chain-ladder method and explore some probabilistic methods. The results from the methods will be compared.

Authors: Joshua Hugen, Michael Hoffman, & Tatjana Miljkovic

Title: k-Component Mixture Modeling

Insurance premiums are determined by both the number of losses and the severity of these losses. These are often modeled using certain well-known single or composite distribution. Previous attempts have been made to model the famous Danish Fire Loss dataset. Cooray and Ananda (2005) tried to model it with a composite Lognormal-Pareto model. The dataset, and loss distributions in general, tend to have heavy tails, so the idea of the lognormal-Pareto composite was to use the heavy tail of the Pareto and the mound shape of the lognormal. This yielded a pretty good fit.

Datasets are often not homogeneous but heterogeneous and a better model might be a composition of several distributions. In that case, in order to appropriately model the dataset, we need a model that is likewise a composite of several distributions. Lee and Lin (2010) did this with k Erlang distributions, but it has never been tried for the aforementioned Danish dataset.

For modeling severity of Danish Fire Losses dataset, we propose a k-component finite mixture model. We will find a number of components, k, such that an optimal fit for the data is achieved. We will make use of the expectation maximization (EM) algorithm for mixing k distributions together.

Author: John Lauman-Beltz

Title: A multiple regression analysis of McDonald’s sandwich serve times, sandwich counts, drive-thru sales, cars served

Service sector businesses, like McDonalds, love to see profits increase. McDonalds does this by selling a variety of items, such as sandwiches, while trying to minimize expenses. The objective of this year’s project is to see if average serve times, number of sandwiches across different one hour periods from 9am – 5pm, and drive thru sales affects numbers of cars served at one local McDonald’s franchisee here in Fargo ND. The data I will be using is from an actual franchisee store. Analysis of variance (ANOVA) and multiple regression will be utilized to show this.

Authors: Christopher McEwen & Yarong Yang

Title: Utilizing the Local Outlier Factor Algorithm as a Refinement Tool for Classification Models

Utilization of machine learning methods to predict medical outcomes has increased in frequency. However, these methods may misclassify data as false positives or false negatives. A method using outlier detection through the local outlier factor (LOF) algorithm was devised to predict whether certain data points may be more likely to be misclassified. A simulation study consistently found that the LOF algorithm could consistently predict many data points misclassified as false positives.

Author: Purbasha Mistry

Title: The Effect of Temperature and Precipitation on the yield of Spring Wheat in North Dakota

The agricultural sector plays a significant role North Dakota's economy and is said to be the leading industry with around $4.5 billion annual turnover. Wheat being the number one crop contributes the most in the increasing economy. A variation in spring wheat yield is influenced by climatic and non-climatic conditions. In this study effect of temperature and precipitation has been studied on the yield of spring wheat in North Dakota region. Data collected for over 43 years and demonstrated in tabular and then graphical form. Although a rise in temperature may be detrimental for crop yield in various regions, global warming has been advantageous for North Dakota as the growing season has been increased by 17.5 days in the past century (Akyuz, 2013). There is an increasing trend (more than 1.35 tons/ha) as far as the wheat yield is concerned. It has been observed that a change in precipitation has a prominent effect on the change in wheat yield from the year 1970 to 2013. A linear regression analysis is attempted to investigate the relationship between the two major climate variables on the spring wheat yield.

Author: Partheeban Selvarajah

Title: Statistical Analysis of Settlement and Rotation of Rocking Shallow Foundations from Centrifuge and Shake Table Experiments

Earthquakes cause damage to structures, such as buildings and bridges, as the seismic energy propagates from soil to structures. Shallow foundations are conventionally designed with fixed-base condition, which allows elastic behavior and prohibits ductile and plastic behavior. Recent research findings have shown that unconventionally designed shallow foundations, with controlled rocking, could significantly reduce the seismic demands on a structure. However, the hurdles that hinder the use of rocking foundation in practice include the prediction of resulting deformations (settlement and rotation) and the uncertainties associated with them. Settlement and rotation data of rocking shallow foundation have been obtained from centrifuge and shake table experiments conducted around the world. The objective of this study is to correlate the permanent settlement and rotation of the foundation with critical contact area ratio of the foundation, which includes the effects of soil parameters, footing geometry, and structural loads, and maximum ground acceleration or root mean square of ground acceleration. This poster presents the statistical analysis of settlement and rotation data of rocking shallow foundations, including data collection method, numerical summary of data, graphical display of data, and inferential method to analyze the data.

Author: Achintyamugdha S. Sharma

Title: Dispersion of Carbon Nanotubes in Aqueous Solutions

Carbon nanotubes (CNTs) are seen to be of ubiquitous presence in various fields utilizing nanotechnology. Over the earlier decades emphasis was laid upon the applications of these nanotubes in many consumer products and technology and their effects on the environment and living organisms including plants have mostly been ignored. Recent studies have concluded with diverse effects of CNTs in different plant species. One of the key aspects of these nanotubes important for studying phytotoxic effects is their poor dispersion in water due to their hydrophobic nature. Various researchers have used surfactants to keep nanotubes in suspension in water. However, these surfactants have ill effects on the plants. Moreover, in order to best simulate the natural set up of plant-nanotube interactions we have used three compounds to study the aqueous dispersion of nanotubes which are found in the soil: Natural Organic Matter (NOM), Humic Acid (HA) and Gum Arabic (GA). A general two factor factorial experiment was conducted to study the normalized absorbance (I/I_0) of carbon nanotubes in aqueous solutions of the three dispersing agents for different concentrations measured by a UV-Visible Spectrophotometer to conclude that optimum response was obtained from NOM at a concentration of 30 mg/L in water.

Author: Suvash Shiwakoti, Kaycie M. Schmidt, & Peter W. Bergholz

Title: Evaluation of Quantitative PCR (qPCR) with Propidium Monoazide (PMA) for the Specific Detection of Live Listeria monocytogenes in Soil

Listeria monocytogenes (Lm) is a foodborne pathogen associated with fresh produce commodities. Environmental models cannot yet account for meteorological effects on Lm abundance in farms. qPCR can quantify this, but DNA from dead cells interferes with this method. After establishing an effective qPCR protocol, we tested PMA to differentiate live and dead Lm in agricultural soils. A probe for the prs gene was effective at quantifying Lm DNA, but the soil matrix rendered PMA ineffective.

Author: Qi Wang, Yarong Yang, & Bin Guo

Title: Using Imputed MicroRNA Regulation Based on Weighted Ranked Expression and Putative MicroRNA Targets to Select MicroRNAs for Predicting Prostate Cancer Recurrence

Imputed microRNA regulation based on weighted ranked expression and putative microRNA targets (IMRE) is a method to predict microRNA regulation from genome-wide gene expression as well as predict microRNA putative targets. A p-value for each microRNA is calculated using the expression of the microRNA putative targets to analyze the regulation between different conditions. The dataset used in this study is GSE10645, which is the gene expression microarray of tumors from 596 patients with prostate cancer. This dataset includes the information of three different phenotypes: PSA (Prostate-Specific Antigen recurrence), Systemic (Systemic disease progression) and NED (No Evidence of Disease) and the expression level of the 1024 unique genes in the prostate cancer specimens of these patients. We used the IMRE method to analysis the GSE10645 dataset and identified several microRNA candidates that can be used to predict PSA recurrence and systemic disease progression in prostate cancer patients.