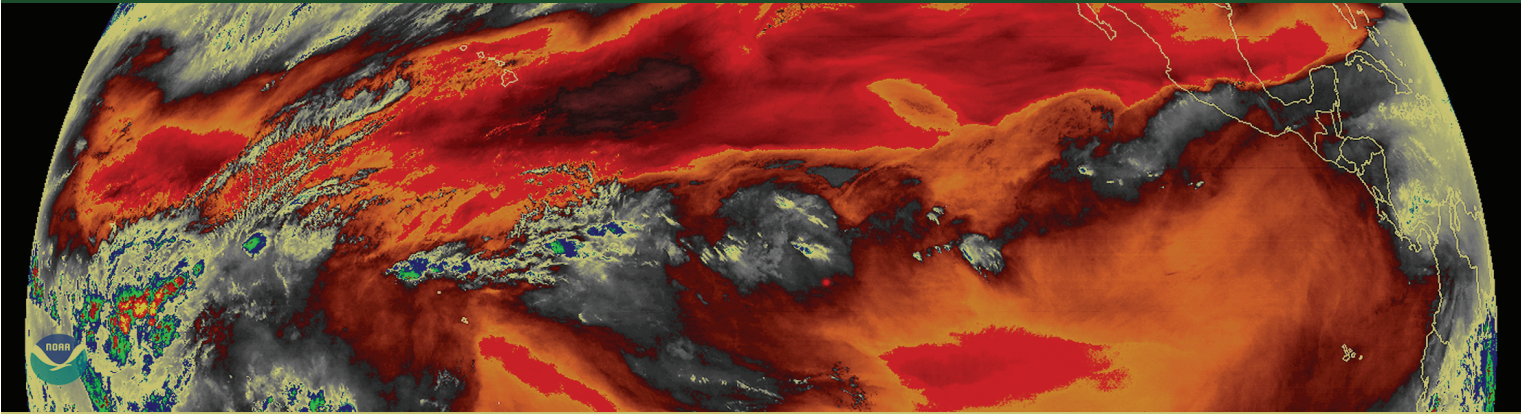


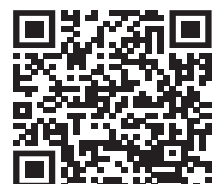
ENVIBAYES WORKSHOP ON COMPLEX ENVIRONMENTAL DATA



SEPTEMBER 18–20, 2023
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO

***Bringing together environmental
and ecological statisticians to share
cutting-edge advancements and
spark deeper knowledge gains***

<https://statistics.colostate.edu/envibayes-workshop>



ISBA



Monday, 18 September, 2023

8:30–9:00am	Continental breakfast	Lory Student Center (University Ballroom)
9:00am–noon	Short Course: Chris Paciorek	<i>Using and developing algorithms for hierarchical models with NIMBLE, including MCMC algorithms and Laplace approximation</i>
noon–1:00pm	Lunch	Not provided
1:00–1:15pm	Opening Remarks	Simon Tavener, Dean of CNS
1:15–2:15pm	Keynote I Veronica Berrocal	<i>Stationarity or non-stationarity: that is the question</i>
2:15–3:45pm	Session I Andee Kaplan Maryclare Griffin Doug Nychka	Fast computing and scalability <i>Improving Bayesian inference for streaming data</i> <i>Log-Gaussian Cox Process Modeling of Large Spatial Lightning Data using Spectral and Laplace Approximations</i> <i>Fast methods for conditional simulation, the key to spatial inference.</i>
3:45–4:00pm	Break	
4:00–5:30pm	Session II Claire Heffernan Michael Schwob Jorge Sicacha-Parada	Student paper winners <i>A dynamic spatial filtering approach to mitigate underestimation bias in field calibrated low-cost sensor air-pollution data</i> <i>Dynamic Population Models with Temporal Preferential Sampling to Infer Phenology</i> <i>New spatial models for integrating standardized detection-nondetection and opportunistic presence-only data: application to estimating risk factors associated to powerline-induced death of birds</i>
5:30–7:00pm	Poster session & Welcome reception	Lory Student Center (University Ballroom)

