

Predicting pollination phenology in lodgepole pine

Susannah Tysor

January 25, 2019

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Introduction

As the climate changes, spring phenological events like budburst and flowering will advance, especially for plants active in rapid seasonal transitions and short growing seasons [@Pau2011], like many high elevation and latitude conifers. This effect is already obvious in many species [@Parmesan2006; @Franks2014]. Changes in pollination phenology can affect fecundity, gene flow, and even range size in a species and have effects on dependent species [@Inouye2008; @Chuine2001a].

[Conifers are important and affected by this problem, but it hasn't been investigated very much] In conifers, pollen is shed from male strobili and must arrive at receptive female strobili for successful pollination. Shifts in the timing of pollen shed and cone receptivity (pollination phenology) in conifers could lead to gene flow changes that hinder or promote adaptation under climate change, decrease fitness, and even affect reforestation via seed orchard productivity declines.

[talk about wind pollination]

[Also affects gene flow now and is important for understanding current spatial genetic structure and local adaptation] Common garden experiments and genetic work reveal extensive local adaptation in many forest tree species, especially boreal and temperate conifers (reviewed in @Alberto2013a). A locally adapted population only grows optimally in a subset of the range and may tolerate a more limited climatic range than the species as a whole. In northern hemisphere conifers, local adaptation often reflects strong trade-offs between avoidance of cold damage and competitive height growth (summarized in @Aitken2008a).

current and future gene flow could have significant adaptive consequences. In order to investigate the effects of gene flow on adaptation in lodgepole pine under climate change, we must first be able to predict pollination phenology.

Lodgepole pine, an ecologically and economically important species in western North America with lots of data on flowering phenology.

Seed orchards in British Columbia produce large amounts of tree seed for reforestation. Interior lodgepole pine (*Pinus contorta* var *latifolia*) is planted more than any other species at rates of [XXX] million/year as of [201X]. In the late 90s, it became apparent that seed yields (filled seeds per cone) at seed orchards in the northern Okanagan were much lower than in orchards near Prince George. To plan for future seed production and orchard establishment and management, studies were undertaken to understand why seed yields at north Okanagan orchards were lower than at the Prince George Tree Improvement Station orchards. Many orchards were already collecting at least some phenological data in order to rate seedlots [@Woods1996], but

standardized collection across all orchards began in 2006 under the Forest Genetics Council's Operational Tree Improvement Program 0722 [@Webber2007]. Pollination phenology data recorded as part of this program included pollen concentrations and dates of pollen shed and cone receptivity.

Predicting pollination phenology will also have practical benefits. Seed orchard managers in British Columbia are particularly concerned about protandry, when all pollen in an area is shed before cones become receptive (*personal communication, Chris Walsh, Former Seed Orchard Manager, Kalamalka Seed Orchards, February 13, 2013*). Protandry occurs in particularly hot and dry years [@Owens2005]. If this pattern holds, protandry could become more common in natural populations, leaving some populations pollen limited and likely hampering local adaption. [Outcrossing! Inbreeding!]

I will use pollination phenology and temperature data to fit a mechanistic model predicting pollen shed and cone receptivity across the entire range of *Pinus contorta* spp. *latifolia* and consider how the date and length of pollination phenology periods vary normally and under several climate change scenarios. Specifically, I will answer 1) What is the relationship between temperature and pollen shed and temperature and cone receptivity timing and length? 2) How strong is the genetic versus environmental effect on pollination phenology? 3) How many days will pollination phenology shift under the next 30, 60, and 90 years of climate change 4) Will protandry become more common?

Methods

Study species

[Need to explain earlier why local adaptation such a big deal for my questions - have to account for/test for it in model] Lodgepole pine is an economically and ecologically important tree species facing multiple threats from climate change [@Schneider2010; @Sambaraju2012; @Hamann2006]. Lodgepole pine has a very large geographical distribution (across 33° latitude and 31° longitude) encompassing a wide range of climates and soils (Figure 1) with widespread and significant local adaptation in many traits. For example, populations from both northern interior British Columbia and northern Idaho can survive in areas with mean annual temperatures between -4 and 6 °C, but the northern British Columbia population survives best where mean annual temperatures are ~ 1 °C and the Idaho population best at ~ 4 °C [@Rehfeldt1999a]. Local adaptation in lodgepole pine can be observed even at relatively small spatial scales when topographic variability is high: in a reciprocal transplant experiment, growth declines were observed when moving high elevation populations just 100m in elevation [@Rehfeldt1983]. Cone serotiny, an adaptive trait which allows for germination after fire events, is much more common in Rocky Mountain populations [@Morgenstern1996a] where fire is frequent. At smaller scales within the Rocky Mountain region, cone serotiny is predicted well by the type of disturbance that initiated the stand; serotinous cones are predominant in fire-initiated stands and open cones in other types [@Muir1985], with sharp boundaries between the two [@Lotan1976].

Several other traits make *Pinus contorta* suitable for this project. Pollen movement is an important vector for identifying gene flow in lodgepole pine because outcrossing is common, pollen dispersal is extensive, and seed dispersal is relatively limited [@Ennos1994]. Selfing is typically infrequent [@Sorensen2001], though when lodgepole pine occurs at low frequencies within a stand, selfing can climb to 21% [@Sorensen1993]. This project also builds on other landscape-scale work on adaptation in lodgepole pine: there is evidence for spatially varying levels of gene flow in the species as populations from areas with higher regional climate heterogeneity have higher genetic variance [@Yeaman2006], and the relative roles of wind, phenology and divergent selection are unclear. Crucial to the scale of this project, decades of high quality reproductive and vegetative phenology data are available from seed orchards producing seed for reforestation for lodgepole pine from many locations throughout the range. Because of lodgepole's importance, the large scale genomics project AdapTree is producing detailed phenological and genetic descriptions of local adaptation in this species that will complement this project.

