Conference on Applied Statistics in Agriculture and Natural Resources

May 15-18, 2023 | Purdue University | West Lafayette, IN

Thank You to Our Conference Sponsors

Cargill

CORTEVA™ agriscience

PURDUE UNIVERSITY | College of Agriculture

STATE | UtahStateUniversity

College of Agriculture and Applied Sciences
Utah Agricultural Experiment Station

PURDUE UNIVERSITY | Department of Statistics

Agenda Sketch

All presentations are at the Beck Agricultural Center
4550 U.S. 52 W West Lafayette, IN 47906

Day 1– Monday, May 15th, 2023 (workshop attendees only)

7:30 am to 8:30 am	Registration and morning refreshments
8:30 am to 5:00 pm	**Workshop:** Scalable Bayesian models and estimation methods for the analysis of big spatial and spatio-temporal data by Profs Andrew Finley and Jeff Doser
10:00 am to 10:30 am	Break with snacks
12:00 pm to 1:30 pm	Group lunch
3:00 pm to 3:30 pm	Break with snacks

Day 2– Tuesday, May 16th, 2023

7:30 am to 8:30 am	Registration, morning refreshments, and poster set-up
8:30 am to 9:45 am	**Session 1** Welcome and Opening Remarks **Keynote address:** Tackling large spatial datasets via dimension reduction, induced sparsity, and distributed computing: A case study in forestry applications by Prof Andrew Finley
9:45 am to 10:15 am	Break with snacks, registration, and poster set-up
10:15 am to 11:35 am	**Session 2**
11:35 am to 1:00 pm	Group lunch
1:00 pm to 2:35 pm	**Session 3** **Invited talk:** Geostatistical capture-recapture models by Prof Mevin Hooten
2:35 pm to 3:00 pm	Break with snacks
3:00 pm to 4:35 pm	**Session 4** **Invited talk:** Designing experiments for large multi-environment genomic studies: Mega-environmental designs (MED) optimization for sparse testing by Prof Lucia Gutierrez
6:30 pm to 10:00 pm	**Reception at the Courtyard Lafayette**, 150 Fairington Ave, Lafayette, IN 47905. Appetizers and open bar (bring your drink tickets)
7:30 pm to 9:00 pm	Country line dancing/lessons

Agenda Sketch

All presentations are at the Beck Agricultural Center
4550 U.S. 52 W West Lafayette, IN 47906

Day 3– Wednesday, May 17th, 2023

7:30 am to 8:30 am	Morning refreshments
8:30 am to 10:05 am	**Session 5** **Invited talk:** Addressing reproducibility in cross-validation schemes for heterogeneous field data by Prof Rob Tempelman
10:05 am to 10:30 am	Break with snacks
10:30 am to 12:05 pm	**Session 6** **Invited talk:** The Carnegie Mellon University Cloud Lab: Automating science for the future by Prof Rebecca Doerge
12:05 pm to 1:30 pm	Group lunch
1:30 pm to 2:45 pm	**Session 7** **Invited talk:** Models and methods for network meta-analysis in the plant and agricultural sciences by Prof Larry Madden
2:45 pm to 4:30 pm	Break with snacks and **in-person poster session**

Day 4– Thursday, May 18th, 2023

7:30 am to 8:30 am	Morning refreshments
8:30 am to 10:05 am	**Session 8** **Invited talk:** Data-driven perspectives in a commercial breeding pipeline by Dr. Maria Almeida Macedo
10:05 am to 10:30 am	Break with snacks
10:30 am to 11:45 am	**Session 9**
11:50 am to 1:00 pm	Group Lunch
1:00 pm to 2:00 pm	**Awards and Closing Remarks**

Special thanks to the organizing committee and local arrangements

- Nora Bello—The Ohio State University
- Bruce Craig—Purdue University
- Xin Dai—Utah State University
- Guillermo Marcillo—West Texas A&M University
- Neil Paton—Cargill
- Hans-Peter Piepho—University of Hohenheim, Germany
- Denis Valle—University of Florida
- Hannah Baker—Dept of Statistics, Purdue university
- Lauren M. Lee Johnson—Beck Center Facility & Events Manager
- Amanda Shields—Purdue Conferences
- Ksheera Sagar KN—Dept of Statistics, Purdue University

Day 1 - Monday, May 15th, 2023

8:30 am – 5:00 pm

WORKSHOP: Scalable Bayesian models and estimation methods for the analysis of big spatial and spatio-temporal data

Andrew Finley, Professor
Michigan State University
Jeff Doser, Postdoctoral Research Associate
Michigan State University

Bayesian hierarchical models have been widely deployed for analyzing spatial and spatio-temporal datasets commonly encountered in forestry, ecology, agriculture, and climate sciences. However, with rapid development of remote sensing and environmental monitoring systems, statisticians and data analysts frequently encounter massive spatial and spatio-temporal data that cannot be analyzed using traditional approaches due to their heavy computing demands. In this course, we will present scalable Bayesian models and related estimation methods that provide fast analysis of big spatial and spatio-temporal data using modest computing resources and standard statistical software environments such as R. We will begin with an introduction to the common types of geo-referenced spatial data, then survey software packages for exploratory and subsequent statistical analysis. We will briefly cover exploratory data analysis techniques like variogram fitting, basics of geo-statistical approaches like kriging, and Gaussian Processes. We will then highlight key computational issues experienced by Gaussian Process models when confronted with large datasets. In this context, we will introduce scalable Bayesian models that can deliver fully model-based inference for massive spatial data. This discussion will focus on the Nearest Neighbor Gaussian Process (NNGP) that yields computational gains while providing rich Bayesian inference for analyzing large univariate and multivariate spatial data. We will also present a comparative assessment of other related methods and strategies for large spatial data including low-rank models. We will demonstrate practical implementation of these models using newly developed spNNGP and spOccupancy R packages. All topics will be motivated using real data and participants will be encouraged to follow along with the analyses on their own laptops. Motivating data will come from forestry, agriculture, and wildlife monitoring applications. The workshop will close with a short focused session on occupancy modeling to assess wildlife species distributions while explicitly accounting for measurement errors common in detection-nondetection data. We will not assume any significant previous exposure to spatial or spatio-temporal methods or Bayesian inference, although participants with basic knowledge of these areas will experience a gentler learning curve.

Day 2 - Tuesday, May 16th, 2023

8:30 am to 9:45 pm Session 1 **Moderator: Bruce Craig**

WELCOME AND OPENING REMARKS

Bernie Engel, Professor
Associate Dean and Director of Agricultural Research and Graduate Education; Professor of Agricultural and Biological Engineering and Environmental and Ecological Engineering
Purdue University

KEYNOTE ADDRESS: *TACKLING LARGE SPATIAL DATASETS VIA DIMENSION REDUCTION, INDUCED SPARSITY, AND DISTRIBUTED COMPUTING: A CASE STUDY IN FORESTRY APPLICATIONS*

Andrew 0 Finley Michigan State University, East Lansing, MI, USA

Spatial process models for analyzing geostatistical data entail computations that become prohibitive as the number of spatial locations increases. Developing statistical and computational methods to tackle this challenge of dimensionality is an active area of research. In practice, workable solutions commonly require a combination of complementary modeling and computing tools. This talk highlights a class of highly scalable Nearest Neighbor Gaussian Process (NNGP) models that provide fully model-based inference for large geostatistical datasets. Presentation of the NNGP is motivated by a study that aims to predict forest biomass at the species-level across the continental US. Here, both the number of locations and number of outcomes at each location is large. The proposed hierarchical model addresses these sources of dimensionality via a spatial latent factor model with NNGPs used as sparsity-inducing priors on each factor. Model parameters are estimated using a computationally efficient Markov chain Monte Carlo (MCMC) algorithm, for which scalability is achieved by avoiding storage or decomposition of large matrices and strategic use of shared memory parallelism.

Presenter Bio: Dr. Finley is a Professor at Michigan State University with a joint appointment in the Department of Forestry and Department of Geography, Environment, and Spatial Sciences. His research advances methodologies and software for monitoring and modeling environmental processes, Bayesian statistics, spatial statistics, and statistical computing.

Contact Information: Andrew O Finley, Department of Forestry, East Lansing, MI Phone: 517-432-7219, Email: finleya@msu.edu

10:15 am to 11:35 am Session 2 — Moderator: William Bridges

A NOVEL TRAINING EXPERIENCE ON THE THEORY AND APPLICATION OF BAYESIAN STATISTICS IN AGRICULTURE FOR AFRICAN RESEARCH PROFESSIONALS

***Sarah Manski*[1]**, *Leonard Johnson[1], Dennis Ikpe[1], Innocensia John[2], Jennifer Green[1], Frederi Viens[3]*
[1]Michigan State University, East Lansing, MI, USA
[2]University of Dar-es-Salaam, Tanzania
[3]Rice University, Houston, TX, USA

Bayesian statistics are becoming more widely used in many areas, including agriculture science, due to their advantages over other methodologies. They are better adapted to the small and low-quality datasets typical in agriculture studies. In Ethiopia, we offered a unique 5-day experience to agriculture scientists from across Africa, teaching the advantages, theory, and application of Bayesian statistics in agriculture research. The 5 days covered basics of learning R, the conceptual framework and computational implementation of Bayesian analyses, and the opportunity for participants to begin applying these ideas to their own agronomy problems. We developed an accessible and inclusive R workshop with curated resources aimed at improving researcher interest, confidence, and ability in learning R. Surveys given over the 5 days showed the R workshop increased participant interest and confidence and facilitated learning new techniques to implement Bayesian methods in the training. We used Manski's prior research as an example to expose participants to the realities of applying Bayesian methods in agronomy. This encouraged them to use the knowledge gained during this learning experience in their own work.

