



Conference on **Applied Statistics**
in Agriculture and Natural Resources

MAY 16–19, 2022 | *Utah State University, Logan, Utah*

Elevate | **EVALUATE**

Thank You to Our Conference Sponsors



Utah Agricultural Experiment Station
UtahStateUniversity



National Institute of Food and Agriculture
U.S. DEPARTMENT OF AGRICULTURE



Ecology Center
UtahStateUniversity.

Drs. Vicky and
Edzard van Santen

This conference is supported by Data Science for Food and Agricultural Systems Program grant no. 2022-67022-36954 from the USDA National Institute of Food and Agriculture.

Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the sponsors.

BLUEZONE WiFi Guest Access Instructions

Visitors and guests may connect to the USU Network and Internet throughout the USU Logan campus, free of charge.

1. Navigate to **https://bluezone.usu.edu**
2. Click on the bottom button labeled '**Guest Access**'
3. You will be taken to a new page. After reading the Conditions of Use, click on '**I Agree**'

1 visit **bluezone.usu.edu**

UtahStateUniversity

Connect to the Internet at USU

By continuing below, you agree to USU's [Acceptable Use Policy](#).

STUDENTS, FACULTY, AND STAFF
A-number and password required.

EDUROAM
Preferred Wireless
Phones, Tablets, Laptops, etc.

BLUEZONE
Desktops, Servers, Ethernet
Smart TVs, Consoles, Printers, etc.

2 GUEST ACCESS
No registration required. Filtered access for 24 hours.

Need Help? [Contact the IT Service Desk](#)

UtahStateUniversity

BlueZone Guest

The Network is:

- Content filtered
- Free to use

Conditions of Use:

- Legal** - network services may not be used for illegal activities, such as downloading or distributing pirated videos, music, or other media.
- Non harmful** - network services may not be used for harmful activities, such as phishing, virus or malware distribution, spamming, or to deliberately mislead or harm others.

Consequences of Misuse:

- Disabled** - your connection will be terminated.

By using a device on USU's network, you agree to abide by these conditions and USU's [Acceptable Use Policy](#).

Expiration: tomorrow at 2am

[Need Help?](#)

3 I Agree

Remember

- protect yourself, your computer and never share information with anyone
- use secure websites (<https://>) whenever possible
- don't download files from untrusted sources
- be an internet skeptic

Agenda

All presentations are in Taggart Student Center (TSC) Ballroom

Day 1 - Monday, May 16, 2022

7:30 am to 8:30 am	Check-in and Morning Refreshments
8:30 am to 10:00 am	Workshop: Regression and Classification Applied to Precision Agriculture by Prof. Guilherme J.M. Rosa
10:00 am to 10:30 am	Break with Snacks
10:30 am to 12:00 pm	Workshop (cont'd)
12:00 pm to 1:30 pm	Group Lunch (TSC/International Lounge)
1:30 pm to 3:00 pm	Workshop (cont'd)
3:00 pm to 3:30 pm	Break with Snacks
3:30 pm to 4:30 pm	Workshop (cont'd)

Day 2 - Tuesday, May 17, 2022

7:30 am to 8:30 am	Check-in, Morning Refreshments, and Poster Set-Up (TSC)
8:30 am to 9:35 am	Opening Ceremony Welcome and Opening Remarks Keynote Address: Deep Neural Network Applied to Genome-Enabled Prediction of Complex Phenotypic Traits: Why its Performance has Been Limited so Far by Prof. Guilherme J.M. Rosa
9:35 am – 10:05 am	Break with Snacks
10:05 am – 11:30 am	Session 1
11:30 am to 1:00 pm	Group Lunch (TSC/International Lounge)
1:00 pm to 2:30 pm	Session 2 Invited Talk: Recursive Computing Strategies Inspire New Model Specifications by Prof. Mevin B. Hooten
2:30 pm to 3:30 pm	Break with Snacks and in-person Poster Session
3:30 pm to 4:30 pm	Session 3
6:00 pm to 9:00 pm	Reception Dinner and Social w/drinks, hors d'oeuvres, and dancing (Marriott Riverwoods Conference Center, off campus at 615 South Riverwoods Parkway, Logan. Please bring valid ID and cash/credit card for cash bar).

Day 3 - Wednesday, May 18, 2022

7:30 am to 8:30 am	Check-in and Morning Refreshments (TSC)
8:30 am to 10:00 am	Session 1 Invited Talk: Advances in Statistics: 2 Steps Forwards, 3 Steps Backwards... by Prof. Normand St-Pierre
10:00 am to 10:30 am	Break with Snacks and in-person Poster Session
10:30 am to 12:00 pm	Session 2 Invited Talk: What are Animal Scientists Learning from Using Deep Learning? by Prof. Juan P. Steibel
12:00 pm to 1:15 pm	Group Lunch (TSC/International Lounge)
1:15 pm to 2:45 pm	Session 3 Invited Talk: Transcriptional Profiling of Two Contrasting Genotypes Uncovers Molecular Mechanisms Underlying Salt Tolerance in Alfalfa by Prof. Rakesh Kaundal
2:45 pm to 4:00 pm	Break with Snacks and Virtual Poster Session
4:00 pm to 5:00 pm	Session 4
5:00 pm to 6:00 pm	Walking Tour of USU Campus & Poster Take-Down

Day 4 - Thursday, May 19, 2022

7:30 am to 8:30 am	Check-in and Morning Refreshments (TSC)
8:30 am to 10:00 am	Session 1 Invited Talk: Model Averaging in Agriculture and Natural Resources: What it is, when it is useful, and when it is a distraction by Prof. Philip M. Dixon
10:00 am to 10:30 am	Break with Snacks
10:30 am to 11:30 am	Session 2: Panel Discussion Challenges and Opportunities of Agricultural Statistics in the Big Data Era
11:45 am to 1:00 pm	Group Lunch (TSC/International Lounge)
1:00 pm – 2:00 pm	Awards Ceremony and Concluding Remarks

Day 1 - Monday, May 16, 2022

8:30 am – 4:30 pm

**WORKSHOP: REGRESSION AND CLASSIFICATION APPLIED TO
PRECISION AGRICULTURE**

Guilherme Rosa, Professor

Department of Animal and Dairy Sciences

Department of Biostatistics & Medical Informatics

University of Wisconsin-Madison

Website: www.gjmrosa.org

Vast amounts of data are routinely collected in agriculture, including traditional operational farm data and, more recently, data from digital tools such as remote and on-site sensing technologies. Moreover, many additional sources of information can be combined with farm data, such as economic and weather variables. The integration and analysis of such data can generate important insights and data-driven decision tools for optimization of agriculture systems. Nonetheless, suitable statistical and data mining tools are required to cope with such large and complex observational datasets, which involve multicollinearities and redundancies, nonlinear relationships, and spatial and temporal dependencies. In this workshop, we will discuss regression and classification tools for both prediction and causal inference with numerical and categorical outcomes, with special interest on agricultural data applications. Specific modelling approaches will include generalized additive models, structural equation models, and mixed-effect models. Useful algorithms and fitting strategies to be discussed will comprise dimension-reduction techniques, regularization approaches such as penalized regression and Bayesian hierarchical methods, and cross-validation strategies for variable selection and model comparison. The methods and applications will be illustrated with examples using infrared spectroscopy data, satellite remote sensing, image analysis, and computer vision, among others.

8:30 am to 8:35 am

WELCOME AND OPENING REMARKS

Chris Davies, Professor

Director, Center for Integrated BioSystems
Associate Director, Utah Agricultural Experiment Station
College of Agriculture and Applied Sciences
Utah State University

8:35 am – 9:35 am

Moderator: Chris Davies

KEYNOTE ADDRESS: DEEP NEURAL NETWORK APPLIED TO GENOME-ENABLED PREDICTION OF COMPLEX PHENOTYPIC TRAITS: WHY ITS PERFORMANCE HAS BEEN LIMITED SO FAR?

Guilherme Rosa, Professor

University of Wisconsin - Madison

High density molecular marker panels have been satisfactorily used in prediction of complex traits in agriculture, both for selection and management decisions. Such marker information generally fit using either regularized regression, kernel methods, or some other machine learning techniques. Deep neural networks (DNN), a particular case of artificial neural networks (ANN) composed by multiple hidden layers, have recently gained attention in many prediction applications using multi-dimensional sets of input variables (predictors). In genome-enabled prediction of complex traits, however, the cross-validation performance of DNN quite often is not striking compared to other methods. In this presentation, we discuss some potential explanations for the often modest performance of DNN in the context of genome-enabled prediction and offer some potential strategies to enhance the results. An example with a sample of about 80K chickens genotyped for 60K genetic markers is used to illustrate the concepts.

Presenter Bio: Guilherme Rosa is a professor at the University of Wisconsin-Madison in statistics and data analysis applied to livestock. His research is focused on quantitative genetics and genomics, analysis of observational data in agriculture, and precision livestock farming.

Contact Information: Guilherme Rosa, Department of Animal and Dairy Sciences, University of Wisconsin-Madison, Madison, WI. Phone: 608-265-8617, www.gjmrosa.org. Email: grosa@wisc.edu.

TWO-DIMENSIONAL P-SPLINE SMOOTHING FOR SPATIAL ANALYSIS OF PLANT BREEDING TRIALS

Hans-Peter Piepho¹, Martin Boer², Emlyn R. Williams³

¹Institute of Crop Science, University of Hohenheim, Stuttgart, Germany

²Biometris, Wageningen University & Research, Wageningen, The Netherlands

³Statistical Consulting Unit, Australian National University, Canberra, Australia

Large agricultural field trials may display irregular spatial trends that cannot be fully captured by a purely randomization-based analysis. For this reason, paralleling the development of analysis-of-variance procedures for randomized field trials, there is a long history of spatial modelling for field trials, starting with the early work of Papadakis on nearest neighbour analysis, which can be cast in terms of first or second differences among neighbouring plot values. This kind of spatial modelling is amenable to a natural extension using splines, as has been demonstrated in recent publications in the field. Here, we consider the P-spline framework, focusing on model options that are easy to implement in linear mixed model packages. Two examples serve to illustrate and evaluate the methods. A key conclusion is that first differences are rather competitive with second differences. A further key observation is that second differences require special attention regarding the representation of the null space of the smooth terms for spatial interaction, and that an unstructured variance-covariance structure is required to ensure invariance to translation and rotation of eigenvectors associated with that null space. We develop a strategy that permits fitting this model with ease, but the approach is more demanding than that needed for fitting models using first differences. Hence, even though in other areas second differences are very commonly used in the application of P-splines, our conclusion is that with field trials first differences have advantages for routine use.

