# Abstract

# Introduction

# Methods

## Data collection

We used historical data collected by O.A. Stevens and others to create a dataset of first flowering days (FFD) for 25 flowering plant species. The observations were made at a tallgrass prairie site in Clay county Minnesota that has been a Nature Conservancy preserve since 1975. Individual data points represent the day of the year on which a given plant species was observed flowering at the site although all species were not observed in all years. The Stevens dataset represents continuous data from 1910 to 1961 and subsequent observations are from 2012 through 2020. Thus, there is a 52-year gap in data at the end of the past century. The plant species analyzed were limited to those that met a series of minimum data requirements. The 25 species chosen had a minimum of five years of data and at least one observation prior to 1962 and one after.

In order to quantify different environmental variables related to annual climate patterns, we used daily climate data collected in Fargo, North Dakota, USA, as part of the National Atmospheric and Oceanic Administration (NOAA) National Climatic Data Center (NCDC) observing network (http://www.ncdc.noaa.gov/oa/ncdc.html). The climate data collection site (46 ° 56’ N, 96 ° 49’ W) is located at the Fargo International Airport 32 km west of the flowering observation site. The dataset includes daily estimates of maximum and minimum temperature, snowpack (0=bare ground) and snowfall beginning in 1942. However, snowpack data is unavailable for 1997 through 2004. As a result, we were able to analyze data for 29 years (1942-1961 and 2012-2020).

## Variables

We used the raw climate data to calculate four variables regarding seasonal patterns of temperature or winter snowfall. Accumulated Degree Growing Unit (AGDU) is an annual estimate of the accumulation of warm temperatures over the first three months of a year and is intended to quantify the relative warmth of spring for a given year. AGDU was calculated as the day of the year on which the sum of the growing units of the year exceeded 300. Growing units were defined as a daily measure of the difference between the average temperature and 35˚ F; units were set to zero if the average temperature was below 35. Thus, in years with warmer spring months the AGDU value will be lower.

Three different precipitation variables were calculated. The winter snowfall amount for a given year (TSNOW) was calculated as the sum of snowfall over the first 90 days. A second variable associated with snowfall was the Date of Bare Ground (DOBG) or the day of the year when snowpack first reached zero. A couple records indicated a short period, one to two days, of snowpack late in the season which were excluded for a more realistic representation of first bare ground. The third correlate of winter snowfall was Snowpack on Day X (SPDX), a variable designed to estimate the extent of snowpack just prior to the growing season. To calculate SPDX for each year we used model selection to identify which day in March represented the optimal day for best predicting the first flowering day (FFD) based on snowpack. The most predictive day was determined separately for each plant species. Once the most predictive day in March was determined for a given species by running individual linear regressions and choosing the model with the lowest AIC, snowpack on that day each year was assigned as the SPDX value. Larger SPDX values indicate greater snowpack on a selected day in March per species.

## Model development

We used the *lavaan* package in R for path analysis of the relationships between the variables previously mentioned. In our initial model, the exogenous variables were AGDU and TSNOW (Fig.1). The endogenous variables were DOBG, SPDX, and FFD. The model included regressions for each endogenous variable, variances within all variables, and residual covariances between the exogenous variables. We considered both direct and indirect regressions. To account for missing data points in the multiple regression model, we applied full information maximum likelihood (FIML) estimation. We used regression estimates for indirect and direct effects to interpret the relationships between latent variables in each of the species.

Since regressions between some variables were weak and not significant for many species, we examined other potential reduced models by eliminating one variable at a time. The reduced models excluded DOBG, SPDX, and AGDU consecutively. We used Akaike Information Criterion (AIC) to select the model that best represented the data.