Presenter Bio: Sarah Manski is a PhD student at Michigan State University in the Department of Statistics and Probability. Sarah's research interests include applied Bayesian analysis in agriculture, probability theory, and statistics education.
Contact Information: Sarah Manski, C512 Wells Hall, 619 Red Cedar Rd, Lansing, MI 48824
Email: manskisa@msu.edu

MITIGATION OF DROUGHT RISK FOR FARMERS VIA ADOPTION OF MORE COMPLEX CROP ROTATIONS IN THE MIDWEST: A BAYESIAN ANALYSIS

*Sarah Manski[1], **Gina Pizzo[1]**, Yvonne Socolar[2], Ben Goldstein[2] Harley Cross[3], Katie Fettes,[3] Aria McLauchlan[3], Leo Pham[1], Timothy Bowles[2], and Frederi Viens[4]*
[1]Michigan State University, East Lansing, MI, USA
[2]UC Berkeley, Berkeley, CA, USA
[3]Land Core, Grass Valley, CA, USA
[4]Rice University, Houston, TX, USA

We build predictive models to quantify the risk-mitigation value of adopting more complex crop rotations. We perform a full Bayesian analysis in Illinois, Minnesota, and Wisconsin predicting yield using rotational complexity, soil quality, and water availability, and generate posterior predictions at the field level for a variety of weather and farm management scenarios. The implications of our work are in the farm lending and crop insurance industries, as risk reduction afforded by soil management practices is not used in current risk assessments for farmers. We discuss how these posterior probabilities can answer a variety of questions and inform decisions for farmers, lenders, and insurers. For example, in 90% of counties in Illinois, over 90% of fields will experience reduction of risk when using 3-crop rotation as opposed to 2-crop rotation in dry conditions. It is widely acknowledged that adoption of soil health practices, including increased rotation, improves climate resilience, stores carbon in the soil, and reduces economic risk. Using Bayesian methodology and our neighborhood approach to spatial statistics, we can evaluate this reduction in risk to encourage adoption of these practices.

Presenter Bio: Gina Pizzo is a PhD student at Michigan State University in the Department of Statistics and Probability. Gina's research interests include applied statistics and Bayesian analysis in agriculture and in sociology.
Contact Information: Gina Pizzo, Michigan State University, East Lansing, MI, Phone: 248-592-1006,
Email: pizzogin@msu.edu

MAKING BETTER USE OF DATA FROM DIVERSE SOURCES: CHALLENGES AND OPPORTUNITIES

***Linda J. Young*[1]**
[1]USDA National Agricultural Statistics Service, Washington DC, USA

National Statistical Agencies (NSAs), in the U.S. and internationally, have relied on probability-based surveys as the foundation for official statistics for more than 50 years. For most survey programs, list frame coverage and response rates have steadily decreased while costs have risen in recent years. Other data sources, such as administrative, weather, earth observation, and geospatial data, have become increasingly available. Consequently, NSAs have increasingly moved to incorporate these alternative data sources into the estimation processes in a statistically principled way. Because farms are inherently geographically based, all non-survey data identified to date are geospatially referenced. Historically, the farm on the USDA National Agricultural Statistics Service (NASS) list frame are not geospatially referenced, making linkage of survey and non-survey data challenging. The Integrated Modeling and Geospatial Estimation System (IMAGES) project, one of NASS's strategic efforts, is charged with modernizing the application of non-survey data by automating the ingestion, integration, modeling, and analysis of non-survey with survey data to produce the best possible agricultural estimates. In this presentation, the opportunities that the project team has identified and the challenges that have been encountered are discussed. Open questions are highlighted.

Presenter Bio: Dr. Young is Chief Mathematical Statistician and Director of Research and Development. 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

SPATIAL AND TEMPORAL ANALYSIS OF CORN RESPONSE TO NITROGEN AND SEED RATES ON-FARM PRECISION EXPERIMENTS

***Carlos A. Alesso*[1]**, *and Nicolas F. Martin[2]*
[1]ICiAgro Litoral, UNL-CONICET, Esperanza, Argentina
[2]Department of Crop Sciences, University of Ilinois, Urbana, IL, USA

In this study, we modeled the within-field variability of corn response to nitrogen (NR) and seed (SR) rates using on-farm precision experiments (OFPE) and geographically weighted regression (GWR). We assessed the spatial agreement and temporal stability of these responses and determined the effect of weather and site-specific features on their spatial distribution. Response maps were estimated by fitting GWR models to yields and as-applied NR and SR from 14 OFPE. A random forest model was trained on a dataset with responses classified as positive (P) and non-positive (NP) spatially joined to weather, soil, and landscape covariates. Responses to NR and SR were detected in all fields, and they were moderately consistent across years due to weather. Spatial agreement between NR and SR response classes for the same year suggests shared controlling factors, with weather variables being the most important predictors followed by landscape and soil attributes. Our findings highlight the importance of weather when modeling crop response for management zone delineation. Overall, our study contributes to the understanding of within-field variability of crop response to inform precision agriculture practices.

Presenter Bio: Dr. Alesso is an Assist. Professor (UNL), with more than 5 years of experience in teaching statistics. He has extensive experience applying spatial statistical methods to spatial data generated in agricultural systems
Contact Information: Dr. Agustín Alesso, ICiAgro Litoral, UNL-CONICET, Esperanza, Argentina, Phone: +54-3496-426400, Email: calesso@fca.unl.edu.ar

1:00 pm to 2:35 pm Session 3 — Moderator: Raul Macchiavelli

INVITED TALK: GEOSTATISTICAL CAPTURE-RECAPTURE MODELS

Mevin B. Hooten[1], *Michael R. Schwob*[1], *Devin S. Johnson*[2] *and Jacob S. Ivan*[3]
[1]The University of Texas at Austin, Austin, TX, USA
[2]NOAA National Marine Fisheries Service, Honolulu, HI, USA
[3]Colorado Parks and Wildlife, Fort Collins, CO, USA

Methods for population estimation and inference have evolved over the past decade to allow for the explicit incorporation of spatial information when using capture-recapture study designs. Traditional approaches to specifying spatial capture-recapture (SCR) models often rely on an individual-based detection function that decays as an individual's activity center is farther from a detection location. Traditional SCR models are intuitive because they incorporate mechanisms of animal space use based on their assumptions about activity centers. We generalize SCR models so that they can accommodate a wide range of space use patterns, including for those individuals that may exhibit elliptical utilization distributions. Our approach uses underlying Gaussian processes to characterize the space use of individuals. This allows us to account for multimodal space use patterns as well as nonlinear corridors and barriers to movement. We refer to this class of models as geostatistical capture-recapture (GCR) models. We adapt a recursive computing strategy to fit GCR models to data in stages, some of which can be parallelized. This technique facilitates implementation and leverages modern multicore and distributed computing environments. We demonstrate the application of GCR models by analyzing both simulated data and a data set involving capture histories of snowshoe hares in central Colorado, USA.

Presenter Bio: Dr. Hooten is a Professor of Statistics and Data Sciences at UT-Austin and elected Fellow of the American Statistical Association. His research is in spatial and spatio-temporal ecological and environmental statistics.
Contact Information: Mevin Hooten, The University of Texas at Austin, Austin, TX, Phone: 512-232-0693, Email: mevin.hooten@austin.utexas.edu

CAPTURE-RECAPTURE ESTIMATION FOR THE CENSUS OF AGRICULTURE

Habtamu K Benecha[1], *Luca Sartore*[1,2], *Grace Yoon*[1], *Bruce A. Craig*[1,3], *Denise A Abreu*[1], *Linda J Young*[1]
[1]USDA National Agricultural Statistics Service, Washington, DC, USA
[2]National Institute of Statistical Sciences, Washington, DC, USA
[3]Purdue University, West Lafayette, IN, USA

USDA's National Agricultural Statistics Service (NASS) conducts the Census of Agriculture every five years. The Census is the only source of uniform, comprehensive agricultural information on the characteristics of United States (U.S.) farms and the people who operate them for every state and county in the U.S. The Census list frame, known as the Census Mail List (CML), is incomplete. To quantify the incompleteness in the CML, NASS uses the June Area Survey (JAS), which is based on an area frame. Census weights are adjusted based on a capture-recapture method that accounts for undercoverage, nonresponse, and misclassification. Historically, only JAS data and CML records that are linked to the JAS have been used to estimate these adjustment weights. This paper proposes an alternative capture-recapture approach that utilizes all the available JAS and Census information for the estimation of adjusted Census weights. In addition, the uncertainty measures associated with the Census estimates are computed using the delete-a-group jackknife and adjusted for calibration. Results from simulation studies and an application of the method to data from the 2017 Census of Agriculture are presented.

Presenter Bio: Dr. Benecha is a Mathematical Statistician at the Research and Development Division, National Agricultural Statistics Service, USDA. Part of his work focuses on model-based estimation from agricultural surveys and the Census of Agriculture.

Contact Information: Habtamu Benecha, USDA National Agricultural Statistics Service, Washington, DC. Phone: 202-375-3871, Email: Habtamu.benecha@usda.gov

UNDERSTANDING HABITAT SELECTION AND RESOURCE USE WITH THE TIME MODEL AND THE TIME-EXPLICIT STEP SELECTION FUNCTION.

Denis Valle[1], *Nina Attias*[1,4], *Joshua A. Cullen*[2], *Aline Giroux*[3], *Luiz Gustavo R. Oliveira-Santos*[3], *Arnaud L. J. Desbiez*[4,5,6], *Robert J. Fletcher Jr.*[1]
[1]University of Florida, Gainesville, FL, USA
[2] Florida State University, Tallahassee, FL, USA
[3] Federal University of Mato Grosso do Sul, Campo Grande, MS, Brazil
[4] Instituto de Conservação de Animais Silvestres, Campo Grande, MS, Brazil
[5] Royal Zoological Society of Scotland, Murrayfield, Edinburgh, UK
[6] Instituto de Pesquisas Ecologicas, Nazare Paulista, SP, Brazil

Understanding the relationship between animals and their environment is a central focus of ecological science, with implications regarding how animals will be affected by environmental change and how these impacts can be mitigated. We describe how complementary information can be obtained by separately assessing the drivers of time to traverse the landscape and the drivers of selection strength. We specify a new model which enables inference on the time taken to traverse landscapes and we show how it can be combined with a selection function to form a time-explicit step selection function (tSSF) model. These models are illustrated using GPS-tracking data from giant anteaters in the Pantanal wetlands of Brazil. The time model revealed that the fastest movements tended to occur between 8 pm and 5 am and that giant anteaters moved faster over wetlands while moving much slower over forests and savannas. We found that wetlands were consistently avoided whereas forest and savannas tended to have higher preference. Finally, we found that selection for forest increased with temperature, suggesting that forests may act as thermal shelters when temperatures are high.

