Final Project

RJ Cass

Table of contents I

Abstract

Tulips form a significant portion of the exports from the Netherlands. We want to understand what changes can be made to ensure tulip growth is not impacted by the changing climate.

We used a logistic model and found:

- Effect of chilling times on germination is dependent on species (some improve with longer chilling, some get worse with chilling)
- Ideal chilling time for each population (for most, range of 8-10 weeks)
- Predicted impact of the chilling time decreasing from 10 weeks to 9 (some do marginally better, but some do much worse)

Context

Tulip Production:

- Tulip products form 25% of agricultural exports from the Netherlands
- Changing climate puts the tulip industry at risk
- Want to understand how to adapt to these changes and protect the industry

Dataset of sample tulip growth populations:

- Year they were grown
- Number of weeks the bulbs were chilled
- Whether or not the bulb germinated
- Indices (can be removed from dataset)

Questions of Interest

We want to use the provided data to answer the following questions:

1. What is the effect of chilling time for the different species of tulips? Is it the same across the species? Which species are the same/different?
2. Is there an ideal chilling time for each species? If so, is it the same for all species?
3. Given climate change conditions, winters are expected to decrease from 10 to 9 weeks in the coming few years. What effect will this decrease in chilling time have on the probability of germination for each species? Is it the same for all species?

EDA - Population 12

None of population 12 germinated. We removed it from the dataset to not dilute the rest of the data

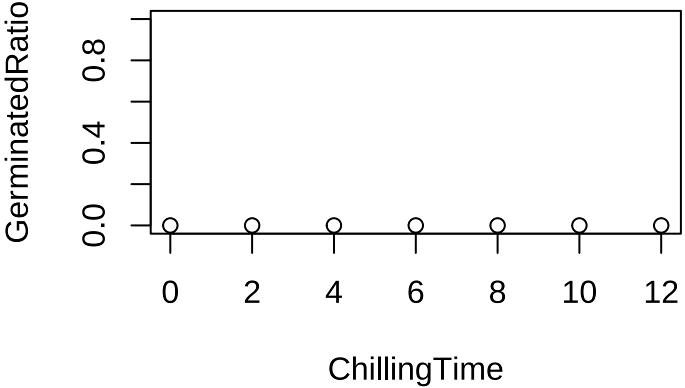


Figure 1: *Population 12 had a 0% germination rate across all chilling times*

EDA - Year

- Each population was tested in only 1 year (ie. no crossing with different years having an effect on one population)
- Physically, given testing conditions, we don't expect year to have an impact on germination
- In variable selection, Year was not important ($p > .05$)
- Removing year from models

EDA - Interactions - Population vs. Chilling

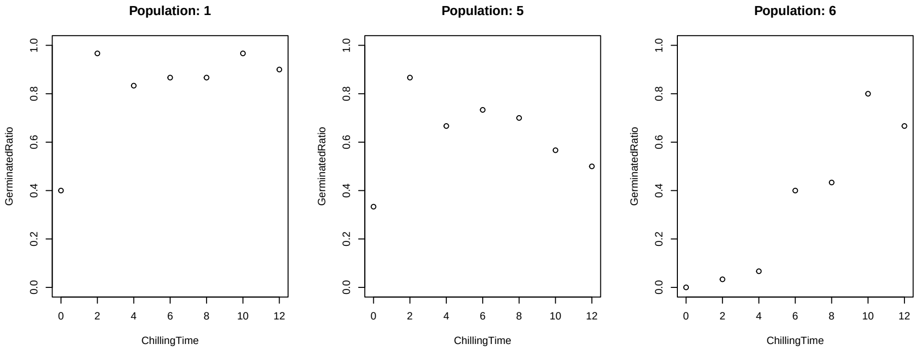


Figure 2: *Sample interaction plots of population and chilling. Some behave similarly, others are vastly different. Note the non-linearity of some populations.*

EDA - Summary

Removing from dataset:

- Population 12
- 'Year' variate

Need to account for the following:

- Interaction between Chilling Time and Population
  - Will need to ensure model includes interactions (either manually, or use a model that explores interactions)
  - If not included, resulting model will not capture the full impact of each variate
- Non-linearity of relationship between chilling time and germination rate
  - Will need to ensure the model handles non-linearity
  - If not included, model will not represent the correct relationship of this variate

Proposed Models - 1

Logistic Model

$$
\begin{aligned}
Y_n = & ln(\frac{p}{1-p}) = \beta_0 + \beta_1 X_{pop} + \beta_c poly(ChillingTime) \\
& + \beta_i (X_{pop} * poly(ChillingTime))
\end{aligned} \quad (1)
$$

- Accounts for interactions of population and year on ChillingTime
- Accounts for non-linearity of ChillingTime

Strengths:

- Captures interactions
- The concept of logistic (change in log-odds) is realtively interpretable

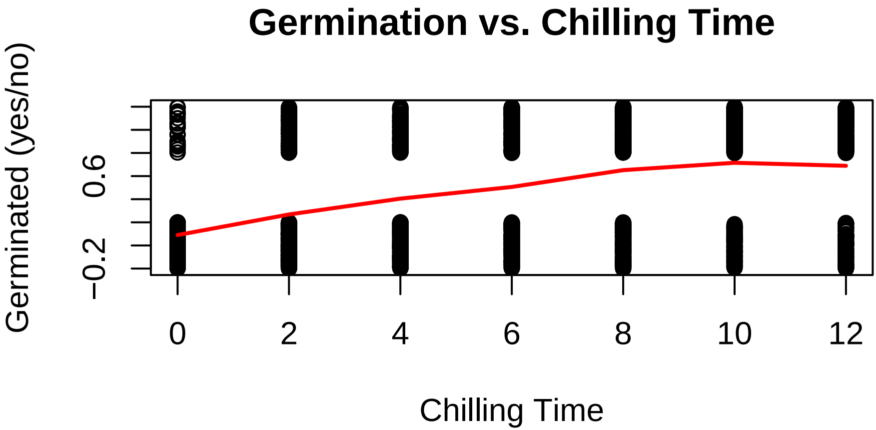
Weaknesses:

- Using splines loses interpretability

Proposed Models - 1 - Cont'd

Assumptions - Indepndence, Monotonocity

- Independence: Assumed due to the design of the experiment
- Monotonicty



Proposed Models - 2

Random Forest

Strengths:

- Relatively explainable (lots of trees, each tree gets a vote, average the votes, compare to cutoff)
- No inherent assumptions (besides the data being 'good')

Weaknesses:

- Can be prone to overfitting

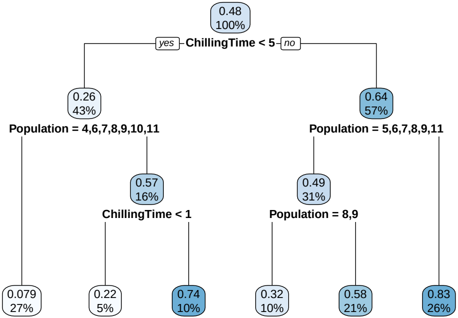


Figure 3: *Sample tree with cp = .01*

Model Evaluation/Selection

Model	In-Sample Accuracy	Out-of-Sample Accuracy
Logistic	0.7927	0.7662
Forest	0.8475	0.7749

- Random Forest does marginally better in accuracy
- Logistic model is much more interpretable
- Logistic model more clearly answers research questions

Will use the Logistic model to answer reserach questions

- Coefficients provided in Appendix

Effect of Chilling Time

3 types:

- Step-functions: above a certain value appears to be steady rate
- Increase up to a value, then decrease
- Primarily decreasing

