its source location due to long-range transport of pollutants.

This study focused on the influences that increased wildfires have on the air quality in Northern Minnesota. Particulate measurements made at Voyageurs National Park, Minnesota, by the Interagency Monitoring of Protected Visual Environments program were used. Average concentrations of elemental carbon (EC), organic carbon (OC), and $PM_{2.5}$ were compared between two periods: 2000–2004 (Period 1) and 2017–2021 (Period 2). During the peak of the wildfire season, in July, EC, OC, and $PM_{2.5}$ show an increase of 32.2%, 43.8%, and 17.8% respectively from Period 1 to Period 2. While each chemical species has multiple anthropogenic and natural sources, they are all associated with wildfires.

Case Study: July 17, 2021, was found to have a clear sky likely affected by smoke plumes based on visible satellite imagery. Enhancement ratios of EC, OC, and $PM_{2.5}$, calculated per an academic publication, were found to be 1.1397, 1.998, and 1.165, respectively.

Back-trajectories of the sampled air, calculated using the Hybrid Single-Particle Lagrangian Integrated Trajectory modeling system, corroborate the transport of air from areas of active wildfires, north of the site.

Protected observational data was most reflective of long-range transport of pollutants to the region. The influence of increased air quality within the past two decades in Northern Minnesota is likely affected by wildfires in the Canadian forests.

Research advisor Gouri Prabhakar writes: "This work provides valuable insight into the influence of wildfires on the air quality of the midwestern United States. Emma Braun and Audrey Shirley performed the analyses independently and wrote this snapshot jointly. This case study will motivate further analyses of observational data in this region."

Model Selection Through Cross-Validation for Supervised Learning Tasks with Manifold Data

Student researcher: *Derek Brown, Senior*

K-fold cross-validation is a popular method used in machine learning for estimating the prediction accuracy of a model. The algorithm starts by partitioning the data into K-folds. The model is first trained on all but one fold, with its accuracy being evaluated on the left-out fold. This is repeated for each fold, with the overall model performance being evaluated by taking the average of each of these accuracies. Notably, each accuracy is correlated with the others since the training folds overlap. This correlation makes establishing large sample asymptotic theorems, such as the central limit theorem, for K-fold cross-validation difficult. Recent work has started to establish central limit theorems for K-fold cross-validation for real-valued random variables.

My research expands these theorems to include the case in which random variables represent angles. Similar to how a unit circle can be defined with a single angle, multiple simultaneously recorded angles lie on a higher dimensional space called a torus, which looks like a doughnut in three-dimensional space. Such spaces are of relevance in biology. For example, in biomechanics, data extracted from locomotion experiments involve joint angles, and in biochemistry, backbone dihedral angles of amino acid residues determine protein structures. Understanding K-fold cross-validation for angular valued variables will provide valuable tools to scientists wanting to use accurate machine learning algorithms in many different areas involving data with complex structures.

Research advisor Alessandro Maria Selvitella writes: "Derek's research aims at understanding the theoretical underpinnings of cross-validation, one of the most popular methods used in machine learning for determining the accuracy of an algorithm and for model selection. His work highlights how important it is to understand the large sample asymptotics of cross- validation for ensuring accurate uncertainty estimation."

Exploring the Evolution of Callose Synthase in Green Plants

Student researcher: *Giovanna Durante, Sophomore*

Callose is a β-1,3-glucan polymer that has been identified as a crucial player in the plant response to fungal and bacterial pathogens. This polysaccharide controls infections by creating a physical barrier at the plant cell

https://doi.org/10.7771/2158-4052.1585