Presenter Bio: Hans-Peter Piepho is an applied statistician with more than 30 years of experience. He is interested in statistical procedures needed in plant genetics, plant breeding, and cultivar testing. Recent interests include spatial methods for field trials and experimental design for various applications.

Contact Information: Hans-Peter Piepho, University of Hohenheim, Germany, Phone: 0049-711-459-22386, Email: piepho@uni-hohenheim.de.

OVERVIEW OF OPTIMAL EXPERIMENTAL DESIGN AND A SURVEY OF ITS EXPANSE IN APPLICATION TO AGRICULTURAL STUDIES

Stephen J. Walsh

Department of Mathematics and Statistics, Utah State University

Optimal Design of Experiments is currently recognized as the modern dominant approach to planning experiments in industrial engineering and manufacturing applications. This approach to design has gained traction among practitioners in the last two decades on two-fronts: 1) optimal designs are the result of a complicated optimization calculation and recent advances in both computing efficiency and algorithms have enabled this approach in real time for practitioners, and 2) such designs are now popular because they allow the researcher to ‘design for the experiment’ by working constraints, cost, number of experiments, and the model of the intended post-hoc data analysis into the design definition, thereby creating designs with more practical meaning than classical or catalogue designs. In this talk, I will review the definition of optimal design, discuss recent computational advancements in this field, and provide a survey of the expanse of this design approach in the agricultural literature.

Presenter Bio: Dr. Walsh is a faculty member in the Department of Mathematics and Statistics at Utah State University. He has over a decade of experience practicing as a quality assurance and experimental design expert in laboratory environments. His Ph.D. provided an adaptation of the Particle Swarm Optimization to solving several difficult optimal design problems. His current research program focuses on the synergistic research bridge between design of experiments and machine learning algorithms.

Contact Information: Stephen J. Walsh, Department of Mathematics and Statistics, Utah State University, Logan, Utah. Email: steve.walsh@usu.edu.

SPECTRUM FOOD DATA CLASSIFICATION WITH KERNEL BASED REGULARIZED PARTIAL LEAST SQUARES LEARNING

Tahir Mehmood¹

¹School of Natural Sciences (SNS) National University of Sciences and Technology (NUST). Islamabad, Pakistan

Qualitative and quantitative aspects of food products vary among varieties and cultivars of food products. Spectroscopy is a rapid tool to determine the structure and functional groups of organic compounds of food product. Food product's varieties and cultivars classification require sophisticated classification methods based on spectrum data. Here we have proposed the kernel based regularized partial least square learning for the food spectrum classification. Partial least squares (PLS) has proved to be an effective method for discrimination and classification. For stable and reliable classification of food varieties and cultivars kernel based regularized PLS (KRPLS) is used. Additionally, KRPLS can do feature selection. Theoretical properties of KRPLS are discussed. Experimental results on three data sets including food products based on olive oils, strawberry, and coffee reveal that KRPLS should be the preferred method for classification purpose, as the classification accuracy of KRPLS in these scenarios was superior to standard PLS for classification.

Presenter Bio: Dr. Tahir Mehmood is currently working as Associate Professor (Statistics) in SNS, NUST, Islamabad, Pakistan. He earned his PhD from Norwegian University for Life Sciences (NMBU), Norway, in 2012. He is the author of several scientific articles and has specific interest in statistical learning.

Contact Information: Tahir Mehmood, Associate Professor, School of Natural Sciences (SNS), National University of Sciences and Technology (NUST), H-12, Islamabad. Phone: 00923315966869. Email: tahime@gmail.com; tahir.mehmood@sns.nust.edu.pk.

THE IMPACT OF COVID-19 ON THE PRODUCTION OF OFFICIAL STATISTICS: A CASE STUDY

Linda J. Young, Tara Murphy, Arthur Rosales, and Denise A. Abreu
USDA National Agricultural Statistics Service (NASS), Washington DC USA

When COVID-19 rates began to rise in early 2020, in-person interviews were stopped, and data collection calling centers were closed. Federal Statistical Agencies moved rapidly to revise data collection processes. The USDA National Agricultural Statistics Service (NASS) chose not to conduct the 2020 June Area Survey (JAS), its largest annual survey, which had always been conducted through in-person interviews. In October 2020, NASS decided to conduct the 2021 JAS without any in-person interviews. Administrative, remotely sensed, and other data was brought together in new data collection tools. Data collection procedures were revised in a myriad of ways, such as mailing maps to producers associated with sampled units in advance of receiving calls from interviewers. After the survey was completed, it was clear that many of the changes led to more efficient processes. Yet, the effect they had on data quality and the subsequent estimates remained an open question. In this presentation, a retrospective comparison of data quality in the pre-pandemic 2019 JAS and the pandemic 2021 JAS is conducted. The potential impact of these revisions on the post-pandemic JAS is discussed. Open questions are highlighted.

Presenter Bio: Dr. Young is Chief Mathematical Statistician and Director of Research and Development at the NASS. She worked extensively with university agricultural researchers prior to joining NASS. Her research has focused on the interface of statistics and science.

Contact Information: Linda J. Young, USDA NASS, Washington DC, Phone: 202-765-6116, Email: Linda.J.Young@usda.gov.

INVITED TALK: RECURSIVE COMPUTING STRATEGIES INSPIRE NEW MODEL SPECIFICATIONS

Mevin B. Hooten¹, Michael R. Schwob¹, Devin S. Johnson², and Jacob S. Ivan³

¹The University of Texas at Austin, Austin, TX, USA

²Pacific Islands Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Honolulu, HI, USA

³Colorado Parks and Wildlife, Fort Collins, CO, USA

Ecologists increasingly rely on Bayesian capture-recapture models to estimate abundance of wildlife populations. Capture-recapture models account for imperfect detectability in individual-level presence data. A variety of approaches have been used to implement such models, including integrated likelihood, parameter-expanded data augmentation, and combinations of those. Recently proposed conditional specifications have improved the stability of algorithms for fitting capture-recapture models. We arrive at similar conditional specifications of capture-recapture models by considering recursive implementation strategies that facilitate fitting models to large data sets. Our approach enjoys the same computational stability but also allows us to fit the desired model in stages and leverage parallel computing resources. Our model specification includes a component for the capture history of detected individuals and another component for the sample size. We demonstrate this approach using three examples including simulation and two data sets resulting from capture-recapture studies of different species.

Presenter Bio: Dr. Hooten is a Professor of Statistics and Data Sciences at UT-Austin. His research focuses on spatial and spatio-temporal statistics for ecological and environmental applications.

Contact Information: Mevin Hooten, Department of Statistics and Data Sciences, The University of Texas at Austin. Email:

mevin.hooten@austin.utexas.edu.

MODELING THE SPATIAL INVASIBILITY OF SLASH PINE FLATWOODS TO CHINESE TALLOW (*TRIADICA SEBIFERA*) INVASION

*Zhaofei Fan*¹, *Shaoyang Yang*

¹Auburn University, Auburn, AL, USA

²Virginia Tech University, Blacksburg, VA, USA

We developed a conceptual model to decipher the mechanism of Chinese tallow invasion based on the community assembly theory, modeling the spatial invasibility of slash pine flatwoods to Chinese tallow using a spatially-explicit point process model. A set of identified factors including the dispersal, overstory, and understory filters as well as tallow abundance and distribution were measured within a total area of 0.86-ha in a slash pine flatwood. The identified filters vary spatially and affect tallow invasion in a nonlinear “concave down” manner. Additive polynomial functions of these identified filters provide a viable approach to model the spatial invasibility of tallow invasion. The full models which include all filters, as well as the sub-models which include only the dispersal filters and either of the overstory and understory filters, seem to have an adequate predictive power based on selected diagnostic statistics. In contrast, the sub-models that include only one of the filters failed to precisely predict the observed invasion outcomes. The results support the importance of dispersal filters and reducing propagule pressure levels for effective tallow control and slowing of spread.

Presenter Bio: Dr. Fan is an Associate Professor in the School of Forestry and Wildlife Sciences, Auburn University. His research focuses on forest ecology and statistics, especially multiscale spatial modeling of biological invasion of invasive plants, and forest decline.

Contact Information: Zhaofei (Joseph) Fan, 602 Duncan Dr., Auburn University, Auburn, AL, Phone: 334-844-1014. Email: zzf0008@auburn.edu.

HANDLING NON-DETECTS WITH IMPUTATION: A SIMULATION STUDY

Rose Adjei¹ and John R. Stevens¹

¹Department of Mathematics and Statistics, Utah State University, Logan, UT, USA

A non-detect is an analytical sample where the concentration is deemed to be lower than could be detected using the method employed by the laboratory. Non-detects are sometimes referred to as censored data or below detection limits. Mostly, observations that fall below the limit of detection are not reported and can create missingness in data, thus necessitating further data analysis. Having a high level of non-detects in a dataset can be problematic as it can complicate the computations of descriptive statistics and testing statistical analyses. Considering the adverse effects non-detects can pose to statistical analysis, it is very crucial that researchers understand how to properly handle non-detects. Even though there are several existing techniques to deal with non-detects, it is essential that one chooses the appropriate technique carefully to avoid any false interpretations since non-detects are considered Missing Not at Random (MNAR) phenomenon. In my presentation, I will briefly outline some challenges associated with having non-detects in data and discuss some of the existing methods used to handle non-detects. A simulation study will be used to assess whether it will be appropriate to address non-detects in data using imputation technique or simply substitute with zeros. I will compare the various methods based on their Type 1 error and power effects at varying magnitudes of non-detects.

Presenter Bio: Rose Adjei is a fourth-year PhD student in Statistics at Utah State University. She earned her MS in Statistics from Northern Arizona University. Her research interests are in Bioinformatics, Biostatistics, and Data Mining.

Contact Information: Rose Adjei, Dept of Mathematics and Statistics, Utah State University, Logan UT, Phone: 435-797-2908. Email: rose.adjei@usu.edu.

DETECTION AND ATTRIBUTION OF EXTREME CLIMATE CHANGE IN UNITED STATES

Wei Zhang, Grace Affram

Department of Plants, Soils and Climate, Utah State University, Logan, UT, USA

Detection and attribution of climate change signals is still challenging due to a lack of sufficient observations and the existence of internal variability in climate data. Here I will report a few recent studies that attribute extreme climate events across the United States to the forcing climate change by the application of statistics and climate models. Specifically, attribution analysis reveals that this decrease in wetter and colder weather is a likely result of anthropogenic forcing. The statistical framework has been applied to the central US major flooding during spring 2019, which is caused by an extremely high frequency of a weather pattern (Midwest Water Hose). The rising trend and the high frequency of the weather regime can be mainly attributed to the rising greenhouse gases caused by human activities, rather than natural forcing.