Presenter Bio: Dr. Valle is an associate professor and his research is at the interface of Bayesian statistics, data science, and ecology. He is interested in developing and applying new models to remotely sensed data (e.g., GPS tracking data and LiDAR).
Contact Information: Denis Valle, Univ. of Florida, Gainesville, FL, Phone: 352-392-3806, Email: drvalle@ufl.edu

SPATIAL GENERALIZED LINEAR MIXED MODELS FOR BIG DATA USING REDUCED RANK MODELS WITH A LAPLACE APPROXIMATION

Eryn Blagg[1], *Jay Ver Hoef*[2], *Philip Dixon*[1], *Paul Conn*[2]
[1]Iowa State University, Ames, IA, USA
[2]National Marine Mammal Laboratory, Seattle, WA

In spatial statistics large spatial fields often have issues with computational times. When dealing with estimates of spatial parameters and covariates, computations of order O(n^3) are need. In this presentation, we define a new way to run computationally large spatial fields using the Hierarchical Generalized Linear Mixed model, with the implementation of the Sherman Morrison Woodbury theorem and reduced rank methods through basis functions. We describe an optimized algorithm for estimation of the random effects and define how Sherman Morrison Woodbury can be implemented to speed up the computation. We then compare speed and accuracy of the estimation of unknown parameters compared to classical Laplace approximation methods. We then apply the methods to a seal data set to predict the number of Steller sea lions in Bering Sea.

Presenter Bio: Eryn Blagg is a PhD candidate at Iowa State University working under the direction of Philip Dixon. She received her Master's in statistics from Iowa State in 2020 and her Bachelor's in Mathematics and Art from Lawrence University in 2018.

Contact Information: Eryn Blagg, Iowa State University, Ames, IA Email: eblagg@iastate.edu

3:00 pm to 4:35 pm Session 4 — Moderator: Hans-Peter Piepho

INVITED TALK: DESIGNING EXPERIMENTS FOR LARGE MULTI-ENVIRONMENT GENOMIC STUDIES: MEGA-ENVIRONMENTAL DESIGNS (MED) OPTIMIZATION FOR SPARSE TESTING

Lucia Gutierrez[1]
[1]University of Wisconsin-Madison, Madison, WI, USA

Optimizing field testing resources is a key component of agricultural evaluations, especially for large experiments such as multi-environment trials (METs). Testing efficiency can be improved by controlling micro- (spatial) and macro-environmental (genotype by environment, GEI) variability. Both, spatial variation, and GEI are common in agricultural field experiments and can largely influence the accuracy and efficiency of large genomic experiments. How to allocate resources within and among environments accounting for both spatial variability and GEI is still not entirely clear. Furthermore, because mixed models can be used to borrow information from relatives or across environments, sparse experimental designs could be superior to balanced designs. The goal of this study was to compare strategies for micro and macro-environmental variability control that include spatial and GEI modeling in a sparse-testing approach to optimize resource allocation in METs. optimizing the mega-environmental sparse testing designs (MEDs) in terms of replication within and across environments and the robustness of MEDs to different covariance structures among environments.

Presenter Bio: Dr. Gutierrez is an Associate Professor in Quantitative Genetics. She has extensive experience with quantitative and statistical genetics applied to plant breeding in cereals.
Contact Information: Lucia Gutierrez, Department of Agronomy, University of Wisconsin-Madison, 1575 Linden Dr., Madison, WI 53706, Phone: 608-890-4348, Email: gutierrezcha@wisc.edu

GENERALIZED FOLDOVER DESIGNS: A CLASS OF DESIGNS FOR IRREGULAR FRACTIONAL 3^F FACTORIAL EXPERIMENTS WITH BIAS PROTECTION PROPERTIES

Vinny Paris[1], *and Max D. Morris*[1]
[1]Iowa State University, Ames, IA, USA

Fractional factorial experiments are often used when the total number of runs in the full factorial experiment is too large. This introduces bias if higher order interaction effects are unaccounted for in the model. The original foldover method proposed by Box and Wilson (1951) allows for complete dealiasing of main effects from unaccounted for second order interactions in fractional 2^f factorial experiments. A new class of designs for irregular fractional 3^f experiments based on a generalization of the foldover method will be presented. By restricting the design space, this new class of designs can dealias main effects from a subset of unaccounted for second order interactions. Simulations will be presented as well as a brief comparison of the currently tabulated *D*-optimal generalized foldover designs against *D*-optimal designs from an unrestricted design space.

Cancelled

Presenter Bio: Vinny is a statistics Ph.D. student at Iowa State University with research in experimental design with interests in bias protection. Fun Trivia: he has been to 17 national parks.
Contact Information: Vinny Paris, Iowa State University, Ames, IA, Email: vinny@iastate.edu

HYPERSPECTRAL PROFILE OF GLYPHOSATE INJURY IN COMMON LAMBSQUARTERS (*Chenopodium album L.*)

Mario Soto[1], *Aurelie Poncet*[1], *Wesley France*[1], *Nilda Burgos*[1], *Juan Velasquez*[1]
[1]University of Arkansas, Fayetteville, AR, USA

Proper understanding of weed response to herbicide application is critical to the development of effective weed control strategies. Different herbicide formulations and application rates may be tested on different species or genotypes in the greenhouse, or in the field. Non-treated controls are included into the experimental design and visual ratings of injury are collected regularly after herbicide application. However, the accuracy and precision of the collected data may vary between raters. The objective of this study was to determine if proximal sensing can be used to provide insights into weed response to herbicide application. An experiment was conducted with Common lambsquarters (Chenopodium album L.), a broadleaf weed of economic importance in Arkansas. Twelve herbicide treatments, consisting of glyphosate with different adjuvants, were applied to 12 cm tall plants in the greenhouse. Hyperspectral measurements of the weed spectral signature were collected -1, 1, 2, 3, 5, 7, 10, 14, 20, and 25 days after treatment application using a spectroradiometer. Different statistical methods were considered to characterize the hyperspectral profile of glyphosate injury in Common lambsquarters.

Presenter Bio: Mario Soto is a visiting scholar at the University of Arkansas. He has experience working with statistical programming and big data analysis in the context of precision agriculture. His current research emphasizes plant stress detection using hyperspectral sensing.
Contact Information: Aurelie Poncet, University of Arkansas, Fayetteville, AR, Phone: 479-575-3979,
Email: poncet@uark.edu

A BAYESIAN APPROACH FOR ESTIMATING AND CHECKING BLOCK DESIGNS IN AGRICULTURAL EXPERIMENTS

Josefina Lacasa*1,2, Ignacio Ciampitti1 and Trevor Hefley2*
1Kansas State University, Dept. of Agronomy, Manhattan, KS, USA
2Kansas State University, Dept. of Statistics, Manhattan, KS, USA

Spatial blocking is a common technique used in designed agricultural experiments. Spatial blocking, however, requires two important assumptions: 1) the spatial blocks can be clearly delineated within a field by an expert; and 2) the area within each block is homogeneous. Statistical analyses of blocked agricultural experiments often show spatially correlated residuals, suggesting one of the assumptions above was not met. We propose a hybrid Bayesian-Machine Learning approach that can account for spatially-correlated residuals in mis-specified block designs. We embed a regression tree within commonly used statistical models for experimental data, which delineates the block boundaries. We illustrate possible applications of this approach with some of the most typical scenarios we have encountered in our applied experience in agricultural research with four synthetic data sets and two field data sets.

Presenter Bio: J. Lacasa is a PhD student in Agronomy and holds an MSc in Statistics from Kansas State University. She is interested in integrating Bayesian statistics to agronomic models.
Contact Information: Josefina Lacasa, Kansas State University, Manhattan, KS, Phone: 785-317-1121,
Email: lacasa@ksu.edu

Day 3 - Wednesday, May 17th, 2023

8:30 am to 10:05 am Session 5 — Moderator: Nora Bello

INVITED TALK: ADDRESSING REPRODUCIBILITY IN CROSS-VALIDATION SCHEMES FOR HETEROGENEOUS FIELD DATA

Robert J. Tempelman[1], *Mohamed Abouhawwash*[1], *and Asiye Yilmaz Adkinson*[1,2]
[1]Michigan State University, East Lansing, MI, USA
[2]Erciyes University, Kayseri, Türkiye

Statistical consultants are being increasingly asked by researchers to assist them with machine learning (ML) techniques for improving predictions. Unfortunately, many of these tools do not readily accommodate data structures that are typical in agriculture and natural resources research, such as spatial, temporal, or genetic correlations as well as blocking and/or different sizes of experimental units. Furthermore, data splits into training and test sets for cross-validation (CV) are often purely random, thereby often leading to overoptimistic assessments of ML tools, particularly as they pertain to predictions in environments not represented in training sets. We assess the implications of CV scheme choices for assessing predictive performance of various ML tools, drawing examples from our own work and from others in animal science. One central theme is that although generalized linear mixed models are not generally considered to be ML tools, their utility along with generalized additive model extensions make them an attractive option to predict responses across environments particularly when, for example, genetic or spatial relationships can be specified between training and test sets.

Presenter Bio: Dr. Tempelman is a Professor of Animal Science and Director of the MSU Agricultural and Natural Resources College Statistical Consulting Center. His research program has been primarily focused on quantitative genetics.
Contact Information: Robert J. Tempelman, Department of Animal Science, East Lansing, MI, Phone: 517-355-8445, Email: tempelma@msu.edu

"BORING" FOR INSIGHTS IN A 28-YEAR SUGARCANE INSECT HERBIVORY DATASET

Quentin D. Read[1], *Hannah J. Pen*[2]
[1]USDA Agricultural Research Service, Southeast Area, Raleigh, NC, USA
[2]USDA Agricultural Research Service, Sugarcane Research Unit, Houma, LA, USA

A sugarcane stalk is a living record of insect herbivory. As the stalk matures, internodes are added at the top and lower internodes harden, allowing the history of damage from herbivores like the sugarcane borer (SCB) to be read chronologically. We used a unique long-term dataset from 28 years of sugarcane yield reduction trials from Louisiana to investigate patterns of SCB damage within stalks, between varieties, and over time. We used a Bayesian binomial GLMM to show that prior conspecific damage increases the probability of future damage on the same stalk. We also found that SCB damage has been steadily decreasing over the past 28 years in Louisiana, and that SCB damage varied over sixfold from the most to least resistant variety. To determine whether variety-level traits explained variation in SCB damage, we performed projection predictive variable selection. Several traits emerged as predictors of SCB damage. In this study, we "bored" into a long-term dataset and derived insights that can be used to investigate mechanisms of SCB resistance and potentially inform sugarcane breeding programs.