# Results

## Variation in first flowering day (FFD)

We identified 24 flowering plant species in the Stevens Data set that met the criteria for analysis described in the methods. The first flowering day (FFD) varies extensively both among years within a species and among species. Median FFD varied across the species from a low of X to a high of Y and included early, mid, and late spring flowering species (Fig. X)

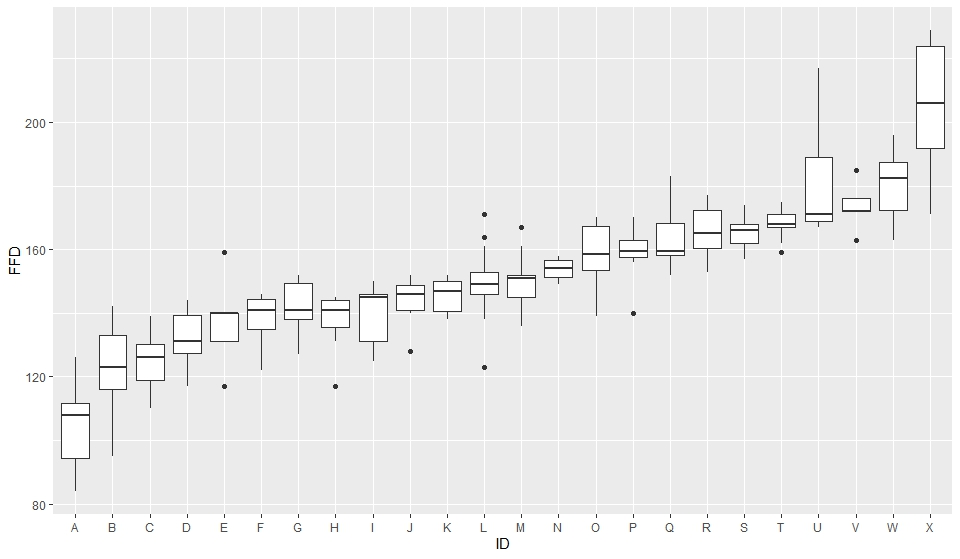


Figure X. Box plots of the first flowering day of 24 plant species from the Bluestem Prairie reserve in Clay county, MN. Observations were made between 1942-1961 and 2012-2020. Box plots indicate distribution quartiles and standard error bars. The species codes are as follows:A=Anemo paten;B=Ranun rhomb;C=Calth palus;D=Ceras arven;E=Ranun abort;F=Oxali viola;G=Sisyr angus;H=Litho canes;I=Trill cernu;J=Litho incis;K=Pedic canad;L=Zizia aurea;M=Vicia ameri;N=Cypri candi;O=Achil mille;P=Anemo canad;Q=Oxytr lambe;R=Rosa arkan;S=Penst grand;T=Penst graci;U=Campa rotun;V=Zigad elega;W=Amorp canes;X=Oenot nutta.

## Snowpack correlation

There were variable relationships between FFD and SPDX among species. R2 values ranged from 0.033 to 0.86 indicating a lot of variation in the explanatory power of SPDX on FFD. Three of 21 species were statistically significant (Cerastium arvense, Amorpha canescens, and Zigandenus elegans).

## Model selection

The model excluding DOBG was selected based on the AIC. This model had the lowest AIC in all species.

## Path analysis

Temperature- Many species had a significant regression coefficient for the relationship between AGDU and FFD. Nineteen species had positive regression coefficients, indicating that warmer temperatures earlier in the year led to earlier flowering. Sixteen of these species were significant. Five species had negative regression coefficients, of which only one was significant. The first four flowering species had strong and significant relationships between AGDU and FFD. Later flowering species typically had weaker, inconclusive relationships and few were significant.

Temperature and SPDX- The relationships between AGDU and SPDX were positive in all species and the coefficients ranged from 0.01 to 0.18. Only four species had a statistically significant relationship between these variables.

Snow – The relationship between TSNOW and SPDX had a positive and significant regression coefficient in all species. TSNOW was expected to be related to SPDX because both describe winter snowfall.

SPDX and FFD – Four out of the 21 species had a significant relationship between SPDX and FFD, two were negative and two were positive.

## A picture containing chart, polygon Description automatically generated

Figure 2. Correlations of FFD and SPDX for each species in order of early flowering species to later flowering species. R2 and p-values reported.

Table

Description automatically generated

Figure 3. Path diagrams with indirect effect estimates labeled. Number of asterisks indicate level of significance for p-value: \*p ≤ 0.05, \*\*p ≤ 0.01, \*\*\*p ≤ 0.001.

# Discussion

Temperature-

Temperature and SPDX-

Snow –

SPDX and FFD –