Presenter Bio: Dr. Zhang is an assistant professor at Utah State University with extensive experience with applying statistics and numerical models to the detection and attribution of climate change.

Contact Information: Wei Zhang, Department of Plants, Soils and Climate, Utah State University, Logan, UT, USA. Phone: 435-797-1101, Email: w.zhang@usu.edu.

3:30 pm to 4:30 pm Session 3 Moderator: Raul Macchiavelli

IN-VITRO ANALYSIS OF CANNABIDIOL AND RESVERATROL ANTIOXIDANTS AGAINST DIESEL EXHAUST PARTICLE-INDUCED CYTOTOXICITY OF HUMAN A549 LUNG CELLS BASED ON MACHINE LEARNING RAMAN SPECTROSCOPY

Ashton Young¹, Emily Brothersen¹, Wei Zhan¹, and Anhong Zhou¹

¹Utah State University, Logan, UT, USA

Urban air pollution is a significant health risk factor, and diesel exhaust particles (DEP) are a contributor to such. DEP exposure enhances pneumonia risk, leads to chronic degradation of alveolar walls, and can cause lung inflammation. Antioxidants such as resveratrol (RES) and cannabidiol (CBD) have been shown to inhibit oxidative damage in cells effectively. Raman Spectroscopy (RS), a modern non-invasive and non-destructive analytical technique, was used in this analysis to measure the antioxidant effect and cytotoxicity of DEP and resveratrol and cannabidiol on human lung cells. RS can generate massive datasets to which machine learning, a subfield of artificial intelligence, can be a powerful tool for developing, testing, and refining experimental results. In our analysis of RS data, we use Potential of Heat-diffusion for Affinity-based Transition Embedding (PHATE) and mutual information (MI) in addition to the conventional ML methods (LDA, kNN, RF and SVM) to investigate the effects of DEP exposure and the protectants RES and CBD on A549 human lung cells.

Presenter Bio: Ashton Young is a student at Utah State University currently enrolled in the PhD Biological Engineering program.

Contact Information: Ashton Young, Utah State University, Logan, UT, Phone: 435-730-2751, Email: A00302384@usu.edu.

VALID PREDICTIONS OF GROUP-LEVEL RANDOM EFFECTS

Nicholas Syring¹, Fernando Miguez², and Jarad Niemi¹

¹Iowa State University, Dept. of Statistics, Ames, IA USA

²Iowa State University, Dept. of Agronomy, Ames, IA USA

³Iowa State University, Dept. of Statistics, Ames, IA USA

Gaussian linear models with random group-level effects are the standard for modeling randomized experiments carried out over groups, such as locations, farms, hospitals, or schools. Group-level effects can be summarized by prediction intervals for group-level means or responses, but the quality of such summaries depends on whether the intervals are valid in the sense they attain their nominal coverage probability. Many methods for constructing prediction intervals are available—such as Student-t, bootstrap, and Bayesian methods—but none of these are guaranteed to be valid, and indeed are not valid over a range of simulation examples. We propose a new method for constructing valid predictions of group-level effects based on an inferential model (IM). The proposed prediction intervals have guaranteed finite-sample validity and outperform existing methods in simulation examples. In an on-farm agricultural study, the new IM-based prediction intervals suggest a higher level of uncertainty in farm-specific effects compared to the standard Student-t based intervals, which are known to under cover.

Presenter Bio: Dr. Syring is an Assistant Professor in the Department of Statistics at Iowa State University. His research interests include the properties of statistical estimators under model misspecification and the construction of exact testing procedures.

Contact Information: Nicholas Syring, Dept. of Statistics, Iowa State University, IA, Email: nsyring@iastate.edu.

PRINCIPAL RESPONSE CURVE ANALYSIS OF COMMUNITY ARTHROPOD ABUNDANCE DATA CONTAINING SPARSE SUBSETS

Changjian Jiang, C. R. Brown, P. Asimwe, Chen Meng, Adam W. Schapaugh

Principal response curve (PRC) analysis was applied to an assessment of the ecological impact of a genetically-modified insect-resistant cotton on predatory communities in the field. Ten arthropod taxa from the community were collected ten times across the season at six sites, in which zero abundance were observed for individual taxon at least 25% of the time (unique site x collection combinations). These complete absences and those nearly so, called sparse subsets of the data in this investigation, were the result of geoclimatic and seasonal variations, which are both independent of the treatment effects of the insecticide application and cotton variety for which the PRC analysis is intended. If the sparse subsets were included in the analysis, the treatment effect would be underestimated. Here, a modified PRC analysis is proposed to remove those sparse subsets and implemented on data with incomplete responses. In the application to the motivating example, four components (PRC1-4) were significant at the 5% level by the modified method, when more than 50% of the data were excluded due to no- or low responses, and five (PRC1-5) by the classical method. While PRC1-2 was highly consistent between two methods, PRC3-5 was largely different because of sparse subsets of the data. Comparing results between two methods indicates that excluding sparse subsets prevented the underestimation of the treatment effect and the subsequent bias of the estimated relationship with the community due to the environmental variation that caused the sparse data. In this regard, the modification should be considered as a supplement of the classical PRC analysis and recommended when abundance data has sparse subsets.

Presenter Bio: Dr. Jiang is the lead statistician at Bayer Crop Science US working in the Regulatory Science, Data Science, and Digital Transformation departments.

Contact Information: Jiang Changjian, Bayer Crop Science US, Saint Louis, MO 63141, Phone: 636-675-1950, Email: changjian.jiang@bayer.com.

8:30 am to 10:00 am Session 1 Moderator: Hans-Peter Piepho

INVITED TALK: ADVANCES IN STATISTICS: 2 STEPS FORWARD, 3 STEPS BACKWARD...

Normand R. St-Pierre¹

¹Perdue AgriBusiness, Salisbury, MD, and The Ohio State University, Columbus, OH, USA

In the 'old days,' statistics were done using a calculator. An experimental designs book was used to find the formulas to compute sums of squares and test statistics. There was only one way to analyze a 5 x 5 Latin square. It was wrong, but it was *uniformly* wrong. Today, advanced methods of data analysis have been developed that provide better estimates and more correct tests when applied properly. The issue is that they appear to frequently be applied improperly by non-statisticians. In a large proportion of published papers, the description of the statistical methods used is so insufficient that a professional statistician could not replicate the analyses. The hunt for ' $P < 0.05$ ' is rampant. For example, the selection of the best error var-cov matrix in repeated measurements is often based on which structure gives the smallest *P-value* for the treatment effect. I will present numerous examples of statistical mis-applications found in papers published in top animal and nutrition journals. I will discuss some of the root causes that made the application of statistical methods worse than it was 50 years ago, and I will suggest ways that professional statisticians can help to remedy this situation.

Presenter Bio: Dr. St-Pierre is Director of R&D for Perdue AgriBusiness and Professor Emeritus at The Ohio State University. He is not a professional statistician, but he tried to translate statistics to non-statisticians throughout his career.

Contact Information: Normand St-Pierre, 5053 Canton Rd., P.O. Box 242, Tangier, VA-23440. Phone:443-437-9526, Email: normand.st.pierre@gmail.com.

WHY DON'T R AND SAS PRODUCE THE SAME RESULTS? UNDERSTANDING WHERE AND WHY THEY DIFFER IN THEIR BEHAVIOR AND HOW TO MAKE THEM AGREE

Simon Riley¹, *Edzard van Santen*¹, *Stephen R. Bowley*², and *Ruth Hummel*³

¹University of Florida, Gainesville, FL, USA

²University of Guelph, Guelph, ON, Canada

³JMP Statistical Discovery LLC, Cary, NC, USA

Researchers seeking to reproduce in R an analysis originally conducted in SAS (or *vice versa*), who operate in “multilingual” software environments, or who otherwise transition frequently between R and SAS, are often frustrated by the extent to which the different software packages generate divergent results when performing seemingly equivalent analytical tasks. The authors, in preparing an applied statistics text to be published by the ASA, CSA, SSSA, have as a result been forced to identify, describe, and “solve” many such discrepancies. This presentation offers to share some of the lessons learned from that process, including when to expect that SAS and R are giving different answers because they’re answering different questions (and, conversely, when to suspect user error), as well as some general guidelines for how to determine whether and how divergent results can be reconciled after encountering a discrepancy for the first time. These lessons are further illustrated via “case studies” of statistical procedures frequently employed by researchers in the agricultural and natural resource sciences and the analysis of designed experiments more broadly.

Presenter Bio: Simon Riley is a doctoral candidate in the Agronomy Department of the University of Florida, where he holds a graduate assistantship appointment at the Statistical Consulting Unit of the Institute for Food and Agricultural Science.

Contact Information: Simon Riley, University of Florida, Gainesville, FL, Email: simon.riley@ufl.edu.

BAYESIAN ANALYSIS OF AGRICULTURAL RISK MITIGATION VIA ADOPTION OF REGENERATIVE SOIL HEALTH PRACTICES

Sarah Manski², Ben Goldstein¹, Yvonne Socolar¹, Harley Cross³, Katie Fettes,³ Aria McLauchlin³, Timothy Bowles¹, and Frederi Viens²

¹UC Berkeley, Berkeley, CA, USA

²Michigan State University, East Lansing, MI, USA

³Land Core, Los Angeles, CA, USA

It is widely acknowledged that adoption of soil health practices (cover cropping, conservation tillage, diverse crop rotations, etc.) improves resilience to flood and drought, stores carbon in the soil, and reduces economic risk. But lenders and insurers do not incorporate these benefits into their risk assessments. We build predictive models to quantify the risk-mitigation value of adopting soil health practices. Initial analysis in Illinois shows several Bayesian regression models are predictive and robust at a county scale. The Bayesian framework allows for accurate and efficient model uncertainty quantification, a key feature to evaluate risk. Corn yield is strongly predicted by a water-availability variable; higher rotational complexity is associated with downside-risk mitigation for 75% of counties in dry conditions and for 60% of counties across all conditions. We are building a tool with these models, in partnership with a Midwestern farm lender, to create the economic rationale to provide financial incentives for producers using climate-resilient de-risking practices, helping with the often cost-prohibitive adoption of these soil health practices for US farmers.

Presenter Bio: Sarah Manski is a PhD student at Michigan State University in the Department of Statistics and Probability. Sarah's research includes applied Bayesian analysis in agriculture, probability theory, and statistics education.