Presenter Bio: Dr.Read has been the USDA-ARS Southeast Area statistician since 2021. He has experience as an interdisciplinary researcher in ecology, environmental science, and economics, and as a data science & stats consultant and teacher.
Contact Information: Quentin D. Read, USDA Agricultural Research Service, Southeast Area, Raleigh, NC.
Website: https://quentinread.com Email: quentin.read@usda.gov

MONITORING CROP GROWTH WITH A HIERACHICAL NONLINEAR MODEL WITH MODEL DISCREPANCY

Spencer G. Wadsworth[1], *Jarad Niemi*[1]
[1]Iowa State University, Ames, IA, USA

A combination of antenna technology from the European Space Agency's Soil Moisture and Ocean Salinity (SMOS) satellite, raw data from SMOS, and data processing algorithms make possible the estimation of vegetation water mass in land crops. It has been shown that over the course of a growing season the water mass estimate data closely resembles that of crop growth. The vegetation water mass over a season can and has been modeled using a nonlinear curve, but such models may fail to capture all existing patterns. They may also produce too narrow predictive intervals for future forecasting. In this presentation, we model the vegetation water mass over several seasons with a nonlinear curve in a hierarchical framework with the addition of a hierarchical model discrepancy element added for capturing systematic deviation and improving model forecasts.

Presenter Bio: Spencer is a PhD student and statistics consultant at Iowa State University. His research focuses on forecast modeling of disease outbreaks.
Contact Information: Spencer Wadsworth, Iowa State University, Ames, IA, Email: sgw96@iastate.edu

SIMULATION COMPARISON OF STATISTICAL METHODS USED IN ASSESSING VACCINE EFFICACY FOR UNBALANCED BLOCKS IN VETERINARY STUDIES

Hui Wu[1], *and Kenneth Wakeland*[2]
[1]Kansas State University, Manhattan, KS, USA
[2]Boehringer Ingelheim Animal Health USA Inc., Duluth, GA, USA

In veterinary clinical studies are designed to assess a vaccine's ability to prevent clinical diseases, which support the licensure. The prevented fraction (PF) is the primary conclusion criterion for a binary outcome. Designs of these studies often consider factors (e.g., housing, or litter) to reduce bias. The ideal study design utilizes a 1:1 allocation rule for randomizing animals to the treatment and control groups across blocks. However, in real-world studies, it is not always possible to achieve balance in each block. Typical methods for analyzing these data include Cochran-Mantel-Haenszel (CMH) and generalized linear mixed model (GLMM). Here, we investigated two methods together with another three methods: Generalized Estimating Equations (GEEs); Bayesian GLMM (BGLMM) with logit and log links. Performance characteristics of these Frequentist and Bayesian approaches were compared under both balanced and unbalanced designs via simulation. Overall, both CMH and BGLMM with a logit link performed well in most scenarios. In certain configurations, CMH did not construct a confidence interval. Coverage could be a concern when true PF is zero for the BGLMM. In conclusion, CMH is the fastest method to apply and performs well. For some cases, such as two or more blocking factors, BGLMM could be an alternative method.

Presenter Bio: Hui Wu holds a MS degree in animal science and is currently a PhD student in statistics at Kansas State University. She strives to apply her multidisciplinary training in support of veterinary clinical studies.
Contact Information: Hui Wu, 108D Dickens Hall, 1116 Mid-Campus Drive N., Manhattan KS, Phone: 785-770-5718, Email: huiwu@ksu.edu

10:30 am to 12:05 pm Session 6 — Moderator: Neil Paton

INVITED TALK: THE CARNEGIE MELLON UNIVERSITY CLOUD LAB: AUTOMATING SCIENCE FOR THE FUTURE

Rebecca Doerge[1]

[1]Carnegie Mellon University, Pittsburgh, PA 15213

The future of science lies at the interface of automation, artificial intelligence, machine learning, data science, the foundational sciences and human ingenuity. Carnegie Mellon University is making this future a reality by merging its strengths in these areas and creating an environment where science is accessible, reliable and limited only by ideas. One way they are making this future a reality is through the construction of the Carnegie Mellon University Cloud Lab, the world's first lab of its kind at a university. Built on the software architecture developed by the CMU alumni-founded Emerald Cloud Lab, the remote-controlled CMU Cloud Lab will provide a universal platform for artificial intelligence-enabled experimentation and revolutionize how academic laboratory research and education are done. The CMU Cloud Lab will make scientific experimentation faster, more transparent, less prone to error and more reproducible. Research will no longer be held back by time, space or access to equipment. The CMU Cloud Lab will democratize the discovery process, making technology available to a community of researchers with diverse experiences and backgrounds, both at the university and in the local community.

Presenter Bio: Rebecca Doerge is the Glen de Vries Dean of the Mellon College of Science at Carnegie Mellon University and a member of the Dietrich College of Humanities and Social Sciences' Department of Statistics and Data Science and the Mellon College of Science's Department of Biological Sciences. Her research program focuses on statistical bioinformatics, an interdisciplinary component of bioinformatics that brings together many scientific disciplines for the purpose of asking, answering, and disseminating biologically interesting information in the quest to understand the ultimate function of DNA and epigenomic associations.
Contact Information: Rebecca Doerge, Carnegie Mellon University, Pittsburgh, PA, Phone 412-268-8156,
Email: rwdoerge@andrew.cmu.edu

EFFECT OF FLIGHT ALTITUDE, CLOUD COVER, SUN ANGLE, AND FRONT AND SIDE OVERLAP ON sUAS MULTISPECTRAL IMAGE ACCURACY

Grant Rothrock[1]*, Aurelie Poncet*[1]*, Wesley France*[1]*, Mario Soto*[1]

[1]University of Arkansas, Fayetteville, AR, USA

Small unmanned aerial systems (sUAS) equipped with multispectral cameras are being increasingly used in the agricultural and environmental sciences to help with scouting and monitoring of spatial variability. Multiple overhead images are collected along a flight path and stitched into a single raster file. The data collection parameters and steps taken during stitching may affect the collected data accuracy. The objectives of this study were to identify the drivers of sUAS multispectral data accuracy and establish best practices for data collection and processing. Imagery was collected above a 0.5 ha study area using a sUAS with RTK accuracy. The following data collection parameters were varied between flights: cloud cover, flight altitude, sun angle, front and side overhead image overlap, and radiometric calibration method. Ground-control points, the achieved ground-sampling distance, and a 6-point greyscale were used to quantify the geospatial, geometric, and radiometric stitched data accuracy. Statistical analysis was computed to compare findings between flights. The density of the point cloud created during stitching was also considered as a predictor of data accuracy in the statistical analysis.

Presenter Bio: Grant Rothrock is a program technician at the University of Arkansas. He has experience collecting data and applying his GIS skills to precision agriculture and wildfire ecology. He is seeking graduate opportunities in environmental or agricultural data science.
Contact Information: Aurelie Poncet, University of Arkansas, Fayetteville, AR, Phone: 479-575-3979,
Email: poncet@uark.edu

CHOOSING BETWEEN PROBABILISTIC MODELS AND MACHINE LEARNING ALGORITHMS TO PREDICT BIOMASS AND LAI FOR MULTISPECTRAL-LIDAR UAV-DATA

Pamela Solano[1,2]*, Luís Pádua*[1,2] *and Domingos Manuel Mendes Lopes*[1,2]

[1]Centre for the Research and Technology of Agro-Environmental and Biological Sciences, University of Trás-os-Montes e Alto Douro, Portugal

[2]Institute for Innovation, Capacity Building and Sustainability of Agri-Food Production, University of Trás-os-Montes e Alto Douro, Portugal

Remote sensing technology has been undergoing development to collect and process simultaneous spectral and 3D geometric information. In forestry, Leaf Area Index and biomass predictions are key indicators to make decisions. However, some limitations have been highlighted in the modeling process, such as high heterogeneity and multicollinearity in the multispectral-LiDAR UAV-based Data. To increase the Leaf Area Index and biomass prediction capability, probabilistic models and robust machine learning architectures have been implemented to suggest a variable selection method suitable for multispectral-LiDAR UAV-based Data, avoiding overfitting, and reducing bias and excess variability. Thirty-four models were applied to two datasets. The first maps the LAI of Castanea sativa Miller Using UAV-Based Multispectral and Geometrical Data. The best model achieved a coefficient of determination (R^2) value of 85%. The Bayesian LASSO results achieved the highest R^2 value 93% with 0.3273 MAE, and an RMSE of 0.3960. These findings reinforce the efficacy of using objective criteria to achieve the best performance in the probabilistic models and robust machine learning algorithms.

Presenter Bio: Pamela Solano is a research statistician with 16 years of experience. She has been working in Ecology, Environmental and Forest Sciences since 2018.
Contact Information: Pamela M. Chiroque-Solano
University of Trás-os-Montes e Alto Douro, 5000-801 Vila Real, Portugal, Phone: +49 15156108261,
Email: pchiroque@ gmail.com

THE USE OF GENERATIVE ADVERSARIAL NETWORKS ON REAL-TIME COMPUTER VISION SYSTEM FOR IMAGE CLASSIFICATION OF BEEF CATTLE IN FEEDLOT

Joseane Padilha[1]*, Arthur F. A. Fernandes*[2]*, Joao R. R. Dorea*[1]*, Tiago S. Acedo*[3]*, Guilherme J. M. Rosa*[1]

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

[2]*Cobb-Vantress Inc.*

[3]*DSM Nutritional Products Brazil*

Our group is developing a 3D image acquisition system to monitor the daily growth of cattle in feedlots. The system involves capturing daily images of each animal in the pen, which are identified through electronic ear tags. A data processing pipeline is used to classify the images as either usable or not, discarding useless images and predicting individual animal body weights using the usable ones. However, image classification requires significant effort to minimize errors, especially given harsh outdoor conditions, unrestrained animals, and varying lighting that make it difficult to obtain good images manually, even in large data sets. To address these challenges, we propose using Generative Adversarial Networks (GANs) to generate fake images and augment our training data set. We generated approximately 30,000 fake images, including images depicting various positions of the animals, to increase the diversity of our training data set. Our results showed a significant decrease in Type 1 error (from 0.09 to 0.03), i.e., classifying an image as usable when it is not, and an increase in overall accuracy (from 0.92 to 0.98) compared to a classifier using only original images. Our approach using GANs has the potential to improve the accuracy of daily monitoring of cattle growth in feedlots, enabling better management practices and improved animal welfare.