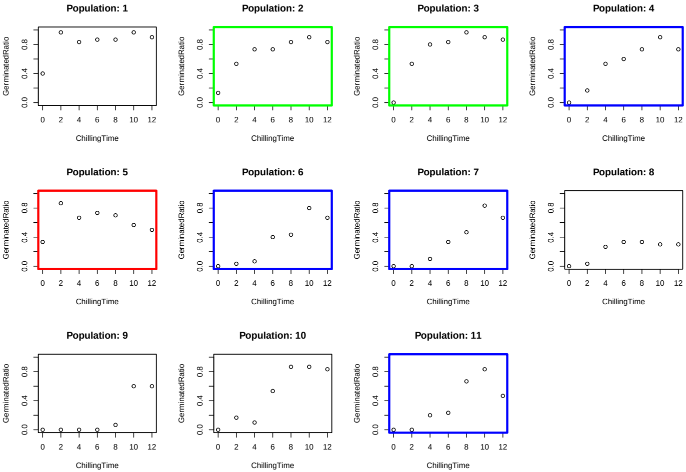


Figure 4: *Plots of chilling time on germination by population*

Ideal Chilling Time

Used model to identify ideal values:

Population	IdealChillWeeks
1	8.072
2	9.009
3	8.216
4	9.514
5	5.742
6	12.000
7	11.003
8	8.697
9	11.039
10	12.000
11	9.526

Effect of Decrease in Chilling Time

Used model to calculate difference in Germination rate at 9 vs. 10 weeks:

Population	GerminationDiff
1	0.77%
2	0.5%
3	1.36%
4	-0.03%
5	6.41%
6	-5.69%
7	-5.07%
8	1.85%
9	-26.03%
10	-4.09%
11	-0.11%

Summary

Used the provided tulips data to:

- Identify that impact of chilling time on germination rate depends on species (most increase with chilling time, some decrease)
- Identify the ideal chilling time for each species (for most, range of 8-10)
- Predict the impact of chilling time decreasing from 10 to 9 weeks on germination rate (some marginal increases, but several large decreases)

Next Steps

To improve our understanding of tulip chilling time on germination rate we suggest:

- Increased resolution of chilling times (ie. get samples of every week, maybe even per day)
- With risk of rising sea levels, test different humidity levels, soil saturation, etc. (effects of more moisture)

Table of Coefficients I

	x
(Intercept)	1.8119287
Population2	-1.0195843
Population3	-0.7643809
Population4	-1.9059727
Population5	-1.1972296
Population6	-3.2707801
Population7	-4.1310271
Population8	-3.3922643
Population9	-15.6040138
Population10	-2.2313094
Population11	-3.7559986
poly(ChillingTime, degree = 2)1	30.6800123
poly(ChillingTime, degree = 2)2	-25.3624386
Population2:poly(ChillingTime, degree = 2)1	16.4673616

Table of Coefficients II

Population3:poly(ChillingTime, degree = 2)1	34.2996294
Population4:poly(ChillingTime, degree = 2)1	39.3329341
Population5:poly(ChillingTime, degree = 2)1	-34.1673984
Population6:poly(ChillingTime, degree = 2)1	76.8147602
Population7:poly(ChillingTime, degree = 2)1	125.4569885
Population8:poly(ChillingTime, degree = 2)1	17.4668247
Population9:poly(ChillingTime, degree = 2)1	613.2091118
Population10:poly(ChillingTime, degree = 2)1	69.6451188
Population11:poly(ChillingTime, degree = 2)1	92.1533916
Population2:poly(ChillingTime, degree = 2)2	-1.5967130
Population3:poly(ChillingTime, degree = 2)2	-25.0972143
Population4:poly(ChillingTime, degree = 2)2	-8.8745982
Population5:poly(ChillingTime, degree = 2)2	1.2779898
Population6:poly(ChillingTime, degree = 2)2	-3.8462368
Population7:poly(ChillingTime, degree = 2)2	-28.4199363
Population8:poly(ChillingTime, degree = 2)2	-5.3469183
Population9:poly(ChillingTime, degree = 2)2	-194.6729684

Table of Coefficients III

Population10:poly(ChillingTime, degree = 2)2	2.8969412
Population11:poly(ChillingTime, degree = 2)2	-34.5792673