Tuesday, 19 September, 2023

8:00–8:30am	Continental breakfast	Lory Student Center (University Ballroom)
8:30–9:30am	Keynote II	
	Amy Herring	<i>Low Rank Longitudinal Factor Regression</i>
9:30–11:00am	Session III	Satellites and Plants
	Maggie Johnson	<i>Tracking plant stress from space: Improving estimates of evapotranspiration through spatiotemporal data fusion</i>
	Henry Scharf	<i>Predicting fine-scale taxonomic variation in landscape vegetation using large satellite imagery data sets</i>
	Yawen Guan	<i>A Bayesian Hierarchical Approach for Modeling Tree Cover Change</i>
11:00–11:15am	Break	
11:15am–12:45pm	Session IV	Exposure and Environmental Health
	Kayleigh Keller	<i>Tropical cyclones and risk of preterm birth: distributed-lag non-linear models in a large-data, time-to-event framework</i>
	Ander Wilson	<i>Heterogeneous Distributed Lag Models to Estimate Personalized Effects of Maternal Exposures to Air Pollution</i>
	Cory Zigler	<i>Bayesian Causal Inference with Uncertain Physical Process Interference</i>
12:45–1:45pm	Lunch	Lory Student Center (University Ballroom)
1:45–6:00pm	Outing: Hike at Lory State Park	Lory Student Center (Lot 310)

Wednesday, 20 September, 2023

8:30–9:00am	Continental breakfast	Lory Student Center (University Ballroom)
9:00–10:00am	Keynote III	
	Amy Braverman	<i>Statistical Challenges for the Next Generation of NASA's Earth Observing Satellites</i>
10:00am–12:00pm	Session V	Machine Learning and Surrogate Models
	Robert Gramacy	<i>Contour Location for Reliability in Airfoil Simulation Experiments using Deep Gaussian Processes</i>
	Dorit Hammerling	<i>Methane emission detection, localization and quantification on oil and gas facilities using continuous monitoring sensors</i>
	Chenlu Shi	<i>Space-Filling Designs for Computer Experiments</i>
	Abhirup Datta	<i>Combining machine learning with Gaussian processes for geospatial data</i>
12:00–1:00pm	Lunch	Lory Student Center (University Ballroom)
1:00pm–3:00pm	Session VI	Extreme and Non-Gaussian Data
	Dan Cooley	<i>Transformed-Linear Methods for Extremes and Fire Season Attribution</i>
	Likun Zhang	<i>Flexible modeling of multivariate extremes with Bayesian networks</i>
	Matt Koslovsky	<i>A Bayesian model for measurement error in multinomial data</i>
	Alexandra Schmidt	<i>Temporal misalignment in geostatistical data</i>
3:00–3:15pm	Break	

3:15–5:15pm

Session VI

Human and Animal Movement

Ephraim Hanks

A Mixture of OU-processes Framework for Jointly Modeling Animal Movement and Species Distribution Data

Albert Orwa Akuno

Inference on a Multi-Patch Epidemic Model with Partial Mobility, Residency, and Demography: Case of the 2020 COVID-19 Outbreak in Hermosillo, Mexico

Vianey Leos Barajas

Incorporating physiology into the analysis of animal movement

Toryn Schafer

Inverse reinforcement learning for animal behavior in the environment

5:15–5:20

Closing Remarks

Abstracts

Keynotes:

Veronica Berrocal

University of California, Irvine

Stationarity or non-stationarity: that is the question

A typical spatial analysis always starts with a decision: how to model the covariance function? A convenient assumption, in some cases warranted by the data, is to assume that the spatial random field is stationary. With spatial statistics handling more and more complex data, there might be a need to revisit this almost automatic assumption. In this talk, we will discuss some of our work in this arena and will insert into a wider lens of methods proposed to handle non-stationarity in spatial data.

Amy Herring

Duke University

Low Rank Longitudinal Factor Regression

A canonical problem in applied statistics is determining the association between a collection of variables measured over time and a subsequent outcome of interest. However, when such measurements are moderate- to high-dimensional and strongly correlated both within and across measurement times, determining the joint relationship with an outcome is challenging, leading researchers often to resort to conducting separate analyses for each variable and/or time. In this work, we develop a novel approach to this problem motivated by studying the long-term health effects of exposure to environmental toxins over the course of pregnancy and early childhood. Our approach handles highly correlated longitudinal exposures using a Bayesian dynamic factor model and presents a novel factor regression approach for the outcome of interest. The regression is set up so that it can collapse on simpler and intuitive submodels when appropriate, while expanding to a quite general quadratic regression on all measured exposures when supported by the data. After demonstrating our model's effectiveness in simulations, we present an application to data from the ELEMENT study.