Contact Information: Sarah Manski, Michigan State University, East Lansing, MI
Email: manskisa@msu.edu.

BIAS PROTECTION IN MIXED-LEVEL FRACTIONAL FACTORIAL EXPERIMENTS

Vinny Paris¹ and Max D. Morris¹

¹Iowa State University, Ames, IA, USA

Fractional factorial experiments are often used when the total number of possible unique trials is too large. This leads to possible bias in the estimates from unaccounted for (and likely inestimable) higher order interactions. The foldover technique, introduced by Box and Wilson, allows for complete de-aliasing of main effects from second order interaction effects for 2-level. This has the advantage that main effect estimates will be unbiased so long as third order and higher interaction effects are zero. Recent advances of extending the foldover technique to regular and irregular mixed-level factorial designs will be discussed. A use case of planning a multi-year factorial crop study will be presented

Presenter Bio: Vinny Paris is a Ph.D. student at Iowa State University with research in experimental design, the focus being bias protected factorial studies. Fun trivia is he has had code ran on Summit, the 2nd strongest supercomputer in the world currently (2/26/22).

Contact Information: Vinny Paris, Iowa State University, Ames, IA, Email: vinny@iastate.edu.

INVITED TALK: WHAT ARE ANIMAL SCIENTISTS LEARNING FROM USING DEEP LEARNING?

Juan P. Steibel¹ and Junjie Han¹

¹Michigan State University, East Lansing, USA.

Deep learning (DL) is a type of machine learning method that uses artificial neural networks for representation learning. DL can be used for supervised learning tasks such as prediction of continuous or ordinal outcomes and classification and for unsupervised learning such as dimension reduction and clustering. DL methods are flexible and can learn cryptic data structures.

Animal scientists and animal breeders are now using deep learning for several applications, including genomic prediction, image classification and segmentation for animal phenomics and genotype imputation. In this talk, we present three different applications of DL in swine management and genetics, covering a range of machine learning tasks including: 1) genomic prediction (prediction of a continuous outcome), 2) image and video classification for behavioral phenotyping (multi-class classification with imbalanced and structured data), and 3) detection of crossbreeding among purebred animals (anomaly detection). Through presenting these examples we illustrate the key aspects of the application of DL, including model tuning and using cross-validation to assess the scope of validity of results.

Presenter Bio: Juan P. Steibel is an Associate Professor of Animal Science and Fisheries and Wildlife at Michigan State University. He investigates the development and application of statistical and computational methods to advance swine breeding through the integration of multi-omics data streams.

Contact Information: Juan P. Steibel, Departments of Animal Science and of Fisheries & Wildlife, Michigan State University, East Lansing, MI. Phone: 517-899-7564, Email: steibelj@msu.edu.

A COMPUTER VISION FRAMEWORK FOR REAL-TIME MONITORING GROWTH OF FINISHING CATTLE IN FEEDLOTS

*Joseane Padilha¹, Arthur F. A. Fernandes², Joao R. R. Dorea¹, Tiago S. Acedo³,
Guilherme J. M. Rosa¹*

¹Department of Animal and Dairy Sciences, University of Wisconsin-Madison,

²Cobb-Vantress Inc.,

³DSM Nutritional Products Brazil

Feed management and marketing decision in beef feedlots are based on visual appraisal of cattle growth and development. Such a subjective approach is used as weighing cattle regularly is labor intensive and also stressful for the animals. An alternative in this regard could be the use of computer vision, to extract animal biometric measurements from 3D images. Such features could then be used as input variables in a prediction model for body weight. Therefore, our group developed an automated system for 3D image acquisition for real-time monitoring of cattle growth in feedlots. However, implementing this technology outdoors and with unrestrained animals poses a great challenge because of harsh conditions for the equipment and varying lighting and animal movement. Here we propose a pre-processing framework to standardize image quality using image composition, a deep learning classifier to select images with only one animal per frame, and a Marker-Controlled Watershed Segmentation to extract the animal body and its biological features for future predictions. A total of 17,754 images were collected, from which 3,304 were deemed usable. The algorithm successfully extracted the animal body from each image and obtained the biological features, with correlations in the range of 0.60-0.90.

Presenter Bio: Joseane Padilha is a PhD student at the University of Wisconsin-Madison. As a statistician, Joseane is interested in modern tools for modeling and understanding complex datasets, in particular artificial intelligence techniques to face challenges in livestock production.

Contact Information: Joseane Padilha, PhD Student, University of Wisconsin-Madison, Animal Science Building, Madison, WI, Phone: (608) 265-8617. Email: jpdasilva@wisc.edu.

PREDICTIVE MODELS FOR WEEKLY CATTLE MORTALITY AFTER ARRIVAL TO THE FEEDLOT USING RECORDS, WEATHER, AND TRANSPORT DATA

Lauren Wisnieski^{1,2}, David E. Amrine^{2,3}, and David G. Renter²

¹Lincoln Memorial University, Harrogate, TN, USA

²Center for Outcomes Research and Epidemiology, Manhattan, KS, USA

³Beef Cattle Institute, Manhattan, KS, USA

To the best of our knowledge, there are no publications on predictive models for weekly all-cause mortality in feedlot cattle. In this study, random forest models to predict weekly mortality for cattle purchase groups from arrival (Day 1) to Day 42, and cumulative mortality from Day 43 until slaughter were built using records, weather, and transport data available at the time of purchase. Models were evaluated by calculating the root mean squared error (RMSE) and accuracy (as defined as the percent of purchase groups that had predictions within 0.25% and 0.50% of actual mortality). The models had high accuracy (>90%), but the RMSE estimates were high (range = 1.0% to 4.1%). The strongest predictors were maximum temperature and purchase weight. The models performed well among purchase groups with low weekly mortality but performed poorly in high mortality purchase groups. The models may potentially have utility as a screening tool for very low mortality purchase groups after arrival. Future studies should consider building iterative models that utilize the strongest predictors identified in this study.

Presenter Bio: Dr. Lauren Wisnieski is an Assistant Professor of Public Health and Research in the College of Veterinary Medicine at Lincoln Memorial University. Her interests include predictive modeling, One Health topics, and statistical analysis.

Contact Information: Lauren Wisnieski, College of Veterinary Medicine, Lincoln Memorial University, Harrogate, TN, Phone: 423-869-6770, Email:

lauren.wisnieski@lmunet.edu.

MALIVHU: A WEB SERVER FOR PREDICTING HUMAN-VIRUS PROTEIN-PROTEIN INTERACTIONS USING MACHINE LEARNING ALGORITHMS

David Guevara-Barrientos¹² and Rakesh Kaundal¹²

¹Utah State University, Logan, UT, USA

²Kaundal Artificial Intelligence and Advanced Bioinformatics Lab, Logan, UT, USA

Covid-19 pandemic has become less of a threat, but it seems that humans will have to learn how to live with the virus around. There is a need to develop treatments against it, not only vaccines, but also medicines. Human-virus protein-protein interactions (PPIs) play a key part in drug-target discovery, but finding them experimentally can be either costly or unreliable. Therefore, computational methods arose as a powerful alternative to predict these interactions, reducing costs, so researchers focus on confirming only certain interactions instead of trying all possible combinations in the laboratory. Malivhu uses machine learning models to predict human-SARS-CoV/SARS-CoV-2/MERS PPIs, running the virus proteins through a 3-phase process, where phase 1 filters ssRNA(-) type virus proteins, phase 2 filters Coronaviridae family proteins and phase 3 filters SARS and MERS species proteins. The performance of the models was measured with Matthews correlation coefficient, f1-score, specificity, sensitivity, and accuracy scores, getting accuracies of 99.08%, 99.8% and 100% for the first three phases respectively, 94.7% for human-SARS-CoV PPI, 94.68% for human-SARS-CoV-2 PPI and 96.97% for human-MERS PPI.

Presenter Bio: David Guevara is a Computer Science PhD student at Utah State University with a bachelor's degree in Computer Engineering and a MSc in Computational Biology. He has worked in software development for 8 years, contributing to the development of the genome assembly module for NGSEP.

Contact Information: David Guevara-Barrientos, Utah State University, Logan, UT, Phone: 435-213-73, Email: david.guevara@usu.edu.

INVITED TALK: TRANSCRIPTIONAL PROFILING OF TWO CONTRASTING GENOTYPES UNCOVERS MOLECULAR MECHANISMS UNDERLYING SALT TOLERANCE IN ALFALFA

Rakesh Kaundal, Naveen Duhan, Biswa R. Acharya, Manju V. Pudusery, Jorge F. S. Ferreira, Donald L. Suarez & Devinder Sandhu

¹Department of Plants, Soils and Climate, Utah State University, Logan, UT, USA.

²US Salinity Lab (USDA-ARS), 450 W Big Springs Road, Riverside, CA, USA.

³College of Natural and Agricultural Sciences, University of California Riverside, Riverside, CA, USA.

Alfalfa is an important forage crop that is moderately tolerant to salinity; however, little is known about its salt-tolerance mechanisms. We studied root and leaf transcriptomes of a salt-tolerant (G03) and a salt-sensitive (G09) genotype, irrigated with waters of low and high salinities. RNA sequencing led to 1.73 billion high-quality reads that were assembled into 418,480 unigenes; 35% of which were assigned to 57 Gene Ontology annotations. The unigenes were assigned to pathway databases for understanding high-level functions. The comparison of two genotypes suggested that the low salt tolerance index for transpiration rate and stomatal conductance of G03 compared to G09 may be due to its reduced salt uptake under salinity. The differences in shoot biomass between the salt-tolerant and salt-sensitive lines were explained by their differential expressions of genes regulating shoot number. Differentially expressed genes involved in hormone-, calcium-, and redox-signaling, showed treatment- and genotype-specific differences and led to the identification of various candidate genes involved in salinity stress, which can be investigated further to improve salinity tolerance in alfalfa. Validation of RNA-seq results using qRT-PCR displayed a high level of consistency between the two experiments. This study provides valuable insight into the molecular mechanisms regulating salt tolerance in alfalfa.

Presenter Bio: Dr. Rakesh Kaundal is an Assistant Professor in the Department of Plants, Soils, and Climate. He is also the Director of Bioinformatics Facility in the Center for Integrated BioSystems (CIB) at Utah State University. His research focuses on collaborative program in bioinformatics & computational biology.

Contact Information: Rakesh Kaundal, Bioinformatics Facility Center for Integrated BioSystems College of Agriculture and Applied Sciences, Logan, UT, Phone: (435) 797-4117, Email: rkaundal@usu.edu.

