



## Spring Meeting 2026

**Date:** Friday, May 8, 2026

**Time:** 9:40 am – 4:30 p.m.

**Host:** Colorado School of Mines

**Location:** Ben H Parker Student Center, Ballroom A, 1516 Maple St, Golden CO, CO 80401

**Parking Information:** [Campus Visitor Parking](#)

**RTD Light Rail:** From the [Jeffco Government Center RTD station](#), take the free [Ore Cart](#) shuttle to Mines' campus (Tungsten Route).

**What is the ASA?** American Statistical Association

**Questions?** Contact Jason Bernstein at [jstatbernstein@gmail.com](mailto:jstatbernstein@gmail.com)

### Program Schedule

#### 9:00 – 9:40

Room becomes available, folks welcome to sit, work, chat, etc.

#### 9:40 – 9:50

Opening remarks:

- Jason Bernstein, President of CO-WY Chapter, 2025 – 2026
- Steve Pankavich, Professor & Department Head, Applied Mathematics and Statistics, Colorado School of Mines

#### 9:50 – 10:10

**From guessing to an ignorant baseline: Dummy kappa statistics for rater agreement**

Wren Lippitt

University of Colorado Anschutz

Quality of agreement between two raters is often quantified using chance-corrected statistics such as Cohen's  $\kappa$ . These measures answer a particular question—*do raters agree more than we would expect if they were guessing?*—but this is often not the question practitioners actually care about. In many applied settings (e.g., medical annotation, diagnostics, labeling rare outcomes), a more relevant question is: *Is their agreement any good?* To this end, a family of dummy kappa statistics which benchmark agreement with dummy classifiers rather than guessing is introduced. Different dummy kappas are found to arise in different applied contexts. They are compared theoretically and empirically to Cohen's  $\kappa$  and argued to align more appropriately with an assessment of 'goodness' of rater agreement, particularly in the case of rare or imbalanced ratings.

**10:10 – 10:30**

**Marginal Distribution of the LatticeKrig coefficients under Spatial Autoregressive model with Spatial Extremes**

Susmit Bhattacharyya

Colorado School of Mines

In gridded, large-scale spatial datasets, a central challenge is constructing an accurate spatial model to predict values at unobserved grid points. LatticeKrig (Nychka et al. 2015, Kleiber et al. 2019) is a widely recognized model in spatial statistics, particularly suited for handling massive, Gaussian datasets. In this framework, the Gaussian process is represented as a sum of basis functions with associated coefficients, which are typically unknown and need to be estimated or simulated for unobserved grid locations.

The complexity of this task increases when dealing with spatially heterogeneous data characterized by heavy-tailed marginal distributions. Simulation-based studies (Rai et al. 2025, Reich & Shaby 2019) have demonstrated that a Spatial Autoregressive (SAR) model is effective in such non-Gaussian settings. The SAR model links each location and its neighbors to a spatially independent random variable, thereby capturing local dependencies effectively. Currently, we are extending the SAR framework by incorporating Generalized Extreme Value (GEV) distributions as independent random variables within a multiplicative structure. However, identifying suitable linear combinations of GEV-distributed variables poses significant analytical challenges. To address this, we implement the SAR model on the logarithm of the coefficients, yielding independent random variables whose exponentials follow a GEV distribution. Our primary goal is to derive theoretical distributions for the coefficients or their logarithms and to estimate the distributional parameters by simulation study. Moreover, including analysis the behavior of the coefficients at extreme values and assessing dependence through the extremal index, particularly at higher values.

- Kleiber, W., Nychka, D. & Bandyopadhyay, S. (2019), ‘A model for large multivariate spatial data sets’, *Statistica Sinica* 29(3), 1085–1104.
- Nychka, D., Bandyopadhyay, S., Hammerling, D., Lindgren, F. & Sain, S. (2015), ‘A multiresolution gaussian process model for the analysis of large spatial datasets’, *Journal of Computational and Graphical Statistics* 24(2), 579–599.
- Rai, S., Nychka, D. & Bandyopadhyay, S. (2025), ‘Modeling spatial extremes using non-gaussian spatial autoregressive models via convolutional neural networks’, In preparation.
- Reich, B. J. & Shaby, B. A. (2019), ‘A spatial markov model for climate extremes’, *Journal of Computational and Graphical Statistics* 28(1), 117–126.