Amy Braverman
Jet Propulsion Laboratory

Statistical Challenges for the Next Generation of NASA's Earth Observing Satellites

In this talk I will discuss the generation of NASA's Earth-orbiting satellites: the Earth System Observatory (ESO). ESO is made of four new missions, one of which is Surface Biology and Geology (SBG). I will focus on SBG since it is in the initial stages of development, and it is the ESO mission for which we statisticians are already significantly involved. After introducing the mission, its science objectives, and data collection, I will delineate three kinds of statistical challenges: science analysis, data processing, and uncertainty quantification.

Listed in order of the session schedule.

Andee Kaplan

Colorado State University

Improving Bayesian inference for streaming data

In the streaming data setting, where data arrive continuously or in frequent batches and there is no pre-determined amount of total data, Bayesian models can employ recursive updates, incorporating each new batch of data into the model parameters' posterior distribution. Filtering methods are currently used to perform these updates efficiently, however, they suffer from eventual degradation as the number of unique values within the filtered samples decreases. We present Generative Filtering, a method for efficiently performing recursive Bayesian updates in the streaming setting that retains the speed of a filtering method while using parallel updates to avoid degenerate distributions after repeated applications. We show that this sampler improves upon existing methods and state conditions under which its convergence is faster. Additionally, investigate the alleviation of filtering degradation through simulation and Ecological species count data.

Maryclare Griffin

University of Massachusetts, Amherst

Log-Gaussian Cox Process Modeling of Large Spatial Lightning Data using Spectral and Laplace Approximations

Lightning is a destructive and highly visible product of severe storms, yet there is still much to be learned about the conditions under which lightning is most likely to occur. The GOES-16 and GOES-17 satellites, launched in 2016 and 2018 by NOAA and NASA, collect a wealth of data regarding individual lightning strike occurrence and potentially related atmospheric variables. The acute nature and inherent spatial correlation in lightning data renders standard regression analyses inappropriate. Further, computational considerations are foregrounded by the desire to analyze the immense and rapidly increasing volume of lightning data. We present a new computationally feasible method that combines spectral and Laplace approximations in an EM algorithm, denoted SLEM, to fit the widely popular log-Gaussian Cox process model to large spatial point pattern datasets. In simulations, we find SLEM is competitive with contemporary techniques in terms of speed and accuracy. When applied to two lightning datasets, SLEM provides better out-of-sample prediction scores and quicker runtimes, suggesting its particular usefulness for analyzing lightning data, which tend to have sparse signals.

Doug Nychka
Colorado School of Mines

Fast methods for conditional simulation, the key to spatial inference.

An advantage of a Gaussian process (GP) model for surface fitting is obtaining the companion estimates of the function's uncertainty. The standard method for assessing uncertainty of a GP estimate is through conditional simulation, a Monte Carlo sampling algorithm of the multivariate Gaussian distribution. Moreover, the conditional sampling step is important because it is common to both a frequentist and a Bayesian analysis. Conditional simulation is a powerful tool, for example allowing for Monte Carlo based uncertainty on surface contours (level sets), a difficult and nonlinear inference problem. This algorithm, however, has two bottlenecks: generating spatial predictions on large, but regular grids and also simulation of a Gaussian process on both a large regular grid and at irregular locations. Accurate approximations are proposed that allow for fast computation of both these steps. The computational efficiency is achieved by relying on the fast Fourier transform for 2D convolution and also sparse matrix multiplication. Under common spatial applications a speedup by a factor from 10 to a 100 or more is obtained and makes it possible to determine uncertainty of GP estimates on a laptop and in often an interactive setting. Besides the practical benefits of this speedup their accuracy are examples of the "screening effect" for spatial prediction and are related to the errors bounds in interpolation when the GP is related to an element in a reproducing kernel Hilbert space. See Bailey, Maggie D., Soutir Bandyopadhyay, and Douglas Nychka. "Adapting conditional simulation using circulant embedding for irregularly spaced spatial data." *Stat* 11.1 (2022): e446.

