# Abstract

# Introduction

# Methods

## Data collection

We used historical data collected by O.A. Stevens and SET to create a dataset of first flowering days (FFD) for 25 flowering plant species. The observations were made at Bluestem Prairie (link), 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 (citation); 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 in this study were limited to those that met a series of minimum data requirements. The focal species had a minimum of five years of observations 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 climate 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 a total of 29 years (1942-1961 and 2012-2020).

## Climate Variables

We used the raw climate data to calculate four variables regarding annual patterns of temperature or winter snowfall. The first climate variable we calculated for each year was intended to quantify the relative warmth of the late winter/early spring season when the earliest flowering on the prairie is initiated. This variable, Accumulated Degree Growing Unit (AGDU), is was calculated as the day of the year on which the sum of the growing units of the year exceeded 300. Growing units are 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 ˚ F (citation). We chose 300 units as the cutoff because this number of units is typically accumulated by the end of March. Thus, in years with warmer spring months the AGDU value will be relatively lower and vice versa.

Three different winter precipitation variables were calculated. The winter snowfall amount for a given year (TSNOW) was calculated as the sum of daily snowfall over the first 90 days of a given year. A second variable associated with winter 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 variable associated with 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 species we used linear regression and model selection to identify which day in March represented the optimal day for best predicting the first flowering day (FFD) for that species from snowpack. The most predictive day was determined separately for each plant species. We ran separate linear regressions where FFD was the dependent variable and snowpack on day X was the independent variable for each day in March. AIC values were determined for each regression and the model associated with the lowest AIC value was chosen and used to assign the day in March consistently used for SPDX in that species. Thus, SPDX values increase with increase and decrease with increasing and decreasing snowpack on the selected day of March. Each of the four climate variables were not independent of each other.

## Model development

Our goal was to use Structural Equation Modelling (citation) to simultaneously assess the relationships between each of the four climate variables and the first flowering day (FFD) of the focal plant species, given the covarying nature of the climate variables. We used the *lavaan* package in R to incorporate path analysis and examine the relationships among the climate variables and the dependence of FFD on each of the climate variables individually. In our initial, full model, we included AGDU and TSNOW as exogenous variables and DOBG, SPDX, and FFD as endogenous variables (Fig. 1). The assumption was that FFD could have direct and indirect effects from both temperature (AGDU) and winter snowfall (TSNOW) through their indirect effects on snowpack in March (SPDX) and the date at which the snow melted (DOBG). The model included regressions for each endogenous variable, variances within all variables, and residual covariances between the exogenous variables (citations). We considered both direct and indirect regressions. To best compensate for missing data points over the course of the 29 years analyzed, we applied full information maximum likelihood (FIML) estimation to determine path coefficients and model statistics. We used regression estimates for indirect and direct effects to interpret the relationships between latent variables in each of the species.

In order to identify the best overall structural equation model for analyzing relationships among climate and flowering variables we used a model selection approach and compared the fit of the full model (above) to three other reduced models that omitted either DOBG, SPDX, and AGDU. After using the lavaan program to conduct path analyses of the three reduced models, we used Akaike Information Criterion (AIC) to select the model of the four that best represented the data based on the lowest AIC value.

# 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. None of the species were observed in every year of the survey; sample size by species ranged from X to Y. The first flowering day (FFD) varied 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)1)

## Model selection

## Model selection comparisons of AIC values among the three reduced models and the full model indicated that the best explanatory model was the reduced model which excluded DOBG indicating that the influence of temperature and snowfall on flowering date was relatively negligent through an indirect effect on when the ground first became bare of snow each spring.

Based on the chi squared statistic estimating goodness of fit of the reduced model for FFD of each species the model was a good representation of the relationships among the exogenous and endogenous variables for all but five species *(Anemone patens*, *Caltha palustris*, *Lithospermum canescens*, *Campanula rotundifolia*, and *Amorpha canescens)*. These five species were removed from further analysis and consideration.

The results of path analysis are presented in Figure 3 for each of the remaining species arranged by order of seasonal flowering sequence. The direct relationship between AGDU and FFD was significant in 12 out of 19 species analyzed suggesting an important role of temperature in determining flowering time for a majority of species. All twelve of species with significant AGDU effects had positive coefficients, indicating that warmer temperatures earlier in the year led to earlier flowering. The five first flowering species had strong and significant relationships between AGDU and FFD. Later flowering species typically had weaker, inconclusive relationships and few were significant. (say something here about general effects of AGDU on SPDX). Winter temperatures (AGDU) also had indirect effects on flowering time (FFD) through intermediary effects on the snowpack in March (SPDX) for 1 species (*Zigadenus elegans*).

The relationship between snowfall in the winter months (TSNOW) and snowpack in March (SPDX) was a predictably strong one. The path coefficient between the two variables was positive and significant for all species. TSNOW was expected to be related to SPDX because both describe winter snowfall. However, only three out of the 19 species had a significant relationship between SPDX and FFD. In one species (*Cypripedium candidum*) the path coefficient was negative indicating that relatively large amounts of winter snowfall led to earlier flowering relative to years when there was less winter snowfall. However in the other two species (*Zigadenus elegans* and *Rosa arkansana*) the coefficient was positive. There were significant indirect effects of TSNOW on FFD for the same three species through its effects on SPDX(Table 1).

# Discussion

This study…

The model for all species improved when we excluded DOBG. We expected that the date of first bare ground would influence first flowering day as was reported by Inouye (2002) for montane species in Colorado. However, only a few of the species had a significant relationship between DOBG and FFD suggesting that when the winter snow melts is not important for determining when plants begin flowering. In contrast, temperature seems to be a consistent determinant of flower timing.