**10:30 – 10:50**

**Constructing a Coordinate System for the Surrogate Variable Analysis Model, with an Application to Yeast Gene Expression**

Riley Skeen-Gaar and John D Storey

Princeton University

In genomics, scientists routinely observe thousands of variables, testing each for correlation with treatment status, or some other aspect of the study design. However, these experiments often contain cryptic sources of variation, such as batch effects, that destabilize multiple hypothesis testing. The Surrogate Variable Analysis (SVA) model can be used to understand this challenge; the model is a multivariate regression with dependence among the response variables due to latent factors. Representing  $k$  response variables across  $n$  individuals as a  $k$  by  $n$  matrix, variation explained entirely by the study design satisfies a row space constraint, while variation explained by the latent factors satisfies a column space constraint. The linear subspace of matrices satisfying both constraints is ambiguous in the sense that it can be explained by both sources of variation. Standard SVA methods use regularization to resolve this ambiguity, arriving at a single point estimate. Instead, we develop a coordinate system for this subspace, without committing to a single estimate of regression coefficients or latent factors. Each basis matrix in this coordinate system is an outer product of two vectors, one from a QR decomposition of the design matrix and another from a QR decomposition of the estimated factor loading matrix. To show what can be gained using this approach, we apply it to a study on the genetics of gene expression in yeast. We show that a pattern of variation that could be explained by a batch effect can alternatively be explained by signal due to variation in the yeast's rate of growth.

**10:50 – 11:00**

Break

**11:00 – 11:20**

**The Rise of Early-Onset Colorectal Cancer: Evidence, Trends, and Drivers**

Ben Cevaal, Adam Dietz, Kit Kennedy, Huyen Nghiem, Emily Speckman, Jackson Tillinghast, Dr. Yanxi Li

MSU Denver

Recently within the last ten years, there has been a push towards colorectal cancer screening for younger groups compared to older populations traditionally targeted for these preventions. This study looked at the possibility of a significant increase of colorectal cancer rates in younger populations (less than 50) and if such an increase is the possible result of potential health and lifestyle changes among the population or rather results of earlier and more effective screening. Using data from the Surveillance, Epidemiology, and End Results (SEER) data set and National Health and Nutrition Examination Survey (NHANES), this study has used multiple statistical models to address the trend of colorectal cancer rates by age as well as these rates in relation to changes in general health in form of obesity rates. The study found that the colorectal cancer rates in the population under 50 are significantly increasing per year, but from the variables of tumor size, presence of tumor deposits as well as obesity rates, it was unable to discern to what degree this increase is due to improved screening techniques or changing health factors causing high cancer rates despite some connection to the release of improved colorectal cancer screening procedures. This study is an aid in the confirmation of increasing colorectal cancer rates among younger individuals under 50 and has detected significant changes in colorectal cancer diagnosis.

**11:20 – 11:40**

**An Exposition on the Importance of Anomaly Detection in AI**

Jacob Johns

University of Colorado Denver

Anomalous behavior being entombed in data is a fact of life that is often overlooked in the rapidly evolving field of AI. While a naive approach to anomalies is strict removal or ignorant dismissal, anomalies themselves can range from key observations or influential points to system errors or device failures. With this variety comes the need for a nuanced look at what anomalies are, how they affect data, how to find them, and how AI models react to their inevitable presence.

We define anomalies, their taxonomy, and highlight unintuitive edge cases. Building off the work of Chen, Fan and Jin, both the good and the bad aspects of anomalies' existence in data are presented. We explore and highlight assumptions, methodology, and computational aspects of several anomaly detection algorithms. Specifically, advantages and limitations of quintessential models will be exhibited in the context of big data. Furthermore, anomaly detection methods are described with their use cases and various applications specifically as part of an AI pipeline.

**11:40 – 12:00**

**Frequency Domain Resampling for Gridded Spatial Data**

Souvick Bera

Colorado School of Mines

In frequency domain analysis for spatial data, spectral averages based on the periodogram often play an important role in understanding spatial covariance structure, but also have complicated sampling distributions due to complex variances from aggregated periodograms. In order to non-parametrically approximate these sampling distributions for purposes of inference, resampling can be useful, but previous developments in spatial bootstrap have faced challenges in the scope of their validity, specifically due to issues in capturing the complex variances of spatial spectral averages. As a consequence, existing frequency domain bootstraps for spatial data are highly restricted in application to only special processes (e.g. Gaussian) or certain spatial statistics. To address this limitation and to approximate a wide range of spatial spectral averages, we propose a practical hybrid-resampling approach that combines two different resampling techniques in the forms of spatial subsampling and spatial bootstrap. Subsampling helps to capture the variance of spectral averages while bootstrap captures the distributional shape. The hybrid resampling procedure can then accurately quantify uncertainty in spectral inference under mild spatial assumptions. Moreover, compared to the more studied time series setting, this work fills a gap in the theory of subsampling/bootstrap for spatial data regarding spectral average statistics.