Claire Heffernan
Johns Hopkins University

A dynamic spatial filtering approach to mitigate underestimation bias in field calibrated low-cost sensor air-pollution data

Low-cost air pollution sensors, offering hyper-local characterization of pollutant concentrations, are becoming increasingly prevalent in environmental and public health research. However, low-cost air pollution data can be noisy, biased by environmental conditions, and usually need to be field-calibrated by collocating low-cost sensors with reference-grade instruments. We show, theoretically and empirically, that the common procedure of regression-based calibration using collocated data systematically underestimates high air-pollution concentrations, which are critical to diagnose from a health perspective. Current calibration practices also often fail to utilize the spatial correlation in pollutant concentrations. We propose a novel spatial filtering approach to collocation-based calibration of low-cost networks that mitigates the underestimation issue by using an inverse regression. The inverse-regression allows incorporating spatial correlations by a second-stage model for the true pollutant concentrations using a conditional Gaussian Process. Our approach works with one or more collocated sites in the network and is dynamic, leveraging spatial correlation with the latest available reference data. We demonstrate in simulations how the spatial filtering substantially improves estimation of pollutant concentrations, and measures peak concentrations with greater accuracy. We apply the methodology for calibration of a low-cost PM_{2.5} network in Baltimore, Maryland, and diagnose air-pollution peaks that are missed by the regression-calibration.

Michael Schwob
The University of Texas at Austin

Dynamic Population Models with Temporal Preferential Sampling to Infer Phenology

To study population dynamics, ecologists and wildlife biologists typically use relative abundance data, which may be subject to temporal preferential sampling. Temporal preferential sampling occurs when the times at which observations are made and the latent process of interest are conditionally dependent. To account for preferential sampling, we specify a Bayesian hierarchical abundance model that considers the dependence between observation times and the ecological process of interest. The proposed model improves relative abundance estimates during periods of infrequent observation and accounts for temporal preferential sampling in discrete time. Additionally, our model facilitates posterior inference for population growth rates and mechanistic phenometrics. We apply our model to analyze both simulated data and mosquito count data collected by the National Ecological Observatory Network. In the second case study, we characterize the population growth rate and relative abundance of several mosquito species in the *Aedes* genus.

Jorge Sicacha-Parada
Norwegian Computing Center

New spatial models for integrating standardized detection-nondetection and opportunistic presence-only data: application to estimating risk factors associated to powerline-induced death of birds

The constant increase in energy consumption has created the necessity of extending the energy transmission and distribution network. Placement of powerlines represent a risk for bird population. Hence, better understanding of deaths induced by powerlines, and the factors behind them are of paramount importance to reduce the impact of powerlines. To address this concern, professional surveys and citizen science data are available. While the former data type is observed in small portions of the space by experts through expensive standardized sampling protocols, the latter is opportunistically collected over large extensions by citizen scientists.

In this paper we set up full Bayesian spatial models that 1) fusion both professional surveys and citizen science data and 2) explicitly account for preferential sampling that affects professional surveys data and for factors that affect the quality of citizen science data. The proposed models are part of the family of latent Gaussian models as both data types are interpreted as thinned spatial point patterns and modeled as log-Gaussian Cox processes. The specification of these models assume the existence of a common latent spatial process underlying the observation of both data types.

The proposed models are used both on simulated data and on real-data of powerline-induced death of birds in the Trøndelag county in Norway. The simulation studies clearly show increased accuracy in parameter estimates when both data types are fusioned and factors that bias their collection processes are properly accounted for. The study of powerline-induced deaths shows a clear association between the density of the powerline network and the risk that powerlines represent for bird populations. The choice of model is relevant for the conclusions that could be drawn from this case study as different models estimated the association between risk of powerline-induced deaths and the amount of exposed birds differently.

Maggie Johnson
Jet Propulsion Laboratory

Tracking plant stress from space: Improving estimates of evapotranspiration through spatiotemporal data fusion

Evapotranspiration (ET) has been identified as a most important science measurement to better understand the impacts of water loss, for drought prediction, and for agricultural water management. Estimation of ET requires near daily and high spatial resolution (<100m) satellite-derived surface reflectance imagery to characterize highly heterogeneous vegetated surfaces. However, no single, non-commercial space mission currently provides such data due to limitations of the spatial, temporal and spectral resolutions of individual instruments. In this talk, we propose a scalable, spatiotemporal data fusion methodology to combine measurements from multiple remote sensing instruments to produce daily, high spatial resolution surface reflectance products with associated uncertainty estimates. Space-time dynamic linear models are used to leverage spatial and temporal dependence for gap-filling between high resolution images, and a local Kalman filter/smoothing is implemented to facilitate processing of billions of measurements on regional to global scales. Finally, we illustrate the impact of the fused products on resulting ET estimates.