Presenter Bio: Dr. Padilha is a PhD student at the Department of Animal and Dairy Sciences at the University of Wisconsin-Madison, and a researcher at the Brazilian Agricultural Research Corporation (Embrapa) in Brasilia, Brazil.
Contact Information: Department of Animal and Dairy Sciences, University of Wisconsin-Madison, Madison, WI 53706. Email: grosa@wisc.edu

1:30 pm to 2:45 pm Session 7 — Moderator: Denis Valle

INVITED TALK: MODELS AND METHODS FOR NETWORK META-ANALYSIS IN THE PLANT AND AGRICULTURAL SCIENCES

Laurence V. Madden[1]
[1]The Ohio State University, Wooster, OH, USA

Meta-analysis, the methodology for analyzing the results from multiple studies, has grown tremendously in popularity over the last 45 years. Although most meta-analyses involve a single effect size from each study (e.g., a mean difference for two treatments or a log ratio), there are often multiple treatments of interest across the network of studies. Multi-treatment or so-called network meta-analysis (NMA) can be used for simultaneously analyzing the results from all the treatments simultaneously. NMA can be based on contrasts with a baseline treatment from each study or directly on means (or transformed means) of treatment arms from each study, with the estimation of contrasts performed after the model fit. The contrast-based approach is more popular, especially in medical research, but results are very similar for contrast- and arm-based methods, and equivalent under some circumstances, if the appropriate mixed model is chosen. Arm-based NMA is much easier to perform with standard mixed-model software. Both frequentist and Bayesian methods are popular for model fitting. The most extensive use of NMA in agriculture has been in the estimation of the effects of chemical treatments (i.e., fungicides) in controlling crop diseases. This presentation will cover issues related to the use of mixed models for arm-based NMA of over 300 disease-control experiments for Fusarium head blight of wheat. An emphasis will be placed on assessment of the stability of treatment effects over two decades.

Presenter Bio: Dr. Madden is an emeritus professor of plant pathology at Ohio State. As a plant epidemiologist, he has spent his career studying and predicting plant disease epidemics using a wide range of statistical methods. For the past 15 years he has focused on the application of mixed models for meta-analysis of disease management (and other) studies.
Contact information: Laurence V. Madden, Dept. of Plant Pathology, The Ohio State University, Wooster, OH.
Phone: 330-749-3806.Email: madden.1@osu.edu

COMPARING AGGREGATE AND INDIVIDUAL META-ANALYSIS ON RESPONSES TO FAT SUPPLEMENTATION IN LACTATING DAIRY COWS

José M. dos Santos Neto*, Robert J. Tempelman, and Adam L. Lock*
[1] Michigan State University, East Lansing, MI USA

Meta-analysis (**MA**) methods traditionally aggregate treatment means from the literature; however, raw data could also be used if available. Our objective was to compare aggregate versus individual MA on responses to fat supplementation in lactating dairy cows. Our database used 8 studies comparing a non-fat-supplemented diet (CON) with a fat-supplemented diet (FA). We performed an aggregate MA using the treatment means from the literature (CON=10, FA=15) and an individual MA using individual cow data (CON=334, FA=436). In both MA, we modeled the fixed effects of diet. The aggregate MA included the random effect of study, such that SEM were used as weights. The individual MA included the random effects of study, cow, and the interaction between study and diet. Both MA lead to similar conclusions: no evidence that FA had effects on dry matter intake, milk yield, or milk protein ($P \geq 0.18$), but that FA increased milk fat ($P < 0.03$). Our findings suggest no difference between aggregate and individual MA to evaluate fat supplementation in dairy cows, provided that study by treatment effects are specified as random to broadly define study as the experimental unit for treatments.

Presenter Bio: Dr. dos Santos Neto is a Research Associate at Michigan State University. He has experience with animal nutrition and statistical analysis.
Contact Information:
Adam L. Lock, Robert T. Tempelman, José M. dos Santos Neto, East Lansing, MI, Phone: 517-355-8445
Email: allock@msu.edu, tempelma@msu.edu, dossan12@msu.edu

A REML METHOD FOR THE EVIDENCE-SPLITTING MODEL IN NETWORK ANALYSIS

Hans-Peter Piepho[1]*, Johannes Forkman*[2]*, Waqas Malik*[1]
[1]University of Hohenheim, Stuttgart, Germany
[2]Swedish University of Agricultural Sciences, Uppsala, Sweden

Checking for possible inconsistency between direct and indirect evidence is an important component of network meta-analysis. Recently, an evidence-splitting model has been proposed, that allows separating direct and indirect evidence in a network and hence assessing inconsistency. A salient feature of this model is that the variance for heterogeneity appears in both the mean and the variance structure. Thus, full maximum likelihood (ML) has been proposed for estimating the parameters of this model. ML is known to yield biased variance component estimates in linear mixed models. The purpose of the present paper, therefore, is to propose a method based on residual maximum likelihood (REML). Our simulation shows that this new method is quite competitive to methods based on full ML in terms of bias and mean squared error. In addition, some limitations of the evidence-splitting model are discussed. While this model splits direct and indirect evidence, it is not a plausible model for the cause of inconsistency.

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, Stuttgart, Germany; phone +049-711-459-22386
Email: piepho@uni-hohenheim.de

3:15 pm to 4:30 Poster Session—Abstracts listed alphabetically on pages 19 through 26

Day 4 - Thursday, May 18th, 2023

8:30 am to 10:05 am Session 8 — Moderator: Guillermo Marcillo

INVITED TALK: DATA-DRIVEN PERSPECTIVES IN A COMMERCIAL BREEDING PIPELINE

Marcia Almeida de Macedo[1]
[1]Syngenta Seeds, IA, USA

The development of cultivars requires many decisions throughout multiple generations of testing and advancement that starts with potentially thousands of candidates and ends, when successful, with an elite few. Limited seed, land, and resources coupled with large numbers of genotypes are some of the constraints that influence the configuration of breeding trials such as plot size and number of testing environments. For this reason, we may encounter interesting statistical problems such as estimating large numbers of parameters based on low numbers of observations, the possible confounding effect of inter-plot competition that contributes to bias and statistical error and using breeding trial data to predict performance on farmers' fields. High throughput phenotypic, genotypic, and environmental data given by new technologies open new possibilities and challenges for analytics. We will present examples of analytical solutions to problems we have encountered and some interesting challenges in search of solutions.

Presenter Bio: Dr. Almeida de Macedo is a senior statistician at Syngenta Seeds where she works with complex design decisions and analyses of large-scale plant breeding field experiments. She holds a doctorate degree in applied statistics and has a background in engineering.
Contact Information: Syngenta Seeds, 2369 330th St, Slater, IA, Phone: 515-230-6592,
Email: marcia.almeida_de_macedo@syngenta.com

MULTIMODAL ANALYTICS: BRINGING TOGETHER THE BEST OF STATISTICS, MACHINE LEARNING AND SUBJECT MATTER EXPERTISE FOR AGRICULTURE

John Gottula[1], *John Berry*[1], *Shelly Hunt*[1] *and Austin McMillan*[2]
[1]SAS Institute, Inc, Cary, NC
[2]SAS Institute, Inc, Indianapolis, IN

Industry research indicates that 60-85% of analytics projects fail to go into production. This is indicative of 1) a frequent lack of alignment of talent to the business goals, 2) the inability to effectively collaborate in data across disciplines and 3) project roadmaps that don't effectively account for messiness inherent in all big data projects. Leaders tasked to manage large scale insight-driven projects have responded by adopting multimodal analytics. Multimodal analytics is characterized by a rich, daily and deep "in the data" collaboration among users of diverse skillsets, from different disciplines and exhibiting multifaceted interests. In this SAS Viya® demonstration we will show how a breeder can explore and visualize, how a data engineer can cleanse and summarize, how a statistical modeler can parameterize and estimate field trial results, and how a machine learning engineer can reduce drone imagery dimensions and predict. By collaborating in the same source data in the same environment, we show how collaborators can extract faster business value, crisper insights, and support data-readiness within a diverse workforce.

Presenter Bio: Hailing from a seventh generation Nebraska farm, Dr. John Gottula is a practicing crop scientist and innovation manager at SAS. John drives digital ag strategies by emphasizing deep data scientist – subject matter expert collaboration through an analytics-first approach.
Contact Information: John Gottula, SAS Institute, Inc., Cary, NC, Phone: 806-441-4427, Email: john.gottula@sas.com

PREDICTING THE EFFECT OF GENETICALLY-ENGINEERED SEEDS ON THE CROP-YIELD DISTRIBUTION IN CHINA BY SYNTHETIC CONTROL METHODS

Yuansen Li[1], *Tor N. Tolhurst*[1], *Matthew Gammans*[2]
[1]Purdue University, West Lafayette, IN, USA
[2]Michigan State University, East Lansing, MI, USA

China recently began to allow for the domestic development and production of genetically engineered (GE) crops. We use synthetic control methods to predict the effect of GE-seed adoption on the time-conditional distribution of crop yields. Intuitively, synthetic control mimics an ideal experiment data from the US (GE-regime) to create a synthetic counterfactual for China (non-GE regime). This allows us to address a major challenge of estimating the long-run effect of allowing GE: for example, the synthetic counterfactual accounts for long-run effects, such as if allowing GE opens new pathways for innovation, biologically or otherwise. We measure the distributional treatment effect of GE using the distributional synthetic control (DSC) estimator of Gunsilius (2020). Analogous to conventional (i.e., mean) synthetic control, the key identifying assumption is that the DSC weights used to estimate the synthetic distribution are constant across the pre- and post-period. In doing so, our approach estimates the consequences of the GE policy divergence on the time-conditional distribution of crop yields. We apply our methods using state-level data on maize yields from China and the US.

Presenter Bio: Yuansen Li is a second-year Ph.D. student in the Department of Agricultural Economics at Purdue University. His main research interests are in trade and political economy, as well as applied econometrics.
Contact Information: Yuansen Li, Purdue University, West Lafayette, IN, Email: li4238@purdue.edu

APPLICATION OF 1D TRANSECT DATA ANALYSIS AND DATA MINING TO IDENTIFY THE DRIVERS OF SITE-SPECIFIC SOYBEAN YIELD VARIABILITY.