SVEN: A NEW TOOL FOR GWAS

*Dongjin Li, **Somak Dutta** and Vivekananda Roy*

Iowa State University, Ames, I.A., USA

Genome-wide association studies require identification of important markers related to a particular trait. However, the number of available markers far exceeds the number of observations. We develop a Bayesian variable selection method, called SVEN, based on a hierarchical Gaussian linear model with priors placed on regression coefficients as well as on the model size. We achieve sparsity by placing degenerate spike priors for the inactive markers and achieve shrinkage of the effect sizes using Gaussian priors. We develop a fast computational framework for fast identification of the posterior high-probability regions, inference, and prediction. An appealing byproduct of SVEN is model weight adjusted trait prediction intervals. The performance of SVEN is demonstrated through a real data example from agriculture with nearly half-a-million markers.

SVEN is implemented in a publicly available R package called *bravo* available at <https://cran.r-project.org/package=bravo>

Presenter Bio: Dr. Dutta is an Associate Professor and a fellow of the Plant Science Institute at Iowa State University. His research interest includes spatial statistics, high-dimensional regression, machine learning, and matrix-free computations.

Contact Information: Somak Dutta, Iowa State University, Ames, IA, Phone: 515-294-4998, Email: somakd@iastate.edu.

PREDICTING BLUEBERRY COLD HARDINESS IN WASHINGTON

Clark Kogan¹, Lisa DeVetter², Nora M. Bello³ and Gwen Hoheisel⁴

¹Washington State University, Spokane, WA, USA

²Washington State University, Mt Vernon, WA, USA

³Ohio State University, Columbus, OH, USA

⁴Washington State University, Prosser, WA, USA

Freezing temperatures in fall, winter, and spring in Washington can cause damage to multiple perennial crops like blueberries, grapes, and cherries. Predictive modeling for lethal temperatures allows producers to make informed decisions about cold mitigation practices. If buds are hardier than air temperatures, mistaken use of propane heaters and wind machines is costly. In turn, aggressive heater and wind machine mitigation under high probability of damaging temperature can prevent crop damage and potential yield loss.

For this study, multiple years of experimental cold hardiness data were collected on four cultivars of highbush blueberry. Freeze chambers were used to systematically reduce bud temperatures, after which buds were dissected and bud survival was assessed. A generalized linear mixed model with a binomial response and logit link was fit to each cultivar to characterize the relationship between bud survival, freezer temperature, recent air temperature and growing degree days. Model simulation was performed to obtain marginal scale lethal temperature estimates. Model error estimation was performed using cross validation.

Presenter Bio: Dr. Kogan is the owner of StatsCraft LLC and an adjunct faculty member at Washington State University in the College of Pharmacy and Pharmaceutical Sciences. He has extensive experience working as a statistical consultant in agriculture, energy and medical settings.

Contact Information: Clark Kogan, Washington State University, Spokane, WA, Phone: 509-389-9649. Email: clark.kogan@wsu.edu.

CORRELATION MODELING OF HIGH-THROUGHPUT PHENOTYPE DATA BY MM ALGORITHM AND ITS APPLICATION TO PHOTOSYNTHESIS

Abhijnan Chattopadhyay¹, David M. Kramer¹, Tapabrata Maiti¹, and Samiran Sinha²

¹Michigan State University, East Lansing, MI, USA

²Texas A&M University, College Station, TX, USA

Quantitative genomics experiments aim to reveal underlying mechanisms that link genotypic variations with multiple biological responses (phenotypes). Interactions among various phenotypes give new insights into how genetic diversity may have tuned biological processes to enhance fitness under diverse conditions. Analyses of high-dimensional datasets generated by high-throughput phenotyping and genome sequencing tools require new computational algorithms and software tools. In this paper, we propose to model the correlations among multiple complex phenotypes as a function of genetic and environmental explanatory variables. We develop an efficient estimation methodology called Correlation Modeling under Pairwise Likelihood Estimation (CMPLE), aided by a novel Minorize-Maximize (MM) algorithm, and provide statistical inference techniques. Simulation studies mimicking biological data show that the method is beneficial to recover pertinent information, including different regulatory pathways, and is computationally efficient to handle many parameters. Our approach is also illustrated by analyzing a motivating dataset from recombinant inbred lines of cowpea (*Vigna unguiculata.*), where we identify certain genetic variations affecting distinct biological mechanisms.

Presenter Bio: Abhijnan Chattopadhyay is a Ph.D. candidate in Statistics, working with the Michigan State University Plant Research Lab. He has worked on numerous projects on analyzing high-throughput phenotype data and has developed interpretable models on thermo-tolerance in plants.

Contact Information: Abhijnan Chattopadhyay, Michigan State University, East Lansing, MI, Phone: 517-940-1736, Email: chattop3@msu.edu.

THE POWER OF DATA VISUALIZATION: REVEALING RANDOM-EFFECT STRUCTURES THROUGH FACETING

Miranda Tilberg¹

¹Iowa State University, Ames, IA, USA

Linear mixed-effect models (LMMs) are commonly used to analyze data across all disciplines. Random effect structures are often inferred from aspects of the experimental design, such as the experimental units associated with repeated measures, but subjective decisions may arise. In this talk, I discuss the benefits of data visualization—in particular, faceting—for determining LMM random-effect structures. The motivating example for this talk comes from an experiment designed to study in-barn heat processing of mass swine mortalities. Six swine carcasses, evenly divided into three weight classes and two barn locations, were fit with temperature sensors in various body parts: brain, pleura, peritoneal, ham, femur bone marrow, and skin surface. Temperatures were recorded every 15 minutes for 14 days to determine the effects of treatments on body temperatures. Preliminary visualizations of the data revealed odd clusters of roughly linear patterns. However, faceting unexpectedly revealed clear linear patterns that ultimately improved the random effect structure of the LMM. As a takeaway, the importance of thorough data visualization for informing an analysis should not be underestimated.

Presenter Bio: Miranda Tilberg received her MS in Statistics from Iowa State University in 2019 and hopes to complete her PhD by the time you're reading this. She has consulted with ISU's statistics department for two years and enjoys collaborations across a variety of disciplines.

Contact Information: Miranda Tilberg (Tilton), Iowa State University, Ames, IA, Phone: 763-244-7901, Email: tiltonm@iastate.edu.

NEW FEATURES AND CAPABILITIES IN JMP V17

Kevin Potcner¹

¹JMP Statistical Discovery, LLC

JMP is a point-and-click GUI statistical software developed by SAS in 1989. JMP is designed so that statistical analyses, from basic to advanced, can be done by myriad users without the need for programming. JMP's exploratory approach to data analyses is aligned with how most analyses are done in practice – through interactive dynamic visualizations and analyses with the results of each step of the analysis leading to a set of tools and options appropriate for the next step in the analysis workflow. JMP runs on Windows and Macintosh operating systems and also functions as an interface to SAS®, R, Python, MATLAB and Excel.

The presenter will showcase new features and capabilities in JMP v17 scheduled to be released this Fall.

Presenter Bio: Kevin Potcner is a statistical scientist at JMP Statistical Discovery, LLC supporting the use of JMP in teaching and research at colleges and universities across the U.S. and Canada.

Contact Information: kevin.potcner@jmp.com.

A ROBUST CLUSTERING METHOD USING COMPOSITIONAL DATA RESTRICTIONS: STUDYING WOOD PROPERTIES IN THE REFORESTATION OF PORTUGAL

Pamela M. Chiroque-Solano¹

²CITAB: Centre for the Research and Technology of Agro-Environmental and Biological Sciences. Universidade de Trás-os-Montes e Alto Douro, Portugal.

In compositional regression, it is important to detect which part of a whole presents high variability to help to understand some patterns about the complete process. This work proposes a procedure to identify clusters with observation with similar composition while also accommodating the data's natural compositional restriction. A probabilistic Dirichlet regression model combining the usual compositional model with a grouping technique was developed. The inference procedure was done under the Bayesian approach using the Hamiltonian Monte Carlo (HMC) method to obtain samples from the posterior marginal distributions of interest. To avoid model identifiability issues, decision theory was used to provide support in choosing one component as a reference. This work was motivated by Populations of *Pinus nigra* planted in Portugal for reforestation of mountainous areas. The physical, chemical, and mechanical properties of *P. nigra* wood were analyzed via robust compositional grouping. The main results allowed us to recognize compositional patterns in different habitats to understand the reforestation process.

Presenter Bio: Pamela Solano is a statistician. She has experience with Ecology, and Environmental and Forest Sciences.

Contact Information: Pamela Solano, CITAB: Centre for the Research and Technology of Agro-Environmental and Biological Sciences, Portugal, Phone: +351 913-966-061, Email: pamela@dme.ufrj.br.

8:30 am to 10:00 am Session 1

Moderator: John Stevens

INVITED TALK: MODEL AVERAGING IN AGRICULTURE AND NATURAL RESOURCES: WHAT IT IS, WHEN IT IS USEFUL, AND WHEN IT IS A DISTRACTION

Philip M. Dixon¹

¹Iowa State University, Ames, IA, USA

Modern statistics provides various answers to the model uncertainty dilemma: how should you make inferences when you do not know which model to use? Model averaging (MA) is appealing because it does not require choosing one model. Most discussions of MA focus on regression models; I will focus on the use of MA with ANOVA models. I will review some of the commonly used, e.g., AIC-weights and Bayesian model averaging, and some less well known, e.g., stacking and bagging, approaches. I apply these to two issues in the analysis of data from factorial experimental designs: the potential value of averaging models with and without interaction terms and the value of “violating hierarchy” in the analysis of Before-After-Control-Impact studies. I show that carefully chosen design matrices produce models that make ecological sense, have interaction terms, and lack one or more main effect terms. I will conclude with general guidance for when MA has been found useful and when I believe it is a distraction.

Presenter Bio: Philip Dixon likes to develop and evaluate statistical methods to answer interesting biological questions. Some of his current projects are developing non-parametric estimates of prediction distributions, modeling physical activity data, and developing model-based visualizations of species composition data.

Contact Information: Philip Dixon, Department of Statistics, Iowa State University, Ames IA, Phone: 515-294-2142, Email: pdixon@iastate.edu.