Henry Scharf
University of Arizona

Predicting fine-scale taxonomic variation in landscape vegetation using large satellite imagery data sets

Accurate information on the distribution of vegetation species is used as a proxy for the health of an ecosystem, a currency of international environmental treaties, and a necessary planning tool for forest preservation and rehabilitation, to name just a few of its applications. However, direct, extensive observation of vegetation across large geographic regions can be very expensive. The extensive coverage and high temporal resolution of remote sensing data collected by satellites like the European Space Agency's Sentinel-2 system could be a critical component of a solution to this problem.

We propose a hierarchical model for predicting vegetation cover that incorporates high resolution satellite imagery, landscape characteristics such as elevation and slope, and direct observation of vegetation cover. Besides providing model-based predictions of vegetation cover with accompanying uncertainty quantification, our proposed model offers inference about the effects of landscape characteristics on vegetation type. Implementation of the model is computationally challenging due to the volume and spatial extent of data involved. Thus, we propose an efficient, approximate method for model fitting that is able to make use of all available observations. We demonstrate our approach with an application to the distribution of three post-fire resprouting deciduous species in the Jemez Mountains of New Mexico.

Yawen Guan

Colorado State University

A Bayesian Hierarchical Approach for Modeling Tree Cover Change

Invasive trees are able to survive in harsh conditions and spread quickly across the landscape. In Nebraska, Eastern Redcedar (ERC) is designated as one of the invasive tree species. However, ERC field observations are scarce and there is no wide-spatial coverage historical records available for monitoring encroachment. Instead, we analyze tree cover estimates derived from Landsat satellite images for understanding tree spread. A Bayesian hierarchical approach with an integration of a mechanistic model is proposed to understand the spatiotemporal dynamics of encroachment. The underlying dynamic process is modeled by diffusion-advection equations which explicitly describe the evolution of tree spread and its association with environmental factors. The proposed approach is applied to the Sandhill region in Nebraska.

Kayleigh Keller

Colorado State University

Tropical cyclones and risk of preterm birth: distributed-lag non-linear models in a large-data, time-to-event framework

Tropical cyclone (TC) related exposures are a severe and recurring threat, causing infrastructure damage and directly threatening human health. Much of the health impacts of TCs, including the potential increased risk of adverse perinatal outcomes associated with these exposures, is not well characterized. In this study we examine the relationship between TC-related wind and flood exposure and risk of preterm birth in over 2 million singleton live births recorded in North Carolina between 1996 and 2017. We address two methodological issues arising in this context: (i) evaluating the impact of different spatial scales of exposure precision and (ii) allowing for flexible relationships with exposure timing and intensity through non-linear distributed lag Cox regression models. Results show evidence of increased risk of preterm birth with TC exposures, most strongly in the middle of second trimester, that is consistent across different degrees of exposure precision.

Ander Wilson

Colorado State University

Heterogeneous Distributed Lag Models to Estimate Personalized Effects of Maternal Exposures to Air Pollution

Children’s health studies support an association between maternal environmental exposures and children’s birth outcomes. A common goal is to identify critical windows of susceptibility—periods during gestation with increased association between maternal exposures and a future outcome. The timing of the critical windows and magnitude of the associations are likely heterogeneous across different levels of individual, family, and neighborhood characteristics. Using an administrative Colorado birth cohort we estimate the individualized relationship between weekly exposures to fine particulate matter (PM2.5) during gestation and birth weight. To achieve this goal, we propose a statistical learning method combining distributed lag models and Bayesian additive regression trees to estimate critical windows at the individual level and identify characteristics that induce heterogeneity from a high-dimensional set of potential modifying factors. We find evidence of heterogeneity in the PM2.5–birth weight relationship, with some mother-child dyads showing a 3 times larger decrease in birth weight for an IQR increase in PM2.5 exposure compared to the population average. Specifically, we find increased vulnerability for non-Hispanic mothers who are either younger, have higher body mass index or lower educational attainment. Our case study is the first precision health study of critical windows.