**12:00 – 1:00**

Lunch

Box lunches provided by Sodexo. Delivery window is 11:30 - 12:00.

**1:00 – 1:20**

## **Psilocybin Trends in States That Decriminalized Use**

Joshua C. Black

Rocky Mountain Poison & Drug Safety, Denver Health and Hospital Authority

Introduction: Amid nationally rising interest and use of psychedelics, the extent to which state-level decriminalization accelerated secular trends in use is not known. We estimate the increase in past year psilocybin use attributable to decriminalization in Oregon (November 2020) and Colorado (November 2022).

Methods: We analyzed data from 15 waves (2018 through 2025) of the Survey of Non-Medical Use of Prescription Drugs, a nationally representative, online, semi-annual survey of U.S. adults, calibrated to national benchmarks. Annual psilocybin use was calculated for each state and wave. The augmented synthetic control method was used to estimate the average treatment effect among the treated (ATT), which is the difference between the observed change in prevalence in Oregon and Colorado following decriminalization and a counterfactual control assuming decriminalization had never happened. Demographics, socioeconomic, health access and behaviors, and social environment predictors from non-decriminalized states were used to construct the control.

Results: Average annual psilocybin use in Oregon increased from 2.0% (95% CI: 1.5- 2.4) before decriminalization in 2020 to 4.5% (3.8-5.1) after. The secular trend in Oregon, estimated by its synthetic control, increased from 1.6% (1.1-2.0) to 2.3% (1.8- 2.9) yielding an ATT of 2.1% (0.7-3.6). In Colorado, average annual use increased from 2.7% (2.1-3.4) before decriminalization in 2022 to 4.3% (2.6-6.1) after. Colorado's synthetic control remained constant at 2.4% before (1.9-2.9) and after decriminalization 2.4% (2.0-2.9) with an ATT of 1.8% (-0.1-3.8).

Discussion: Findings indicate 70,000 to 90,000 additional Oregonians used annually following decriminalization, far exceeding the number using regulated services, signifying most increased use occurred in unregulated settings. Colorado showed a similar magnitude, though had more uncertainty due to the shorter post-intervention period. States considering decriminalizing psilocybin should anticipate increases in unregulated use, and plan for appropriate public health surveillance and harm-reduction strategies.

**1:20 – 1:40**

### **Novel Clinical Trial Design with Stratum-Specific Endpoints and Global Test Methods for Rare Diseases with Heterogeneous Clinical Manifestations**

Emily Shives

Center for Drug Evaluation and Research (CDER), Food and Drug Administration (FDA)

Many rare disease clinical trials are underpowered to detect a moderate treatment effect of an investigational product due to the limited number of participants available for the trials. In addition, given the complex, multisystemic nature of many rare diseases, it is challenging to confidently prespecify a single primary efficacy endpoint that is applicable to all trial participants with a heterogeneous clinical manifestation of their disease. Traditional trial designs and analysis methods often used in more common diseases to analyze the same endpoint(s) for all patients may be in-

efficient or impractical for a rare disease with heterogeneous clinical manifestations. To address these issues, we propose a novel trial design and analytic approach that allows for an evaluation of stratum-specific efficacy endpoints in a broader population of participants. We develop several non-parametric global test methods which can accommodate the novel design and provide global evaluation of treatment effects. Using a case example in patients with a rare disease, our simulation studies illustrate that the novel design evaluated using the global test methods may be more sensitive to detect a treatment effect compared to the traditional design that uses the same endpoint(s) for all patients.

**1:40 – 2:00**

**General semantic positional encoding with ggregions, or: Mapmaking that’s as easy as ggplot2 scatter plotting**

Gina Reynolds

University of Denver

Maps are some of the most compelling and intuitive data visualizations, and they are some of the first visualizations people will come across in their lives. There is great support for mapping in R and in the data visualization library ggplot2 which supports many R packages.

However, producing a simple map, like a choropleth, feels harder than producing other plots, like a scatterplot for example in ggplot2 due to the fact that with current tools data joins are the responsibility of the data analyst.

The ggregions framework and package proposes paths to allow regional semantic identifiers, e.g. 'Texas', 'North Carolina', 'WY', 'Colo' to be mapped to a new positional aesthetic 'region' for fluid and intuitive mapping with ggplot2.