EXPLORATION OF ENSEMBLE APPROACHES FOR CREATING AUTOMATED MAPS of SOIL MOISTURE AND SNOW WATER EQUIVALENT IN THE STATE OF UTAH

Brennan Bean¹, Logan Schneider¹, Ethan Ancell¹ and Jon Meyer²

¹Department of Mathematics and Statistics, Utah State University, Logan UT

²Department of Plants, Soils, and Climate, Utah State University, Logan UT

Utah's water management planning and drought forecasting rely heavily on accurate characterizations of springtime snowpack and soil moisture conditions. Federal agencies provide daily estimates of both variables at various resolutions, but the accuracy of the model estimates provided by these agencies in a topographically complex place like Utah has not been well explored. This presentation describes two independent modeling efforts to improve upon national-level characterizations of Utah's snowpack and soil moisture content using Utah-specific maps created from land-based weather station networks. Both efforts use ensemble approaches where the "local" maps are combined with the national maps using weights dynamically determined based upon the spatial density of the local station network and the overall bias detected in the national data products. The annual performance (i.e. root mean square error) of the local, federal, and ensemble models are compared for both variables via spatial cross-validation using land-based measurements available at weather stations scattered throughout the state of Utah. The resulting model performance measures highlight the general success, though occasional failure, of the ensemble models to improve upon the best available federal-level model estimates of soil moisture and snow water equivalent in the state of Utah. The presentation concludes with a description of the ongoing efforts to incorporate these ensemble maps into automated data-products hosted by the Utah Climate Center.

Presenter Bio: Dr. Bean is an Assistant Professor of Statistics at Utah State University whose research program focuses on the applications of spatial statistics in civil engineering and natural resources.

Contact Information: Brennan Bean, Utah State University, Logan, UT Phone: 435-797-4130, Email: brennan.bean@usu.edu.

pySeqRNA: AN AUTOMATED PYTHON PACKAGE FOR RNA SEQUENCING DATA ANALYSIS

Naveen Duhan^{1,2} and *Rakesh Kaundal*^{1,2,3}

¹Bioinformatics Facility, Center for Integrated BioSystems, Utah State University, Logan, UT

²Department of Plants, Soils, and Climate, Utah State University, Logan, UT

³Bioinformatics lab, Department of Computer Science, Utah State University, Logan, UT

With the advent of Next-Generation Sequencing (NGS) technologies, numerous data is being generated every day, however, streamlined analysis remains a big hurdle to efficiently use the technology. A large number of algorithms, statistical methods, and software tools have been developed in recent years to perform the individual analysis steps of various NGS applications. Some NGS applications data analysis procedures are therefore very complex, requiring several program tools to be downloaded for their various processing steps. There is a significant room for the development of scalable computing environments that link the individual software components to automated workflows to efficiently and reproducibly conduct complex genome-wide analyses. We have developed a Python package (pySeqRNA) that is capable of running the NGS data analysis from start to finish reproducibly and efficiently. This package provides a uniform workflow interface and support for running on the High-Performance Computing Cluster (HPCC) as well as on local computers. It is an extensible pipeline for performing end-to-end analysis with automated report generation for various NGS applications like RNA-Seq, single-cell RNA-Seq, and dual RNA-Seq, etc. To simplify the analysis of these applications, the package provides pre-configured analysis and report templates. pySeqRNA workflow consists of quality check and pre-processing of raw sequence reads, accurate mapping of millions of short sequencing reads to a reference genome including the identification of splicing events, quantifying expression levels of genes, transcripts, and exons in two ways: (i) Uniquely mapped reads, (ii) Multi-mapped groups, and Differential analysis of gene expression among different biological conditions, biological interpretation of differentially expressed genes, including functional enrichment analysis. This package accelerates the retrieval of reproducible results from NGS experiments. By integrating several command-line tools and custom Python scripts, it allows effective use of existing software and tools with newly written scripts in Python without restricting users to a collection of pre-defined methods and environments. pySeqRNA is freely available at <http://bioinfo.usu.edu/pySeqRNA/>.

Presenter Bio: Naveen Duhan is a PhD student in the Department of Plants, Soils and Climate at Utah State University. He works on developing computational tools and pipelines for biologists.

Contact Information: naveen.duhan@usu.edu.

ENVIRONMENTAL PARAMETERIZATIONS FOR PREDICTIVE AGRICULTURE

Shelly Hunt^{1,2} *Daniela Jones*², **John Gottula**¹,

¹ SAS Institute, Inc, 100 SAS Campus Drive, Cary, NC USA, 27513

² Department of Biological and Agricultural Engineering, North Carolina State University, Raleigh, NC 27695

The single largest source of variation in modern multi-environment agricultural studies is ‘environment.’ Partitioning environmental variation into its components enables researchers to discover causal relationships, deconstruct interaction components and predict future outcomes. While considerable effort has been undertaken to prosecute environmental parameterizations within specific research programs, there are few examples that draw concurrently from the most advanced statistical and machine learning algorithms, or from domain-specific advances like Additive Main Effect and Multiplicative Interaction (AMMI) research (breeding), reference product comparison (biologicals), or crop models (agronomy). Incorporating findings from diverse research contexts, we propose a structure to prepare, engineer and test environmental features in a multi-model framework. Our results demonstrate that flexible feature selection iterated with multi-model comparisons and layered with subject matter expertise can generate environmental parameterizations of reasonable interpretability and with superior predictive accuracy.

Presenter Bio: John Gottula is an AgTech innovator with an applied crop science background and a passion for analytics. John is the co-creator of #AgileAg, a movement to link Agile processes with crop science to drive analytics adoption in the agriculture sector.

Contact Information: John Gottula, SAS Institute, Inc, 100 SAS Campus Drive, Cary, NC USA, 27513, Phone: 806-441-4427, Email: john.gottula@SAS.com.

10:30 am to 11:30 am Panel Session Moderator: Bruce Craig

CHALLENGES AND OPPORTUNITIES OF AGRICULTURAL STATISTICS IN THE BIG DATA ERA

Ann Stapleton¹, James Reecy², Jennifer Clarke³, Juan Steibel⁴, Rakesh Kaundal⁵

¹ National Institute of Food and Agriculture, US Department of Agriculture, Kansas City, MO 64105, USA

² Office of the Vice President for Research, Iowa State University, Ames, IA 50011, USA

³ The Quantitative Life Sciences Initiative, University of Nebraska – Lincoln, Lincoln, NE 68508, USA

⁴ Michigan State University, East Lansing, MI 48824, USA

⁵ Kaundal Artificial Intelligence and Advanced Bioinformatics Lab, Utah State University, Logan, UT 84322, USA

In-Person Poster Presentations

Day 2 - Tuesday, May 17 & Day 3 – Wednesday, May 18, 2022

2:30 pm to 3:30 pm (05/17) & 10:00 am – 10:30 am (05/18)

A SEMINAR COURSE ON SYSTEMATIC REVIEWS IN AGRICULTURE AND THE ENVIRONMENT

Bruce Craig¹, *Sylvie Brouder*¹, *Jeff Volenec*¹, *Chao Cai*¹, *Jane Yacilla*¹, *Scott Murrell*², *Matt Ruark*³, *Carol Barford*⁴

¹Purdue University

²African Plant Nutrition Institute

³University of Wisconsin-Madison

As part of an AFRI NIFA grant titled “FACT: An Innovative Cyber-Framework Integrating Public/Private Data for Evidence-Based Recommendations,” we developed a seminar course for PhD students who want to develop a publishable literature review for their dissertation topic and/or in future work. The course goal is for students to understand the distinguishing features of systematic review (SR) methodology and be able to construct a research question that is answerable by an SR. Course module were taught by instructors from agriculture, information and library sciences, and statistics, and was co-taught simultaneously at two universities. In this poster, we summarize the learning outcomes and course structure, and share what we learned through its challenges and successes.

Presenter Bio: Bruce Craig is a professor of the Department of Statistics at Purdue University.

Contact Information: 250 N. University Street, Purdue University, West Lafayette IN 47907-1267, Phone: (765) 494-6043, Email: bacraig@purdue.edu.

QUANTITATIVE GENETIC ANALYSIS OF BODY FEATURES IN DAIRY CALVES EXTRACTED USING COMPUTER VISION

Ligia C. Moreira¹, Joseane P. Silva¹, Laura L. Hernandez¹, Joao R. R. Dorea¹, and Guilherme J. M. Rosa¹

¹Department of Animal and Dairy Sciences, University of Wisconsin-Madison, Madison, Wisconsin, USA

The present study was undertaken to explore the phenotypic and genetic relationships of body features in dairy calves extracted using computer vision. A total of 211 Holstein calves were raised under commercial conditions at Arlington Agricultural Research Center. Images and body features (body weight, height front, height back, rump, and body length) were collected once a week for seven weeks. The images were captured using an Intel RealSense D435 camera, and the Marker-Controlled Watershed Segmentation was used to extract the animal body and its features on MATLAB. A half-sib families mixed model was fitted for each outcome variable using the lme4 package in R, and the intraclass correlation (ICC) and heritability (h^2) were calculated. The ICC of height back, rump and body length were very small, as well as the heritability of these variables. Body weight had the highest heritability of 0.2454. The Pearson correlation between variables showed that height front and body length had positive small to moderate correlation with predicted transmitting ability for production traits. The results indicated that body features of dairy calves can be efficiently extracted using computer vision and this system can also be automated and implemented on commercial farms, providing an additional tool for selection of replacement heifers.

Presenter Bio: Dr. Moreira is a Research Associate at the University of Wisconsin-Madison with extensive experience in data analysis of farm data and meat sciences.

Contact Information: Ligia C. Moreira, Dr. Rosa's lab, Madison, WI, Phone: 608-960-3787, Email: lmoreira@wisc.edu.

Water quality monitoring of streams in Puerto Rico using GAMM

Cristian Perdomo¹, Raúl E. Macchiavelli², Miguel Vázquez² and Gustavo Martínez²

¹Dept. of Mathematical Sciences, University of Puerto Rico, Mayaguez, Puerto Rico

²Dept. of Agroenvironmental Sciences, University of Puerto Rico, Mayaguez, Puerto Rico

Water quality monitoring is a crucial process to assess the health of an ecosystem. Decisions on natural resource management must be based on current and historical conditions, and the effect of policy changes must be assessed comparing water quality and trends before and after the changes. In this paper we analyze 57 streams in Puerto Rico to study the trends over time of water quality parameters such as pH, water temperature, dissolved oxygen and phosphorus concentration, etc. Data were obtained from US Geological Service records and additional studies conducted in the last 15 years. The data available span from 1958 through 2019, with significant gaps in several years. Generalized additive mixed models were fitted using the mgcv library (R v. 4.0.5). The model incorporated random effects to account for the longitudinal nature of these data and meaningful covariates to study differences between regions and the impact of water treatment plants. Smooth functions to study these effects over time were fitted using P-splines. With this approach we were also able to assess statistically significant changes in trends due to time-specific interventions and predict current trends in water quality parameters.