Cory Zigler

The University of Texas at Austin

Bayesian Causal Inference with Uncertain Physical Process Interference

Causal inference with spatial environmental data is often challenging due to the presence of interference: outcomes for observational units depend on some combination of local and non-local treatment. This is especially relevant when estimating the effect of power plant emissions controls on population health, as pollution exposure is dictated by (i) the location of point-source emissions, as well as (ii) the transport of pollutants across space via dynamic physical-chemical processes. In this work, we estimate the effectiveness of air quality interventions at coal-fired power plants in reducing two adverse health outcomes in Texas in 2016: pediatric asthma ED visits and Medicare all-cause mortality. We develop methods for causal inference with interference when the underlying network structure is not known with certainty and instead must be estimated from ancillary data. We offer a Bayesian, spatial mechanistic model for the interference mapping which we combine with a flexible non-parametric outcome model to marginalize estimates of causal effects over uncertainty in the structure of interference. Our analysis finds some evidence that emissions controls at upwind power plants reduce asthma ED visits and all-cause mortality, however accounting for uncertainty in the interference renders the results largely inconclusive.

Robert Gramacy

Virginia Tech

Contour Location for Reliability in Airfoil Simulation Experiments using Deep Gaussian Processes

Bayesian deep Gaussian processes (DGPs) outperform ordinary GPs as surrogate models of complex computer experiments when response surface dynamics are non-stationary, which is especially prevalent in aerospace simulations. Yet DGP surrogates have not been deployed for the canonical downstream task in that setting: reliability analysis through contour location (CL). Level sets separating passable vs. failable operating conditions are best learned through strategic sequential design. There are two limitations to modern CL methodology which hinder DGP integration in this setting. First, derivative-based optimization underlying acquisition functions is thwarted by sampling-based Bayesian (i.e., MCMC) inference, which is essential for DGP posterior integration. Second, maximal uncertainty acquisition criteria, such as entropy, are famously myopic to the extent that optimization may even be undesirable. Here we tackle both of these limitations at once, proposing a hybrid criteria that explores along the Pareto front of entropy and (predictive) uncertainty requiring evaluation only at strategically located “triangulation” candidates. We showcase DGP CL performance in several synthetic benchmark exercises and on a real-world RAE-2822 transonic airfoil simulation.

Dorit Hammerling

Colorado School of Mines

Methane emission detection, localization and quantification on oil and gas facilities using continuous monitoring sensors

Methane, the main component of natural gas, is the second-largest contributor to climate change after carbon dioxide. Methane has a higher heat-trapping potential but shorter lifetime than carbon dioxide, and therefore, rapid reduction of methane emission can have quick and large climate change mitigation impacts. Reducing emissions from oil and gas production facilities, which account for approximately 14% of total methane emissions, turns out to be a particular promising avenue due to the rapid development in continuous emission monitoring technology. We present a statistical framework for quick emission detection, localization and quantification using continuous methane concentration data measured by multiple monitoring sensors on oil and gas production facilities, and show its performance in a test setting using controlled release data where we know the emission rates. We also demonstrate its effectiveness under real-world conditions and discuss ideas for future directions, including a Bayesian model for multi-source scenarios.

Chenlu Shi

Colorado State University

Space-Filling Designs for Computer Experiments

Computer models are powerful tools used to study complex systems from almost every field in natural and social sciences. However, the complexity of the computer model results in a high computational cost for the investigation. This issue calls for computer experiments that aim at building a statistical surrogate model based on a set of data generated by running computer models. Space-filling designs are the most accepted designs for computer experiments. A broad introduction to space-filling designs will be given in this talk. Thanks to the guaranteed space-filling property in the low dimensional projections of the input space, a class of space-filling designs, so-called general strong orthogonal arrays, is appealing. We will introduce this class of designs and show their usefulness in computer experiments in the talk.

Abhirup Datta

Johns Hopkins University

Combining machine learning with Gaussian processes for geospatial data

Spatial generalized linear mixed-models, consisting of a linear covariate effect and a Gaussian Process (GP) distributed spatial random effect, are widely used for analyses of geospatial data. We consider the setting where the covariate effect is non-linear and propose modeling it using a flexible machine learning algorithm like random forests or deep neural networks. We propose well-principled extensions of these methods, for estimating non-linear covariate effects in spatial mixed models where the spatial correlation is still modeled using GP. The basic principle is guided by how ordinary least squares extends to generalized least squares for linear models to account for dependence. We demonstrate how the same extension can be done for these machine learning approaches like random forests and neural networks. We provide extensive theoretical and empirical support for the methods and show how they fare better than naïve or brute-force approaches to use machine learning algorithms for spatially correlated data.

Dan Cooley
Colorado State University

Transformed-Linear Methods for Extremes and Fire Season Attribution

In this talk, we will introduce transformed linear methods for extremes. By using the tail pairwise dependence matrix (TPDM) in place of the covariance matrix, and by employing transformed linear operations, extreme analogues can be developed for familiar linear statistical methods.