Ujjwal Sigdel[1], *Aurelie Poncet*[1], *Wesley France*[1], *Grant Rothrock*[1], *Jeremy Ross*[2], *Mario Soto*[1], *Kris Brye*[1]
[1]University of Arkansas, Fayetteville, AR, USA
[2]Cooperative Extension Services, Little Rock, AR, USA

Most of Arkansas soybean are cultivated on raised-beds. Irrigation is applied between the beds and water is delivered by gravity. Heterogeneous amounts of water are applied across the field and interactions with in-field changes in topography and soil properties may increase in-field yield variability and the potential loss in profitability. The objective of this study was to identify the relative contribution of spatial dependencies, field conditions, management, and measurement error on soybean yield. Different seeding rates and fertilization treatments were applied in strips: 247,000, 308,000, and 370,000 seeds ha^{-1} with phosphorus (P) and potassium (K) fertilization, and 370,000 seeds ha^{-1} with and without P and K fertilization. Yield, yield components, and soil fertility metrics were collected every 15 m along the middle of each treatment strip. Correlations between yield and the collected data for each transect were investigated using linear regression analysis when no significant spatial dependencies were found between measurements, and state-space modeling otherwise. A meta-analysis of the results obtained between transects was performed using various data mining techniques.

Presenter Bio: Ujjwal Sigdel is a master's student in Agronomy and Crop Science, concentrating on precision agriculture.
Contact Information: Aurelie Poncet, University of Arkansas, Fayetteville, AR, Phone: 479-575-3979,
Email: poncet@uark.edu

10:30 am to 11:45 am Session 9 **Moderator: Xin Dai**

TREE SPECIES SHIFT ACROSS CONUS IN THE PAST 20 YEARS

Jianmin Wang[1], *Brady Hardiman*[1], *Jonathan Knott*[2], *Brian Walters*[2], *Leah Griffin*[1], *Rebekah Shupe*[1], *Songlin Fei*[1]
[1]Purdue University, West Lafayette, IN, USA
[2]USDA Forest Service, St. Paul, MN USA

Tree species distribution can shift as climate change. Understanding the dynamics of these shifts is crucial for predicting future changes in forest ecosystems and for developing effective management strategies. Previous inventory-based studies have found divergent tree shifts that depend on factors such as species, region, period, and inventory method. In this study, we examine the tree shift for 150 species across the CONUS in the last 20 years using the recently available remeasured annual forest inventories of a consistent method from USDA Forest Inventory and Analysis (FIA) program. We found that more tree species in the western US have shifted upwards while more species in the eastern US have shifted northwestward and northeastward. Saplings have shifted faster than adult trees in all directions. Interestingly, obligate wetland plants tended to experience a downward, eastward, and southward shift as the water availability. Further analysis is ongoing to explore how these shifts are driven by climate change and are related to species traits such as dispersal type and seed weight.

Presenter Bio: Dr. Wang is a postdoctoral research associate in forestry and remote sensing at Purdue University. He has extensive experience in satellite remote sensing such as data fusion and phenology detection, and is familiar with forest inventory data.
Contact Information: Jianmin Wang, FORS 204, 195 Marsteller Street, West Lafayette, IN 47907.
Email: wang5736@purdue.edu

COMPARISON OF DIFFERENT STATISTICAL APPROACHES TO ASSESS THE APPLICABILITY OF VARIABLE-RATE SEEDING FOR ARKANSAS SOYBEAN

Wesley France[1], *Aurelie Poncet*[1], *Ujjwal Sigdel*[1], *Jeremy Ross*[2], *Grant Rothrock*[1]
[1]University of Arkansas, Fayetteville, AR, USA
[2]Cooperative Extension Services, Little Rock, AR, USA

Arkansas soybean seeding rates are selected based on soil properties and management. Recommendations were created to optimize whole-field management but an increasing number of farmers are using them for variable-rate seeding (VRS) applications without proof that they maximize profitability and efficiency within-field. The objective of this study was to compare the use of different statistical and data mining techniques to evaluate the relevance of current recommendations for VRS in Arkansas. Soil fertility metrics, soil texture, population counts, and yield data were collected in two production fields that were strong candidates for VRS due to significant in-field variability. In each field, soybeans were cultivated on raised-beds. Irrigation was delivered between beds and five seeding rate treatments were established to bracket the normal range: 85K, 247K, 308K, 370K, and 432K seeds/ha. Drivers of yield variability were identified using linear regression analysis, factor analysis of mixed data, clustering, and random forest to identify the drivers of variability. Comparison of results provided insight on the parameters that could be accounted for to optimize VRS for soybean in Arkansas.

Presenter Bio: Wesley France is a weed scientist by training and works as a program associate in precision agriculture. He has experience conducting on-station and on-farm research trials and working with precision technologies.
Contact Information: Aurelie Poncet, University of Arkansas, Fayetteville, AR, Phone: 479-575-3979,
Email: poncet@uark.edu

USING IMPRECISE MEASURES OF LANDSCAPE DIVERSITY TO QUANITFY THE MARGINAL EFFECT OF DIVERSITY ON CROP YIELDS

Katherine S. Nelson[1], *Emily K. Burchfield*[2], ***Brennan L. Bean***[3]
[1]Department of Geography and Geospatial Sciences, Kansas State University, Manhattan, KS, USA
[2]Department of Environmental Sciences, Emory University,
Atlanta, GA, USA.
[3]Department of Mathematics and Statistics, Utah State University, Logan, UT, USA.

Diverse landscapes tend to foster better ecological outcomes, including agricultural yields, as compared to less-diverse landscapes. One issue with quantifying the positive effect of diversity on yields is the ambiguity associated with measuring diversity. To address this issue, this paper defines several metrics of diversity for the Conterminous United States (CONUS) and creates ensembles of Bayesian generalized additive models, where each member of the ensemble uses a different metric to represent diversity. The result is a collection of posterior distribution estimates that are used to develop a mixture distribution that allows for the calculation of within-metric and across-metric variance in the marginal effect of diversity on yield prediction. Included also is a discussion of weighted mixture modeling approaches that account for correlations among the various diversity metrics. The results show that all ensemble members estimate a positive relationship between yield and diversity across CONUS, albeit with a relatively large cross-metric variance in the estimates.

Presenter Bio: Dr. Bean is an Assistant Professor of Statistics with a research focus of applied spatial statistics in climate and engineering. Dr. Bean has been active in updating snow related design provisions in US building codes and standards.
Contact Information: Brennan Bean, Department of Mathematics and Statistics, Utah State University, Logan, UT.
Phone: 435-797-4130, Email: brennan.bean@usu.edu

AN ALGORITHM TO SUMMARIZE HIGH-RESOLUTION VOLUMETRIC SOIL WATER CONTENT SENSOR DATA

Aurelie Poncet[1], *Resham Thapa*[2], *Dennis Timlin*[3], *Harry Schomberg*[3], *David Fleisher*[3], *Chris Reberg-Horton*[4], *Steven Mirsky*[3]
[1]University of Arkansas, Fayetteville, AR, USA
[2]Tennesse State University, Nashville, TN, USA
[3]USDA-ARS, Beltsville, MD, USA
[4]North Carolina State University, NC, USA

Volumetric water content sensors are being increasingly used in research and production agriculture to describe temporal changes in soil water recharge and uptake. Sensors can be installed at one or multiple depths, and data loggers may be used to automate data collection. Such systems provide unique opportunities for real-time monitoring of soil water status, but the data processing and analysis remains challenging. The objective of this study was to automate high-resolution volumetric water content sensor data processing. An algorithm was created to identify soil water recharge events from volumetric soil water content data collected at a temporal resolution of 60 min or less. The identified events are then associated with rainfall events defined from high-resolution rainfall data. The results are summarized into a long data table that identifies periods of soil water recharge following rainfall, and periods where drainage, evapotranspiration under low stress conditions, and evapotranspiration under drought stress conditions dominate water flow in the soil profile. The information provided in the long table can be used to compare the data collected in different treatments or locations.

Presenter Bio: Dr. Poncet is an Assistant Professor of precision agriculture and remote sensing. She has experience working with high-resolution spatial and temporal data. Her research focuses on the development of best management practices for optimized crop management.

Contact Information: Aurelie Poncet, University of Arkansas, Fayetteville, AR, Phone: 479-575-3979,
Email: poncet@uark.edu

1:00 pm to 2:00 pm Awards and Closing Remarks

Day 3 - Poster Presentations

LEARNING NETWORK STRUCTURE FOR POTENTIAL DETERMINANTS OF FOOD SAFETY PRACTICES IN CAMBODIA'S INFORMAL VEGETABLE MARKETS

Vrinda Ambike[1]*, Sabrina Mosimann*[2]*, Keorimy Ouk*[3]*, Paul Ebner*[2] *and Nora M. Bello*[1]
[1]The Ohio State University, OH USA
[2]Purdue University, IN USA
[3]Royal University of Agriculture, Phnom Penh, Cambodia

The Capability, Opportunity, Motivation-Behavior (COM-B) model is a theoretical framework for understanding determinants of human behavior that may be used to develop context-specific interventions to promote behavioral change. Our objective in this study was to search for potential causal network structures among said behavioral determinants specific to food safety practices in informal vegetable markets in Cambodia. Data consisted of responses from 169 participants to 18 survey items related to COM-B constructs for food safety and measured on a 1-to-7 Likert scale. Inductive causation based on Spearman rank correlations was used to determine (conditional) dependence and independence amongst items, yielding an undirected dependency graph. Directed-separation (d-sep) was then used to identify unshielded colliders and orient edges. Results indicated a partially oriented graph depicting closely interconnected survey items. The data-informed interconnections between behavioral determinants of food safety practices in Cambodian vegetable markets were consistent with expectations from the COM-B model. However, only a limited number of network edges could be oriented based on these data. Additional empirical work is warranted to further refine the COM-B model.