**2:00 – 2:15**

Break

**2:15 – 3:00**

**Keynote Talk**

**The Explainability Gap: Why “It Works” Isn’t Enough For Real World AI**

Kassie Fronczyk

Leidos

Many enterprise analytic deployments fail not because models perform poorly, but because performance alone is insufficient for real-world adoption. The "explainability gap" describes the distance between a model that works statistically and one that earns the operational trust needed to actually be used. This talk presents a practical framework for closing that gap, arguing that explainability is not a nice-to-have feature but a core deployment requirement, and offering three diagnostic questions every team should be able to answer before shipping a model to production.

**3:00 – 3:20**

## **medicalcoder: A Unified and Longitudinally Aware Framework for ICD-Based Comorbidity Assessment in R**

Peter DeWitt

University of Colorado, Anschutz

Comorbidity algorithms derived from International Classification of Diseases (ICD) codes are central to risk adjustment and cohort characterization in clinical research. However, existing implementations often fragment across packages, inconsistently handle mixed ICD-9 and ICD-10 data, and typically rely on encounter-level aggregation that may under-ascertain chronic disease burden.

We present `medicalcoder`, an R package providing a unified, longitudinally aware framework for applying multiple variants of the Charlson, Elixhauser, and Pediatric Complex Chronic Conditions algorithms. The package includes an internal ICD database, supports full and compact codes, accommodates mixed ICD versions within a dataset, and integrates present-on-admission and primary diagnosis indicators.

Unlike encounter-level approaches that simply aggregate flags, `medicalcoder` implements cumulative longitudinal methods that propagate qualifying diagnoses forward in time, increasing sensitivity and improving detection of disease severity. The package is self-contained ( $R \geq 3.5.0$ ) and designed for portability in restricted computing environments while dynamically leveraging modern R workflows when available.

This talk will demonstrate longitudinal sensitivity gains, mixed-version handling, and practical workflows for reproducible comorbidity assessment in real-world EHR data.

**3:20 – 3:30**

Break

**3:30 - 3:50**

## **Confounding Robust Reinforcement Learning: A Causal Approach**

Junzhe Zhang

Syracuse University

Standard off-policy reinforcement learning algorithms assume that observed data are free from unmeasured confounding. This assumption is routinely violated in practice—for example, when the demonstrator’s behavior policy has access to information not recorded in the data, or when there is a mismatch between the demonstrator’s and the learner’s sensory capabilities. In this talk, I present a line of work that addresses confounding bias in offline RL through causal inference and partial identification. I introduce the Confounded MDP framework, extend Bellman’s equation to derive informative bounds on value functions from confounded observations, and develop model-free temporal difference algorithms using causal eligibility traces that work even when there is no common support between behavior and target policies. I then show how these ideas scale to high-dimensional domains: first through Causal DQN, which learns robust policies from confounded visual observations across twelve Atari games, and then through Causal Flow Q-Learning, which combines the

partial identification framework with flow-matching generative models to handle continuous actions in twenty-five pixel-based robotic tasks. Together, these results demonstrate that principled causal reasoning leads to practical, scalable algorithms for robust decision-making from biased offline data.

**3:50 – 4:10**

**Expanding LatticeKrig: Introducing a Tensor Product Cubic B-Spline Basis Function to the Multiresolution Gaussian Process Model LatticeKrig**

Ryan H. Peterson, Douglas Nychka, and Dorit Hammerling  
Colorado School of Mines

Remotely sensed observations of the atmosphere play an important role in climate research since they have more extensive spatial coverage than surface measurements. However, there are multiple challenges that arise from the large quantities of data needed to provide the necessary spatial coverage. A useful framework for spatial models involves expanding the field using basis functions and making distributional assumptions about the basis coefficients. This approach forms the foundation of the successful fixed rank kriging (FRK) methodology, which has been adopted and extended by models such as LatticeKrig. We introduce a tensor product cubic B-spline basis for representing a multi-resolution Gaussian process model. Surprisingly, the use of B-splines as spatial basis functions has not been extensively explored, despite their several advantages. The cubic B-spline basis function presented here is compactly supported, preserving the efficient sparse linear algebra used in LatticeKrig. In addition, we leverage the partition of unity property of splines to reduce basis function artifacts and develop more accurate numerical integration over irregular spatial regions for change-of-support methods. This work is illustrated through an application to total column carbon monoxide (CO) data over eastern Australia, collected from satellite instruments in the form of footprints that vary in size and degree of overlap between successive observations.

**4:10 – 4:30**

**Chapter Business**

- Maurice Davies Awards
- David Young Awards
- Chapter officer elections
- Discussion of chapter finances and potentially raising membership fees a dollar

**After Meeting**

**No-host social**

Following the meeting, we can have a no-host social at somewhere in Golden, location TBD.