Here, we will focus on developing transformed linear time series models to capture dependence in the upper tail. These models are extremal analogues to familiar ARMA models. We apply these models to perform attribution for seasonal wildfire conditions. To focus on change in fire risk due to climate, we model the fire weather index (FWI) time series. We use our fitted model to perform an attribution study. According to our fitted model, the 2020 Colorado fire season is many times more likely to occur under recent climate than under the climate of 50 years ago.

Likun Zhang
University of Missouri

Flexible modeling of multivariate extremes with Bayesian networks

When extremal behavior is observed in environmental and financial processes, it is usually observed in multiple processes simultaneously. Although past and current work discusses modeling such multivariate extremal dependence, they mostly focused on the limiting max-stable and generalized Pareto models that suffer inflexible dependence structure. In this work, we propose a scale mixture model of marginally transformed multivariate Gaussian distributions, where the scaling factors can be either heavy- or light-tailed, depending on parameters that are estimated from the data. We build in additional flexibility by incorporating a latent directed network structure in the scaling factors to represent causal relationships among different variables. The interplay between the scaling parameters and the directed graph permits smooth transitions between asymptotic dependence and asymptotic independence for any pair of components of the multivariate distribution. We build a standard adaptive Metropolis algorithm for model fitting and perform extensive simulation studies to validate its ability to infer important dependence parameters and the network adjacency matrix from data.

Matt Koslovsky
Colorado State University

A Bayesian model for measurement error in multinomial data

Measurement error in multinomial data is a well-known and well-studied inferential problem that is encountered in many fields, including engineering, biomedical and omics research, ecology, finance, and social sciences. In wildlife monitoring studies, measurement error, often referred to as imperfect detection, occurs when a species that is present goes undetected (i.e., false negative) and/or an observed individual is misclassified as the wrong species (i.e., false positive). Motivated by multispecies monitoring data collected in ecological research, we provide a unified framework for accommodating both forms of measurement error using a Bayesian hierarchical approach. We demonstrate the proposed method's performance on simulated data and apply it to acoustic bat monitoring data.

Alexandra Schmidt
McGill University

Temporal misalignment in geostatistical data

Due to the high costs of monitoring environmental processes, measurements might be taken at different temporal scales. When observations are available at different temporal scales across different spatial locations, we name it temporal misalignment. Rather than aggregating the data and modeling it at the coarse scale, we propose a model that simultaneously accounts for observations at the fine and coarse temporal scales. More specifically, we propose a spatiotemporal model that accounts for the temporal misalignment when one of the scales is the sum or average of the other. We discuss two cases: when observations at the fine scale follow a normal distribution, and the case when observations are either positive or equal to zero. In the latter, we propose a hurdle gamma model at the fine scale. The proposed models allow the estimation of the process at the fine scale only when coarse measurements are available. The motivating example consists of measurements of different types of pollen concentration across Toronto, Canada. The data were recorded daily for some sites and weekly for others. This is joint work with Sara Zapata-Marin.

Ephraim Hanks

Pennsylvania State University

A Mixture of OU-processes Framework for Jointly Modeling Animal Movement and Species Distribution Data

Joint modeling of individual animal telemetry data and animal species distribution data requires models that can be scaled from the individual to population scales. This has typically been done through scaling individual models to their differential equation limits, but this leads to computationally challenging models for inference. We present a novel framework for scaling individual models to the population level through a mixture of Ornstein Uhlenbeck processes, with each individual animal moving based on an OU-process with individual parameters, and the population of animals defined by a hierarchical mixture. We show that if this mixture is multivariate Gaussian, that the population-level dynamics are analytically tractable, with fully deterministic temporal dynamics which are computationally efficient to compute. This provides a straightforward probabilistic framework for joint analysis of telemetry and species distribution data. We illustrate this approach through a joint analysis of golden eagle telemetry data and eBird species distribution data in western North America. We show that our joint approach allows for more accurate estimation of movement parameters, and better identifiability of population dynamics than would be possible through individual analyses of either data stream alone. We show how this approach can be used to identify the spatio-temporal landscape of potential risks to migratory birds, as well as clarify spatio-temporal regions of most import when managing migratory species.

Albert Orwa Akuno

Centro de Investigación en Matemáticas, A.C.