Presenter Bio: Dr. Macchiavelli has more than 40 years of experience teaching, doing research, and serving as consulting statistician in several countries. He has worked in modelling longitudinal and other correlated data in many areas of Agricultural and Environmental Sciences.

Contact Information: Raúl E. Macchiavelli, College of Agricultural Sciences, University of Puerto Rico, Call Box 9000, Mayaguez, PR 00681-9000, Phone: 787-832-40404 ext 5975, Email: raul.macchiavelli@upr.edu.

TOOLS AND RESOURCES FOR R USERS IN AGRICULTURAL RESEARCH

Julia Piaskowski¹ and William Price¹

¹University of Idaho, Moscow, ID, USA

The language ‘R’ has become quite popular for analysis among agriculture researchers. R is powerful and agile language that is constantly undergoing improvements. As an open-source language, its utility is constantly expanded upon by user-contributed packages and additional developer resources. These include whole software packages like Rstudio, as well as tutorials, and cheatsheets. Keeping up with changes and advancements in this dynamic ecosystem is difficult. We have sought to provide resources for R users in agricultural research. We have established a central website to gather all related information, www.agstats.io that includes workshops, tutorials and short instructional blog posts. We also have built an accompanying “R Universe” repository, <https://idahogstats.r-universe.dev>**Error! Hyperlink reference not valid.** intended to serve as a guide to existing R packages serving different facets of agricultural research. Our goal is to help agricultural researchers fully utilize the array of available R resources to reach their goals.

Presenter Bio: Dr. Julia Piaskowski is a consulting statistician for the Idaho Agricultural Experiment Station at the University of Idaho. She provides guidance on statistical analysis, data management, and workflow.

Contact Information: Julia Piaskowski, Statistical Programs, University of Idaho, Moscow, ID. Email: jpiaskowski@uidaho.edu.

INFORMING MANAGEMENT THROUGH LIMITED DATA: AN AHP-DERIVED HABITAT SUITABILITY MODEL FOR GOSHAWKS IN UTAH'S NATIONAL FORESTS

Sarah Bogen¹ and Marilyn Wright¹

¹Utah State University, Logan, UT USA

Habitat suitability models predict species distributions based on observable environmental covariates. HSMs play a vital role in ecosystem management and often require spatial input data. National forests in Utah designate the goshawk as a Management Indicator Species and have intensively monitored the species for more than a decade. As the US Forest Service shifts its focus from MIS- to ecosystem-based management, fewer resources will be allocated to monitoring for individual species such as the goshawk.

The Analytic Hierarchy Process is often used in the business world to solve complex problems by combining a variety metrics. The AHP has also been employed in some HSM studies to generate covariate weights based on species biology rather than spatial input data. However, few studies apply this method in the context of a large, highly varied study area. Here, we present and evaluate an AHP-derived HSM for goshawks in national forests throughout Utah. While we do not suggest AHP as a replacement for more sophisticated models based on field-collected data, we propose this approach as a simple and user-friendly alternative to inform management of well-studied species in cases where data are limited.

Presenter Bio: Sarah Bogen is pursuing a PhD in applied mathematics at Utah State University and is an NSF Climate Adaptation Science trainee. Her current work focuses on statistical and mechanistic modeling approaches for investigating ecological space use under global change.

Contact Information: Sarah Bogen, Department of Mathematics and Statistics and the Seedscape Ecology Research Group, Utah State University, Logan, UT. Email: sarah.bogen@usu.edu.

MACHINE LEARNING METHODS FOR IDENTIFYING RESISTANT, RESILIENT AND SUSCEPTIBLE SHEEP TO GASTROINTESTINAL NEMATODE INFECTIONS

Luara A. Freitas^{1,2,3}, Rodrigo P. Savegnago⁴, Anderson A. C. Alves³, Ricardo L. D. Costa⁵, Guilherme J. M. Rosa³, Claudia C. P. Paz^{2,6}

¹Supported by the São Paulo Research Foundation (FAPESP) #2020/03575-8, #2018/01540-2 and #2016/14522-7, SP, Brazil

²University of Sao Paulo, SP, Brazil

³University of Wisconsin-Madison, WI, USA

⁴Michigan State University, MI, USA

⁵São Paulo Agency of Agribusiness and Technology, SP, Brazil

⁶Animal Science Institute, SP, Brazil

The aims were to investigate the feasibility of using easy-to-measure phenotypic traits to predict immune response to gastrointestinal nematodes and to compare the classification performance of logistic regression, linear discriminant analysis (LDA), random forest, and artificial neural network. Animals were classified into classes of the immune response: resistant, resilient and susceptible, according to fecal egg count and packed cell volume. The classification methods were fitted using 3,654 records of 1,250 Santa Inês sheep, with age class, month of record, farm, sex, Famacha® score, weight and body condition score as predictors, and the immune response as the response variable. The LDA presented the best performance in predicting susceptible (recall=0.80±0.12) and resistant animals (precision=0.82±0.02). The results suggest that scoring easily measurable traits may provide useful information for management decisions on the farm, contributing to reducing parasitic infection and production costs, reducing expenses with the use of measures that require laboratory resources in the disease monitoring. Animals identified as resistant may be incorporated as selection candidates in breeding programs.

Presenter Bio: Luara Freitas is a PhD student at the University of Sao Paulo developing research on image analysis, genomic selection, and machine learning for solving practical problems in sheep. She is in exchange period at the University of Wisconsin-Madison.

Contact Information: Luara A. Freitas, University of Sao Paulo, Ribeirão Preto, SP, Brazil, Phone: 6086401307, Email: luaraa.freitas@gmail.com.

POWER APPROXIMATIONS FOR GENERALIZED LINEAR MIXED MODELS IN R USING STEEP PRIORS ON VARIANCE COMPONENTS

Sydney Geisler¹ and John R. Stevens¹

¹Utah State University, Logan, UT, USA

Many researchers are interested in power approximations from mixed effects models. The probability distribution method requires an exemplary dataset of a given sample size and holding constant the variance components to approximate power of the test for a specific fixed effect. Currently, the option to hold the variance components constant exists in SAS but it is unavailable in R. We present here an R implementation for power approximation that will allow variance components to be essentially held constant by putting steep priors on those parameters. The results obtained are identical to what SAS would produce. Examples of this will be shown with a Gaussian, binomial, and Poisson distribution of the response variables, although this R implementation is certainly not limited to only these family types.

Presenter Bio: Sydney Geisler is a graduate statistics student at Utah State University in the final semester of her master's degree. She has experience with R programming and creating new functions and R packages.

Contact Information: Sydney Geisler, Logan, UT, Phone: 801-231-8280, Email: sydneykgeisler@gmail.com.

ANALYSIS OF AGGRESSIVE DYADIC INTERACTIONS IN GROUP-HOUSED GROW-FINISH PIGS

Junjie Han, Janice Siegford, Gustavo de los Campos, Robert Tempelman, Cedric Gondro, and Juan Steibel

Michigan State University, East Lansing, MI, USA

This work proposes statistical models for dyadic data that arise from pig aggression studies. Our dataset was collected from 797 growing pigs freshly mixed into 59 social groups, resulting in 10,058 records of dyadic interactions (single sided attacks). The response variable was the time in seconds that each animal spent delivering attacks on another group mate. Three generalized mixed models were fit. All models included the fixed effects of sex and individual weight, and indicators of prior social relationships at nursery and at litter stages. Random effects included: aggression giver, aggression receiver, dyad, and social group. A Bayesian framework was used for parameter estimation and for posterior predictive model checking. An overdispersed Poisson mixed model produced the best fit. Prior litter mate effects were the only significant ($p < 0.05$) fixed effect. The dyad variance explained 11.4% of the total variance and aggression giver explained 13.8 % whereas other random effects explained less than 1% each. Further work is aimed at improving model fit, including modeling zero-inflation and accounting for genetic relationships in the modeling of animal-specific and dyadic effects.

Presenter Bio: Junjie Han is a PhD candidate in Animal Science & Computational Mathematics, Science and Engineering at Michigan State University. He is interested in analyzing swine genomics, phenomics, and interactive behaviors through state-of-the-art algorithms in computer science and statistics.

Contact Information: Junjie Han, Quantitative Genetics Group, Michigan State University, East Lansing, MI 48824, Phone: 517-944-7316, Email:

hanjunji@msu.edu.

HUMAN APPROPRIATION OF NET PRIMARY PRODUCTION ACROSS THE UNITED STATES

Kaeli Mueller¹, Suman Paudel², Chris Lant³ and Gustavo Ovando-Montejo¹

¹⁻⁴Utah State University, UT, USA

Human Appropriation of Net Primary Production (HANPP) measures the environmental impact of harvesting the biosphere. It is frequently used as a socio-ecological metric by measuring the grams of organic carbon that are used by humans. HANPP quantifies the amount of Net Primary Production (NPP) that is appropriated by humans and therefore removed from the energy flow of the ecosystem. When NPP, the energy produced by plants through photosynthesis, is removed from the environment, there is less available energy for other organisms. Humans use huge amounts of HANPP globally. Some studies estimate around 1/4 of global NPP is co-opted every year¹.

HANPP use varies greatly across different regions. Our research examines HANPP harvest across the conterminous US. For our study, we calculated HANPP by adding the amount of NPP used in crop harvest, timber harvest, and livestock grazing. Additionally, we also calculated the HANPP of specific products within each of these categories, for example wheat, corn, and soybeans. HANPP was calculated for each county, state, and Land Resource Region in the US. Our novel dataset helps observe trends and patterns in land use between different regions. We will present a poster that uses spatial and graphical visualizations of HANPP trends across the US.

Presenter Bio: Kaeli Mueller is a graduate student in the MS Geography program in the Environment and Society Department in the Quinney College of Natural Resources at Utah State University.

Contact Information: Kaeli Mueller, Utah State University, Logan, UT, Phone: 720-425-2681, Email: kaeli.mueller@usu.edu.