Presenter Bio: Vrinda Ambike is a first-year Ph.D. student in the Department of Animal Sciences at the Ohio State University working with Dr. Nora Bello. Her research interests include the application of statistical techniques for modeling animal production data.
Contact Information: Vrinda Ambike, Department of Animal Sciences, OSU, OH, Phone: 614-930-8759, Email: ambike.2@buckeyemail.osu.edu

USING PUBLICLY AVAILABLE CLIMATIC DATA TO IMPROVE HINDCASTS OF GOPHER TORTOISE NEST TEMPERATURES FOR DATA IMPUTATION

Alicia Arneson[1]*, Kevin Loope*[1]*, and Leah Johnson*[1]
[1]Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

The purpose of this work is to design a modeling framework to hindcast subterranean nest temperatures back to an initial lay date, given hourly nest temperature data from the date of nest discovery forward. The hindcasts from this model will be used in a future project investigating the effect of nest thermal environment on hatching success. Publicly available climatic data from the Daymet Daily Surface Weather and Climatological Summaries and the PERSIANN (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks) system developed by the Center for Hydrometeorology and Remote Sensing (CHRS) at the University of California, Irvine (UCI) were used to improve the quality of hindcasts produced from the incomplete series of iButton data logger recorded temperatures alone. Several models were fitted and tested on one nest to find the best approach. Then, the process of model fitting using the most informative predictors was automated to expand model fitting and testing to the other 199 nests. Hindcasts produced from a linear model and adjusted using hindcasted residuals produced using an AR(2) model of the residuals proved to be reasonable accurate for most of the nests.

Presenter Bio: Alicia Arneson is a first year PhD student in the quantitative ecological dynamics (QED) lab led by Dr. Leah Johnson and has worked with the Statistical Applications and Innovations Group (SAIG) on campus for two years.
Contact Information: Alicia Arneson, Virginia Polytechnic Institute and State University, Blacksburg, VA, Phone: 540-290-8672, Email: aga98@vt.edu

TEMPORAL ANALYSIS OF THE EFFECTS OF URBANIZATION ON LYME DISEASE RATES IN THE UNITED STATES, 2008 - 2020

Daphne Fauber[1] *and Hsin-Yi Weng*[2]
[1]Purdue University, Department of Agricultural and Biological Engineering, West Lafayette, IN, USA
[2] Purdue University, Department of Comparative Pathobiology, West Lafayette, IN, USA

Lyme disease (LD) is a bacterial infection caused by *Borrelia burgdorferi* and transmitted to people through the bite of infected ticks. In the United States, LD has an estimated 476,000 people diagnosed each year according to insurance claims data analyzed by Schwartz et al. LD has the highest incidence rates in the Northeast, northern Midwest, and Northwest regions, though cases of LD have been documented in nearly every state. The known areas of risk have expanded outward over time to include wider geographic ranges. Urbanization has been posited as a contributor to LD risk and could be a factor in the geographic spread over time. The goal of this research is to identify trends in LD rates as compared to urbanization indicators between 2008 and 2020. The urbanization indicators used include Rural-Urban Continuum Codes, Rural-Urban Commuting Codes, USDA Land Use and Crop Cover Data, and USDA County Typology Codes for the entire United States. Poisson regression and zero-inflated negative binomial regression are used to identify space-time clusters and to correlate with the urbanization indicators. Results from this study can be used to inform public health policy of risks associated with LD.

Presenter Bio: Daphne Fauber (she/her) is a Master's student at Purdue University doing research in biotechnology and bioinformatics workforce development with an additional interest in data accessibility. She is also a certified 5-12 educator in a variety of STEM disciplines.
Contact Information: Daphne Fauber, dfauber@purdue.edu; Hsin-Yi Weng, weng9@purdue.edu

DECISION-SUPPORT TOOL DEVELOPMENT AT THE NEXUS OF STATISTICS AND PROGRAMMING FOR BETTER INTEGRATION OF RESEARCH WITH EXTENSION

Wesley France[1]*, Aurelie Poncet*[1]*, Larry Purcell*[1]*, Trenton Roberts*[1]*, Jason Kelley*[2]
[1]University of Arkansas, Fayetteville, AR, USA
[2]Cooperative Extension Services, Little Rock, AR, USA

Optimized crop management requires adequate understanding of site-specific crop needs and response to management. Today, large amounts of crop, soil, and equipment performance-related data are collected through precision agriculture. Advances in statistical programming and the growing interest for spatial statistics, machine learning, and artificial intelligence have resulted in the development of data-driven recommendations proven to address some of the inefficiencies that affect modern crop production systems. However, the data are often not directly accessible to producers and integration into decision-support tools is becoming an essential component of the land-grant mission. Our team is finalizing the development of a decision-support tool prototype that assesses mid-season corn nitrogen status from remote sensing imagery. The objective of this presentation is to discuss our approach to decision-support tool development and lessons learned. The following topics will be addressed: model development, process automation, integration into a user-friendly interface, field validation, addition of new functionalities, and considerations for tool deployment.

Presenter Bio: Wesley France is a weed scientist by training and works as a program associate in precision agriculture. He has experience conducting on-station and on-farm research trials and working with precision technologies.
Contact Information: Aurelie Poncet, University of Arkansas, Fayetteville, AR, Phone: 479-575-3979, Email: poncet@uark.edu

(VIRTUAL) STATISTICAL INFERENCE FOR INTERVAL MONITORED STEP-STRESS ACCELERATED LIFE TEST UNDER PROGRESSIVE TYPE-I CENSORING

***David Han*[1]** *and Tianyu Bai*[1]
[1]The University of Texas at San Antonio, TX, USA

In order to examine the lifetime of a given test item, a standard life test at normal conditions is not suitable when the lifespan is significantly long. Examples of test items in an agricultural context include plants and animals and the stress variables could be the dosage of a chemical, the amount of fertilizer, temperature, and the like. Accelerated life tests, in contrast, provide a solution to this problem by subjecting the test items to more extreme stress levels for quicker and more reliable results. In this work, we consider a step-stress accelerated life test under progressive Type-I censoring when the continuous monitoring is infeasible but inspections at particular time points is possible. In addition to the accelerated failure time model to explain the effect of stress changes, a general scale family of distributions is considered for flexible modeling by allowing different lifetime distributions at different stress levels. Assuming that the inspection points align with the stress-change time points, the maximum likelihood estimators of the scale parameters and their conditional density functions are derived explicitly. If the inspection points do not align with the stress-change time points, the parameter estimates can be obtained numerically.

Presenter Bio: Dr. Han is a statistics professor with more than 15 years of experience. He has extensive expertise in lifetime analyses with more than 100 publications and technical reports.
Contact Information: Dr. David Han, The University of Texas at San Antonio, TX, Phone: 210-458-7895
Email: david.han@utsa.edu

(VIRTUAL) ORDER-RESTRICTED BAYESIAN INFERENCE & OPTIMAL DESIGN FOR SIMPLE STEP-STRESS ACCELERATED LIFE TEST UNDER PROGRESSIVE TYPE-I CENSORING

Crystal Wiedner[1] *and* ***David Han*[1]**
[1]The University of Texas at San Antonio, TX, USA

This work looks into the order-restricted Bayesian estimation and design optimization for step-stress accelerated life tests with an exponential lifetime under both continuous and interval inspections. For example, in the agricultural field, the test items can be plants or animals and the stress variable can be the amount of fertilizer, chemicals, and temperature. Using the three-parameter gamma distribution as a conditional prior, it is ensured that the failure rates will increase with the increase in stress levels. The conjugate-like structure of the prior allows us to determine the joint posterior distribution of the parameters without having to use an expensive MCMC sampling. We then propose several Bayesian estimators for the model parameters and their individual/joint credible intervals. The Bayesian design optimization under various design criteria such as Shannon information gain and posterior variance-covariance matrix are discussed. The performance of the proposed inferential methods is assessed and compared between the continuous and interval inspections using Monte Carlo simulations.

Presenter Bio: Dr. Han is a statistics professor with more than 15 years of experience. He has extensive expertise in lifetime analyses with more than 100 publications and technical reports.
Contact Information: Dr. David Han, The University of Texas at San Antonio, TX, Phone: 210-458-7895
Email: david.han@utsa.edu

BIG DATA ANALYTICS, DATA SCIENCE, ML&AI FOR CONNECTED, DATA-DRIVEN PRECISION AGRICULTURE AND SMART FARMING SYSTEMS: CHALLENGES AND FUTURE DIRECTIONS

***David Han*[1]**
[1]The University of Texas at San Antonio, TX, USA

Big data and data scientific applications in the modern agriculture are rapidly evolving as the data technology advances and more computational power becomes available. The adoption of Big data has enabled farmers to optimize their agricultural activities sustainably with cutting-edge technologies, resulting in eco-friendly and efficient farming. Wireless Sensor Networks (WSNs) and Machine Learning (ML) have had a direct impact on smart and precision agriculture, with Deep Learning (DL) techniques applied to data collected via sensor nodes. Additionally, robotics, Internet of things (IoT), and drones are being incorporated into farming techniques. Digital data handling has amplified the information wave, and information and communication technology (ICT) have been used to deliver benefits for both farmers and consumers. This work highlights the technological implications and challenges that arise in data-driven agricultural practices as well as the research problems that need to be solved.

Presenter Bio: Dr. Han is a statistics professor with more than 15 years of experience. He has extensive expertise in lifetime analyses with more than 100 publications and technical reports.
Contact Information: Dr. David Han, The University of Texas at San Antonio, TX, Phone: 210-458-7895
Email: david.han@utsa.edu

UTILIZATION OF RANDOM FOREST AND ARTIFICAL NEURAL NETWORKS IN ESTIMATING PIE CRUST COLOR COLOR UNDER VARIABLE LIGHTING CONDITIONS

***Harrison Helmick*[1]**, *Kara Benbow*[1], *Troy Tonner*[2], *Jozef Kokini*[1]
[1]Purdue University, West Lafayette, IN, USA
[2]Purdue Northwest University, Calumet, IN, USA

Estimating colorimeter values in the CIELAB color space under variable lighting conditions from image data is a difficult task. Here, RGB images of 127 color swatches from bright and dark lighting conditions were split into 10 segments, and the average R,G,B,L*,a* and b* values were extracted using Python's cv2. The slope of the lightness values in the horizontal and vertical directions were also calculated and used as indicators of glossiness / lighting conditions. These data were used for training (N = 1016) and testing (N = 254) RF and artificial ANNs using SciKit Learn. Grid searching was conducted on 28 models per lighting condition using different input data. On average, RF models had lower RMSE values than ANNs, and the slope of lightness improved model performance. A weighted average of light and dark conditions calculated through Euclidian distances after PCA also led to reasonable estimates at intermediate lighting conditions. The model was validated for food uses on pie crusts brushed with pea protein - glycerol coatings intended to replace egg-washes for consumers seeking more sustainable products with plant-based protein. This work may contribute to computer vision systems used with variable lighting conditions in difficult to access areas, such as inside bakery ovens.