Inference on a Multi-Patch Epidemic Model with Partial Mobility, Residency, and Demography: Case of the 2020 COVID-19 Outbreak in Hermosillo, Mexico

Most studies modeling population mobility and the spread of infectious diseases, particularly those using meta-population multi-patch models, tend to focus on the theoretical properties and numerical simulation of such models. As such, there is relatively scant literature focused on numerical fit, inference, and uncertainty quantification of epidemic models with population mobility. In this research, we use three estimation techniques to solve an inverse problem and quantify its uncertainty for a human-mobility-based multi-patch epidemic model using mobile phone sensing data and confirmed COVID-19-positive cases in Hermosillo, Mexico. First, we utilize a Brownian bridge model using mobile phone GPS data to estimate the residence and mobility parameters of the epidemic model. In the second step, we estimate the optimal model epidemiological parameters by deterministically inverting the model using a Darwinian-inspired evolutionary algorithm (EA)—that is, a genetic algorithm (GA). The third part of the analysis involves performing inference and uncertainty quantification in the epidemic model using two Bayesian Monte Carlo sampling methods: t-walk and Hamiltonian Monte Carlo (HMC). The results demonstrate that the estimated model parameters and incidence adequately fit the observed daily COVID-19 incidence in Hermosillo. Moreover, the estimated parameters from the HMC method yield large credible intervals, improving their coverage for the observed and predicted daily incidences. Furthermore, we observe that the use of a multi-patch model with mobility yields improved predictions when compared to a single-patch model.

Vianey Leos Barajas

University of Toronto

Incorporating physiology into the analysis of animal movement

A long-sought goal in ecology is to connect movement with population dynamics. Especially for ungulates, there is a known link between condition (e.g. fat reserves) and the probability of survival and reproduction. Assuming a particular genetic makeup and physiology, condition reflects the history of behavioral decisions, including movement and habitat use. However, the condition of an animal can also have a direct implication on the types of movements that it performs and the habitats that it visits. Using Merino sheep as a case study, we present a model that allows for interaction of movement and condition over time. For the movement dynamics, we use a discrete-time, finite-state hidden Markov models (HMMs) with the positional data of the sheep serving as the observation process and the underlying state process serving as a proxy for behaviors of interest. To incorporate condition as a potential covariate affecting the movement, and thus behavioral, process, we make use of physiological equations that describe the evolution of body fat in order to predict daily values of the condition process, that are typically recorded once a month. The physiological equations are expressed as a function of the states inferred by HMM, as well as the distance that the sheep travels.

Toryn Schafer
Texas A&M University

Inverse reinforcement learning for animal behavior in the environment

Animal movement is a complex spatio-temporal process due to interactions of environmental and social cues. The strength of the cues in driving movement varies among life history strategies. We analyze varying types of animal movement processes with the linearly-solvable Markov decision process (LMDP) from optimal control and inverse reinforcement learning. Instead of making simplifying assumptions across all anticipated scenarios, inverse reinforcement learning estimates the short-term rules governing long term behavior policies by using properties of a Markov decision process. Simulations illustrate trial and error learning by agents (animals) with memory and complex interactions between the environment and agents. Inference under the LMDP is done in a Bayesian framework with approximate inference obtained by variational inference in large state space settings.

Posters

- Mahshid Ahmadian (Virginia Commonwealth University)
- Renato Berlinghieri (Massachusetts Institute of Technology)
- David Burt (Massachusetts Institute of Technology)
- Danielle Demateis (Colorado State University)
- Lane Drew (Colorado State University)
- Lucas Godoy (University of Connecticut)
- Matthew Hofkes (Colorado School of Mines)
- Gabriel Huerta (Sandia National Laboratory)
- Seongwon Im (Colorado State University)
- Kevin Korsurat (Colorado State University)
- Vijay Kumar (Columbia University)
- Liz Lawler (Colorado State University)
- Wyatt Madden (Emory University)
- Conor Osborne (University of Edinburgh)
- Maddie Rainey (Colorado State University)
- Andrew Roberts (Boston University)
- Braden Scherting (Duke University)
- Muyang Shi (Colorado State University)
- Troy Wixson (Colorado State University)
- Wilson Wright (Colorado State University)

Organizing Committee

Andee Kaplan, Colorado State University (chair)
Ben Shaby, Colorado State University

Organizational Liaison

Martin Sweeny, Colorado State University

Financial Support Provided by

National Science Foundation
CSU College of Natural Sciences
CSU Department of Statistics