¹ Imhoff, M. L., Bounoua, L., Ricketts, T., Loucks, C., Harriss, R., & Lawrence, W. T. (2004). Global patterns in human consumption of net primary production. *Letters to Nature*, 429(June), 870–873. <https://doi.org/10.1038/nature02685>. Published

Virtual Poster Presentations

Day 3 - Wednesday, May 18, 2022

3:00 pm to 4:00 pm

Moderator: Bill Price

IMPACT OF EXPERIMENTAL DESIGN ON RESPONSES TO FAT SUPPLEMENTATION IN LACTATING DAIRY COWS

José M. dos Santos Neto and Adam L. Lock

Michigan State University, East Lansing, MI USA

Some authors have recommended excluding change-over designs (COD), such as Latin-Square or crossover designs, from meta-analyses related to animal nutrition. Our objective was to perform a meta-analysis to evaluate whether experimental design affects production responses to fat supplementation in lactating dairy cows. We used peer-reviewed papers that contained a comparison between a non-fat-supplemented diet (CON) with a fat-supplemented diet (FA). We classified experimental designs as COD or continuous designs (CD [completely randomized and randomized block design]). Our database had 113 peer-reviewed papers with 220 observations from COD (CON=117 and FA=103) and 214 from CD (CON=102 and FA=112). We used PROC MIXED of SAS, including the random effect of study and the fixed effects of diet (CON vs. FA), experimental design (COD vs. CD), and the interaction between diet and experimental design. Standard errors were used as weighting factors. We did not observe interactions between diet and experimental design for any variable ($P \geq 0.43$). Our findings indicate no evidence that experimental design affects production responses to fat supplementation in lactating dairy cows.

Presenter Bio: Dr. dos Santos Neto is a Research Associate at Michigan State University. He has experience with animal nutrition.

Contact Information: Adam L. Lock, East Lansing, MI, Phone: 517-802-8124, Email: allock@msu.edu.

RANDOM REGRESSIONS FOR MODELING SEMEN FERTILITY IN HF PUREBRED AND CROSSBRED BULLS USING A BAYESIAN FRAMEWORK

Vrinda Ambike¹, R. Venkataramanan¹, S.M.K. Karthickeyan¹, K.G. Tirumurugaan¹, Kaustubh Bhawe² and M. Swaminathan²

¹Tamil Nadu Veterinary and Animal Sciences University, Chennai, Tamil Nadu, India

²BAIF Research Foundation, Pune, Maharashtra, India

Data on insemination records of Holstein Friesian (HF) purebred (n=45497) and crossbred (n=58497) collected from BAIF Research Foundation were utilized. Conception rate was modeled as a binary trait, using linear repeatability models. Random regression models (RRM) were used to obtain the trajectory of variance components across age of the bulls. Legendre Polynomials up to order of fit of 4 were used for the random effects of additive genetic and permanent environmental effects. 200,000 Gibbs samples were generated with a burn-in of 20,000 and thinning interval of 50 using THRIBBS1F90 program. Heritability estimates were very low (<0.1) in both breeds but peaked at the extreme ages. Heritability and repeatability estimates ranged between 0.038 (8 years) to 0.627 (10 years) and 0.060 to 0.809 respectively in purebreds. Narrower ranges of 0.010 (4 years) to 0.087 (11 years) and 0.532 to 0.832 for heritability and repeatability respectively, were obtained in crossbreds. Effect of permanent environment was low in purebreds compared to the crossbreds. The study using RRM was able to provide genetic parameter estimates of fertility for all ages throughout the productive lifespan of bulls.

Presenter Bio: Dr. Vrinda Ambike is a Master's Student at the Dept. of Animal Genetics and Breeding in Tamil Nadu Veterinary and Animal Sciences University. Animal production modeling and large data handling are her areas of interest.

Contact Information: Vrinda Ambike, Department of Animal Genetics and Breeding, Chennai, Tamil Nadu, Phone: +91 9820063304, Email: vrindaambike@gmail.com.

BAYESIAN ANALYSIS FOR PREDICTING RICE YIELD AT RAINFED LANDS

Abdul Aziz, Komariah, Dwi Priyo Ariyanto, Sumani and Widyatmani Sih Dewi

Department of Soil Science, Faculty of Agriculture, Universitas Sebelas Maret, Surakarta, Central Java, Indonesia

Rice consumption level increases along with the increase in population. Rice is the staple food for most of the Asian population in the world. In Indonesia, rice yield from the rainfed land helps to meet the rice demand. However, the rainfed lands are generally poor in nutrients, frequent in drought, and other socioeconomic problems surroundings. Therefore, rice yield from rainfed is fluctuated and hard to predict. Specific-location input and management, as well as socio-cultural support are important factors to support optimum rice yield at the rainfed lands. In order to support optimal field management, the harvest prediction models as a reference for adaptation during cultivation is important, since agricultural production factors are influenced by soil, environment, and management (Pant et al., 2021). Yield prediction has become one of the most challenging problems in precision agriculture. Efforts to develop a crop-yield-prediction-model program are carried out using various inputs from several indicators to increase the accuracy of the prediction model. In this study, we developed a precision rice yield prediction model based on Bayesian Neural Network (BNN) using multiple data sources, including Specific-location input, soil property, climate observations, socio-cultural support and management. According to Drury et al. (2017), Bayesian approach is suitable for the agricultural sector because it can be represented interdependent causal elements or factors, provide an overview of partial or uncertain information, combine new information, and make new conclusions with new information. The developed BNN model achieved an average coefficient of determination (R^2) of 0.46 with Mean Square Error (MSE) of 0.54. This study provides a robust framework for the precision prediction of crop yield and highlights the need to obtain an understanding concept of the environmental stress effect to agricultural productivity and crop yield estimation.

Presenter Bio: Abdul Aziz is a second-year master's degree student in soil science major of agriculture faculty at Universitas Sebelas Maret, Surakarta, Indonesia. His interests include yield prediction modelling and precision agriculture. He also has expertise in climate forecasting.

Contact Information: Abdul Aziz, Department of Soil Science, Faculty of Agriculture, Universitas Sebelas Maret, Surakarta, Indonesia, Phone: +62-8953-6465-2474, Email: abdul19azizz@student.uns.ac.id.

THE FEASIBILITY OF UTILIZING RANDOM FOREST ALGORITHM TO ASSESS LOSS IN PICTURE-BASED CROP INSURANCE IN INDONESIA

Gerald Bryan¹, **Aurellia Christie**¹, Dian Kusumaningrum¹ and Agung Alfiansyah¹

¹Universitas Prasetiya Mulya, IDN

Agricultural insurance, which is one of the essential risk mitigation tools in Indonesia, still has many shortcomings. Huge costs for the loss assessment process, geographical issues, long duration of claim processing, and accessibility issues in many areas are examples of the shortcomings. As a result, many research are done in order to improve agricultural insurance in Indonesia. This research aims to analyze the feasibility of adopting picture-based crop insurance (PBI) to Indonesia with some modifications. PBI currently uses agricultural experts' evaluation of the crop field images to detect loss. It has reduced loss assessment costs but is still inefficient since it requires more time to do the assessment. To address this issue, this research modifies the current PBI by analyzing the viability of using Random Forest algorithm to detect loss from the extracted RGB data of the crop field images. Random Forest successfully detects loss in the validation dataset with the accuracy above 80%. The expected premium of the modified PBI is then compared to the current crop insurance premium in Indonesia to see whether the modified PBI fits to be adopted in Indonesia.

Presenter Bio: Aurellia Christie is a senior with 2 years of apprenticeship experience in the data science field. During her study, she also has received 2 out of 7 The Society of Actuaries of Indonesia certifications.

Contact Information: Aurellia Christie, Universitas Prasetiya Mulya, Banten, Indonesia, Phone: (+62)83848526269, Email: aurelliachristie77@gmail.com.

MIXED-EFFECT LOGISTIC REGRESSION MODELS FOR VIGOR TEST ANALYSIS OF BIOLOGICAL SEED TREATMENT EFFECT ON ORGANIC CORN HYBRIDS

Amin Shirazi¹

¹Iowa State University, Ames, IA, USA

Assessing seed vigor under severe lab conditions has emerged in recent decades as one of the most effective methods to imitate harsh field conditions. The vigor tests have helped seed producers and consumers predict the seed performance in less than ideal field conditions to minimize product loss. Simple regression models are commonly used for vigor test analysis to determine a single combination of seed and hybrid levels that would perform the best in fields. This study proposes a mixed-effect logistic regression model to assess the main and interaction effects of treated seeds and hybrid levels and compares their odds to identify the most appropriate combination for fields. The results show that the mixed-effect logistic regression is a more suitable model to describe the performance of treated seeds for a vigor test than a simple linear regression model.

Presenter Bio: Amin Shirazi is a Ph.D. candidate in statistics at Iowa State University. He has been a research assistant in the statistical consulting group for more than a year and has considerable experience in providing statistical consulting.

Contact Information: Amin Shirazi, Iowa State University, Ames, IA, Phone: 515-708-6372, Email: ashirazi@iastate.edu.

COMPARISON OF BAYESIAN AND FREQUENTIST APPROACHES OF ANALYSIS OF BALANCED AND UNBALANCED MULTI-ENVIRONMENTAL TRIAL DATA

Siraj Osman Omer

School of Accounting, Finance, Economics, and Decision Sciences, Western Illinois University

Multi-environmental trials (METs) are essential for effective breeding lines selection and cultivar recommendation. Genotype-by-environment interaction (GEI) is widely used in crop variety trials for desirable genotypes and stable yield. This paper outlines Bayesian approach to estimate variance components, heritability, and genetic advance in comparing with traditional approach. Our methodology is based on a real dataset on sorghum (sorghum bicolor) trials for balanced factors and simulated data for unbalanced factors involved in data generation, for covering broad application scenarios to diverse and representative the similarities and differences. In this study, half-normal, uniform, and half-t priors' distribution were used. The three priors' sets provide Bayesian posterior information on genetic variability and predicted means with their statistical features. Half-normal prior distribution was found the best for the balanced dataset, while uniform prior distribution for the unbalanced dataset. Bayesian approach gave higher accuracy compared to frequentist approach and with more diverse statistical information. For the real data applications on crop variety trials, the WinBUGS codes are available for Bayesian analysis to similar situations to ensure accurate results and easy implementation.

Presenter Bio: Dr. Siraj is a Biometrician with more than twelve years of experiences. Siraj's research areas include Bayesian statistics and Biometry in agricultural research. Dr. Siraj has published over 25 research articles in applications of statistical methods.

Contact Information: Siraj Osman Omer, School of Accounting, Finance, Economics, and Decision Sciences, Western Illinois University, Email: soo-mohammed@ywiu.edu., sirajstat@yahoo.com. [Tel:3123832279](tel:3123832279).

Announcing the 2023 Conference

Next year's Conference on Applied Statistics in Agriculture and Natural Resources will take place at Purdue University, West Lafayette, Indiana. Tentative dates are May 15-18, 2023.

More information will follow soon. The local contact is Professor Bruce Craig (bacraig@purdue.edu).



Utah Agricultural Experiment Station
UtahStateUniversity