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
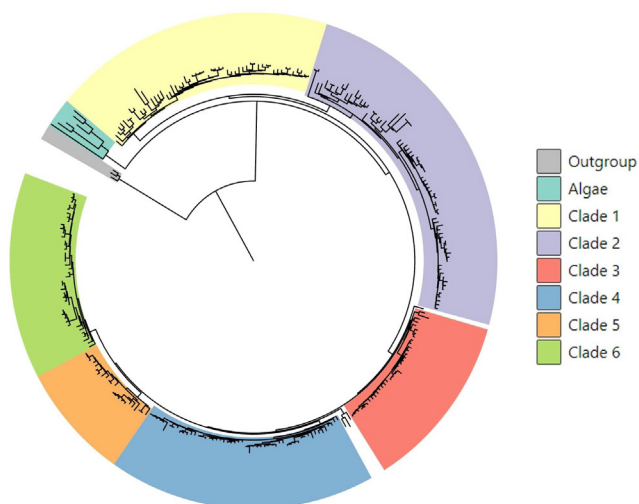
Phylogenetic reconstruction of the callose synthase gene family using iTOL.

wall, inhibiting the spread of infections into healthy plant tissue. Callose synthase is the enzyme responsible for the synthesis of this critical polymer.

Neither the evolution of callose synthase in the plant response to pathogens nor its modification across green plants has yet been investigated. This study explores the callose synthase enzyme family across green plant species to understand how plant immune systems evolved throughout their 500-million-year evolutionary history.

To achieve this, 33 plant genomes representing different plant groups were analyzed. BLAST and HMMER searches, using callose synthase genes previously characterized in *Arabidopsis* as the query, identified 294 unique callose synthase proteins. Multiple sequence alignment using MAFFT revealed the relationship between these homologs. The resulting sequences were trimmed and used to build a phylogenetic tree using IQTree maximum-likelihood estimation.

This phylogeny showed that green algae form a clade by itself and that the remaining six callose synthase clades contain only land plant taxa, suggesting that the function of callose synthase diverged when plants evolved for land and became critical to their survival at that point. Additional changes among clades reflect significant modifications in plant pathogen defense strategies over time and identify likely targets of related proteins, which

can be further investigated to develop novel methods to enhance the innate pathogen response of plants.

Research advisor Chao Cai writes: "Gianna's work explores the evolutionary history of callose synthase, which is an important player in the process of plant response to infection. Identifying callose synthases that are responsible for disease resistance in a variety of plant species will help with the design of disease control strategies, especially in crop plants."

The Cross-Linguistic and Cross-Situational Association Between Accentedness and Its Impact as Rated by Speakers

Student researcher: *Makaila Groves, Senior*

Accentedness is the extent of a foreign accent in one's speech, which depends on how listeners perceive this speech to be different from their own variety. The scope of practice of speech-language pathologists (SLPs) includes providing services for accent modification to culturally diverse populations. Although SLPs are taught that accentedness is a language *difference*, not a *disorder*, training to provide accent-related services is not currently required, leaving clinicians to draw on their experience with disorders and assume that accent *modification* means accent *reduction*. Gaining understanding of the situation-specific ways in which accentedness can impact individuals would improve the SLPs' ability to serve a variety of diverse clients.

To determine the relationship between accentedness and its impact on speakers, 43 multilingual adults were asked to rate the consequences of their accentedness, from their own perspective, across a variety of situations (e.g., work, home, etc.). Responses were analyzed quantitatively by ranges of magnitude of accentedness and ranges of impact on the speakers' lives (positive, negative, or neutral) and examined within the context of languages involved.

As hypothesized, this study confirms that the impact of accentedness is not limited to the magnitude of the accent. It also relates to the languages involved, and it is context/situation specific. A qualitative analysis of open-ended answers provided details of the sociolinguistic and sociocultural considerations that may have