AGDU and FFD had a positive and significant relationship in most of the species. A higher AGDU means a colder spring suggesting that temperature is important for growth and development. This was especially the case for earlier flowering species. These results mirror those of other studies of both upper Midwestern prairie plants (citations) and those in other temperate communities (citations).

Snowpack in March was largely unimportant, with the exception of three species. Two species (*Rosa arkansana* and *Zigadenus elegans*) had positive regression coefficients meaning the deeper the snow on day X in March, the later the species flowered. This outcome would be expected if snow cover impaired earlier flowering. One species has a negative regression coefficient suggesting that the amount of moisture from snow melting is important in determining the flowering of *Cypripedium candidum.* All three species had significant indirect effects between winter snowfall and first flowering date through depth of snow on day X. Interestingly, these three species flower later in the season, after June.

Overall, our results suggest that snowpack does not have a strong relationship with flower timing. Snowpack does not inhibit flowering because, even with late start growing plants can catch up. We expected snowpack to impact early flowering species and not later flowering species but, our results indicate that growth and flowering begin regardless of snow cover. However, air temperature has a much stronger effect on when plants flower. Even early flowering species

* + Interestingly, those three species all flowered later in the season. None of the species flowering before June were affected by snowpack. Why?
  + Air temperature in late spring is far bigger determinant of speed of growth

Further research

* More complete data sets, more species
* Other climatic conditions altered by climate change in this region- precipitation.

Table 1. Statistical summary of regression coefficients for direct and indirect effects for reduced model. AD = indirect effect for TSNOW on FFD mediated by SPDX, BD = indirect effect for AGDU on FFD mediated by SPDX. Number of asterisks indicate level of significance for p-value: \*p ≤ 0.05, \*\*p ≤ 0.01, \*\*\*p ≤ 0.001.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Species | A  SPDX~TSNOW | B  SPDX~AGDU | C  FFD~AGDU | D  FFD~SPDX | AD | BD |
| *Ranunculus rhomboides* | 0.471\*\* | 0.178 | 0.737\*\*\* | 0.295 | 0.139 | 0.053 |
| *Cerastium arvense* | 0.281\*\*\* | 0.066 | 0.427\*\*\* | 0.348 | 0.098 | 0.023 |
| *Ranunculus abortivus* | 0.508\*\*\* | 0.098 | 0.595\*\*\* | 0.679 | 0.345 | 0.066 |
| *Oxalis violacea* | 0.405\*\*\* | 0.139\* | 0.319\* | 0.146 | 0.059 | 0.020 |
| *Sisyrinchium angustifolium* | 0.281\*\*\* | 0.066 | 0.334\*\* | 0.405 | 0.114 | 0.027 |
| *Trillium cernuum* | 0.553\*\*\* | 0.056 | -0.211 | 1.288 | 0.712 | 0.072 |
| *Lithospermum incisum* | 0.553\*\*\* | 0.056 | 0.486\*\*\* | -0.061 | -0.033 | -0.003 |
| *Pedicularis canadensis* | 0.151\*\* | 0.083\* | -0.048 | 1.359 | 0.205 | 0.113 |
| *Zizia aurea* | 0.427\*\*\* | 0.102\* | 0.652\*\*\* | 0.307 | 0.131 | 0.031 |
| *Vicia americana* | 0.134\*\*\* | 0.081\*\* | 0.371 | -0.452 | -0.060 | -0.037 |
| *Cypripedium candidum* | 0.398\*\*\* | 0.074 | 0.113 | -0.438\* | -0.174\* | -0.032 |
| *Achillea millefolium* | 0.257\*\*\* | 0.036 | 0.529\*\* | 0.042 | 0.011 | 0.002 |
| *Anemone canadensis* | 0.398\*\*\* | 0.074 | 0.491\*\*\* | -0.092 | -0.037 | -0.007 |
| *Oxytre lambe* | 0.493\*\*\* | 0.008 | 0.133 | 0.573 | 0.283 | 0.004 |
| *Rosa arkansana* | 0.398\*\*\* | 0.074 | 0.364\*\* | 0.700\*\* | 0.279\*\* | 0.052 |
| *Penstemon grandifloras* | 0.438\*\*\* | 0.097 | -0.077 | 0.382 | 0.167 | 0.037 |
| *Penstemon gracilis* | 0.459\*\*\* | 0.106 | 0.241\* | 0.067 | 0.031 | 0.007 |
| *Zigadenus elegans* | 0.144\*\*\* | 0.091\*\* | 0.301\*\*\* | 0.467\*\*\* | 0.067\* | 0.043\* |
| *Oenothera nuttallii* | 0.340\* | 0.178 | -0.177 | 1.236 | 0.421 | 0.220 |

**FIGURE LEGENDS**

Figure 1. Box plots of the first flowering day (FFD) 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=Ranun rhomb;B=Ceras arven;C=Ranun abort;D=Oxali viola;E=Sisyr angus; F=Trill cernu;G=Litho incis;H=Pedic canad;I=Zizia aurea;J=Vicia ameri;K=Cypri candi;L=Achil mille;M=Anemo canad;N=Oxytr lambe;O=Rosa arkan;P=Penst grand;Q=Penst graci; R=Zigad elega; S=Oenot nutta.

Figure 2. Path diagrams with direct effect estimates labeled. Number of asterisks indicate level of significance for p-value: \*p ≤ 0.05, \*\*p ≤ 0.01, \*\*\*p ≤ 0.001.

Figure 3. Simple linear regressions of FFD as a function of SPDX for each species with best-fit lines based on least-square estimates. Results for species are organized in order of flowering sequence over the season from early spring to late summer

**FIGURES**

Chart, box and whisker chart

Description automatically generated

Figure 1.



Figure 2.

A picture containing diagram

Description automatically generated

Figure 3.