Presenter Bio: Harrison Helmick is a PhD candidate in the department of food science at Purdue University. He studies the intersection of food science and data science with a focus on structural bioinformatics and analyzing data from diverse sources including video, image, and text.
Contact Information: Harrison Helmick, Kokini Laboratory Research Group. West Lafayette, IN
Email: hhelmick@purdue.edu

A FRAMEWORK TO AUTOMATE THE STATISTICAL DESIGN OF FIELD EXPERIMENTS FOR MODERN FARM MANAGEMENT PRACTICES

***Sneha Jha**, Yaguang Zhang, James V. Krogmeier and Dennis R. Buckmaster*
Purdue University, West Lafayette, IN, USA.

Modern precision agriculture technology has provided farmers with a wide variety of field management practices based on irrigation conditions, weather, land use, crop requirements etc. On-farm trials help farmers gain a better understanding of the efficiency of these modern field management techniques and tools when applied in their local conditions. However, the traditional randomized block design used for the statistical design of field trials is of limited utility when the natural variability of the fields influences the outcome from changes in nutrient management, seeding, crop management etc. Therefore, refinement of experiments is necessary to better understand the characteristics and limitations of these new farm management techniques in local conditions. This work integrates natural field variations, including elevation, slope, aspect, soil classification, drainage class, etc., with historic yield data to suggest appropriate design of experiments for modern farm management practices.

Presenter Bio: Sneha Jha is a graduate research assistant in the department of Agriculture and Biological Engineering at Purdue University.
Contact Information: Sneha Jha, Agricultural and Biological Engineering, 225 South University Street, West Lafayette 47907, Email: jha16@purdue.edu

A PRACTICAL APPROACH TO THE EVALUATION OF PROXIMAL SOIL SENSOR DATA ACCURACY

***Aurelie Poncet**[1], Francisco Velasquez[1], Wesley France[1], Andy Mauromoustakos[2], Nathan Slaton[3]*
[1]University of Arkansas, Fayetteville, AR, USA
[2]Agriculture Statistics Laboratory, Fayetteville, AR, USA
[3]Agricultural Experiment Station, Fayetteville, AR, USA

Soil data from field sampling are widely used in Arkansas to measure in-field changes in soil pH and nutrient availability. Most field samples are collected using 2.5- or 5.0-acre grids without evidence that the chosen resolution appropriately describes in-field variability. The optimal sampling resolution may vary between and within fields and proximal soil sensing may be used to inform field sampling strategies and soil fertility management providing adequate data accuracy. The objective of this study was to develop a method that evaluates proximal soil sensor data accuracy. Gamma-ray spectrometer data and field samples were collected in a selection of fields to characterize in-field changes in soil pH and specific nutrient availability. Comparison of the cumulative empirical density functions determined if the sensor provided an accurate measurement of the central tendency, skewness, and variability of the ground-truth data. The Moran's I indices and spatial linear regression analysis were used to characterize spatial associations within the data and determine if the sensor provided an accurate measurement of the spatial distribution of the ground-truth data.

Presenter Bio: Dr. Poncet is an Assistant Professor of precision agriculture and remote sensing. She has experience working with high-resolution spatial and temporal data. Her research focuses on the development of best management practices for optimized crop management.

Contact Information: Aurelie Poncet, University of Arkansas, Fayetteville, AR, Phone: 479-575-3979,
Email: poncet@uark.edu

OPTIMIZING THE SCALE OF SPATIAL AGGREGATION IN THE ANALYSIS OF MINIRHIZOTRON DATA

***Simon Riley**[1], James Colee[1], Edzard van Santen[1]*
[1]Statistical Consulting Unit, Institute of Food and Agricultural Science, University of Florida, Gainesville, FL, USA

Minirhizotrons play an important role in crop root systems research, serving as one of the few vehicles for taking non-destructive measurements of root length and diameter over a cross-section of the soil profile under field conditions. It is common for these measurements, taken at e.g. 9.55mm intervals, to be aggregated into depth classes prior to analysis, in order to reduce the computational resources required for model fitting, to reduce the (otherwise very high) proportion of zeroes in the data, and to help ameliorate the characteristic heavy right skew. No research has yet been undertaken to ascertain what effect this aggregation - which introduces the modifiable areal unit problem (MAUP) – has on the efficiency of the subsequent analysis. This research thus seeks to determine, for various degrees of serial autocorrelation, how statistical power is affected by the choice of depth class and, on that basis, to identify if possible, the optimum scale of aggregation. The study is undertaken using simulation, and the findings subsequently applied to the analysis of a minirhizotron study of peanut (*Arachis hypogaea* L.).

Presenter Bio: Simon Riley is a data analyst at the Statistical Consulting Unit of the University of Florida's Institute of Food and Agricultural Science (IFAS SCU). He is currently completing his doctorate in agronomy, his current research focusing on statistical modeling of crop roots.
Contact Information: Simon Riley, 404 McCarty C, PO Box 110500, Gainesville, FL 32611,
Email: simon.riley@ufl.edu

COMPARISON OF DIFFERENT SCOUTING TECHNIQUES TO MONITOR THE EMERGENCE AND INTENSITY OF YIELD-LIMITING STRESS IN SOYBEANS

***Ujjwal Sigdel**[1], Aurelie Poncet[1], Wesley France[1], Grant Rothrock[1], Jeremy Ross[2], Mario Soto[1], Kris Brye[1]*
[1]University of Arkansas, Fayetteville, AR, USA
[2]Cooperative Extension Services, Little Rock, AR, USA

Scouting with precision technologies has become a critical component of optimized crop management. Proximal and remote sensing are being increasingly used to measure in-field changes in crop vegetative development and plant physiological stress as proxies for plant health. The objective of this study was to evaluate the performance of different sensing-based scouting methods for Arkansas soybeans. Different seeding rates and fertilization treatments were established in four production fields with strong in-field changes in topography and soil properties. The treatments were established in strips. Leaf area index, chlorophyll fluorescence, stomatal conductance, and remote sensing-based vegetation indices computed from Sentinel-2 satellite imagery were collected along the middle of each strip during critical stages. Correlations between yield and the sensor data were investigated by transect using linear regression analysis when no significant spatial dependencies were identified along the transect, and state-space modeling otherwise. A meta-analysis of the results across transects and scouting methods was performed to identify best scouting practices.

Presenter Bio: Ujjwal Sigdel is a master's student in Agronomy and Crop Science, concentrating on precision agriculture.
Contact Information: Aurelie Poncet, University of Arkansas, Fayetteville, AR, Phone: 479-575-3979,
Email: poncet@uark.edu

APPLIED STATISTICS WITH CHATGPT: BOONS AND BANES

John R. Stevens[1] *and Maha Moussa*[1]
[1]Utah State University, Logan UT, USA

When ChatGPT was first launched in November 2022, it opened an era of widely-accessible AI chatbots that have the potential to drastically change how interdisciplinary statisticians train students, collaborate with researchers in various fields, and approach their own statistical research questions. We outline some issues to consider when using ChatGPT in these capacities, and demonstrate the kinds of opportunities and pitfalls this technology presents. Depending on how we use (or ignore) them, these AI chatbots can either empower or frustrate our objectives in professional activities ranging from assigning homework to designing studies to writing research presentations.

Presenter Bio: Dr. Stevens is Professor of Statistics and Interim Department Head, with a research program involving a range of applications and methodological developments of statistics in the biological and agricultural sciences.
Contact Information: John Stevens, 3900 Old Main Hill, Dept of Mathematics and Statistics, Utah State University, Logan UT 84322-3900; Phone: 435-797-2810, Email: John.R.Stevens@usu.edu

BEHIND THE DECLINE OF ALASKA SNOW CRAB – EXPLORING CLIMATE CHANGE FACTORS & POLICY OPTIONS

Jingjing (Tina) Tao[1]***,*** *Weicheng Wang*[1]
[1]Purdue University, West Lafayette, IN, USA

For the first time in state history, Alaska fisheries officials announced the shutdown of the snow crab harvest due to a catastrophic shrink of the sizable crustaceans in the Bering Sea. Billions of Alaska snow crabs have disappeared in the past two years, and local fishers are estimated to lose millions of dollars. Crab farmers, also known as sea crabbers, are calling for explanations and solutions for economic resilience when facing this blow. Yet the reason behind this vast collapse, nearly a 90% drop in the snow crab population, remains unclear. As a vital element of Alaska's economy and a global source of seafood, this crush is a seismic shock to the state's crab business. To better understand the mechanisms of the mysterious decline of the Alaska snow crab so as to bring insights to policymakers and stakeholders further, this paper plans to explore how the ecosystem, especially climate crisis factors across the regions of the Eastern Bering Sea and Alaska coast, influence the snow crab supply and crabber's income; and then, provide policy implications and suggestions concerning the scenario.

Presenter Bio: Jingjing is a first-year master's student at the Agricultural Economics department. Weicheng is a junior student in Data Science at the department of Computer Science.
Contact Information: Jingjing (Tina) Tao, Purdue University, West Lafayette, IN, Phone: 765-775-9026, Email: tao121@purdue.edu

GEOSPATIAL ANALYSIS OF OPTIMAL GNSS PLACEMENT IN PRECISION FORESTRY

Aishwarya Chandrasekaran[1]*, Max Hess*[2]*,* ***Audrey Ward***[3]
[1]Purdue University, Forestry and Natural Resources
[2]Purdue University, Civil Engineering
[3]Purdue University

Locating individual trees spatially is a critical component of efficient forest management. However, conventional forest inventorying procedures provide limited to no information on the location of individual trees. Using Global Positioning System (GPS) units within a forest requires multiple expensive base stations outside the forest perimeter while resulting in limited accuracy due to interference from the tree canopy. In this study, we examine multiple Global Navigation Satellite System (GNSS) models to obtain clusters of GPS coordinates on the perimeters of forest plots. These coordinate measurements will be used to compare the precision and accuracy of the GNSS units. They will also allow us to model the effects of survey equipment placement relative to the trees and predict optimal locations for establishing temporary benchmarks with known global coordinates. Ultra-wideband transmitters can then be used to triangulate the individual tree's local position relative to the benchmarks, resulting in a robust and cost-efficient, location-tagged tree inventory system.

Presenter Bio: Audrey Ward is a homeschool student finishing their first year of research at Purdue. They have presented earlier results of this research at the Purdue Undergraduate Research Symposium and the Purdue Undergraduate Research Expo.
Contact Information: Audrey Ward, Purdue University, Phone: 765-491-8482, Email: ward294@purdue.edu