While lodgepole pine has four subspecies (*P. contorta* spp. *contorta*, *P. contorta* spp. *murrayana*, *P. contorta* spp. *bolanderi*, and *P. contorta* spp. *latifolia*), I will be focusing solely on *latifolia*. The smaller distributions

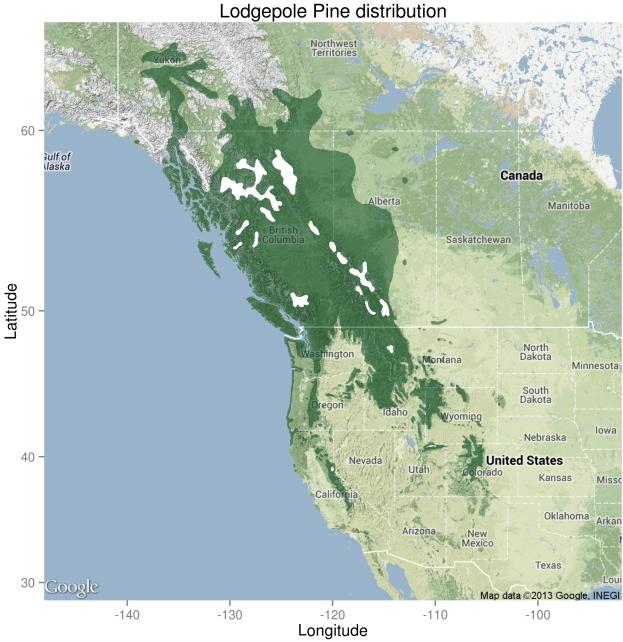


Figure 1: Rangemap of lodgepole pine. After @Little1971 .

and lower economic values of the *contorta*, *murrayana*, and *bolanderi* subspecies unfortunately limit data availability. Hereafter, lodgepole pine refers to *P. contorta* spp. *contorta* only unless otherwise noted. This includes references to the range of the species; only spp. *contorta* is under consideration.

In the first stage, I will model pollination phenology.

Data

Weather data

[build model with local data from orchards and weather stations, apply across landscape with PCIC daily gridded data

Phenology Data

Trees selected from across the British Columbia portion of the lodgepole pine range are grown in both large scale provenance trials and seed orchards, which are similar to common gardens, as part of tree breeding programs. There are two types of orchards: first generation and advanced. Each orchard contains trees that are the descendants of seeds and scions collected from mother trees in one particular provenance. First generation orchards have not undergone selective breeding but may be subject to selection through the process of choosing mother trees (called “plus” trees by tree breeders), testing and selection of mother tree offspring, or culling of unfavorable trees in the orchard. First generation orchard trees are clones of mother trees and their offspring grafted onto rootstock. There are three types of first generation orchards subject to different levels of selection: 1.0, 1.5, and 1.75 generation orchards. All orchards are subject to selection based on mother tree selection. In 1.5 generation orchards, superior offspring of mother trees are selected for the orchard or poor trees are culled from the orchard. This selection may be strong; more than 50% of offspring can be rejected in this process due to poor growth form [@Ukrainetz2011]. In 1.75 generation orchards, data from 10 year tests of mother tree offspring are used to select 40 of the best genotypes for a given provenance. Offspring from controlled crosses between trees in first generation orchards are tested in field trials and the best trees are then cloned and grafted onto root stock in advanced generation orchards.

```

orchards_in_data <- read.csv("~/Documents/research_phd/data/PhenologyAndPollenCounts/data_formatted_and_
  select(Orchard, Year, ) %>%
  group_by(Orchard) %>%
  dplyr::summarise(Years = n_distinct(Year))
orchard_gen <- read.csv("~/Documents/research_phd/data/OrrchardInfo/OrrchardGen.csv", header=TRUE, stringAsFactors=FALSE)

orchard_meta <- dplyr::left_join(orchards_in_data, orchard_gen) %>%
  unique()

## Joining, by = "Orchard"
knitr::kable(orchard_meta)

```

Orchard	Years	Generation
218	2	1.5
219	1	1.5
220	15	1.75
222	2	1.5
223	14	1.75
228	15	1.75
230	4	1.5
234	2	Advanced
237	3	Advanced
240	2	Unknown
307	4	1.75
308	1	1.75
310	1	1.5
311	1	1.5
313	1	1.5
338	2	Advanced
339	2	1

I contacted Seed Orchard Managers and other forestry professionals across British Columbia in 2012 and received pollination phenology data from C. Walsh, previously at Kalamalka Seed Orchards (now retired), R. Wagner at the Prince George Tree Improvement Station, and J.E. Webber previously at the Glyn Road Research Station (now retired).

I obtained records for 17 of the 26 lodgepole pine seed orchards in British Columbia. Orchards include the offspring of trees from 6 of the 7 BC seed planning zones (SPZs) (Figure 3) grown at 7 sites across BC. SPZs are biogeoclimatic and political units used for seed planting purposes by British Columbia. SPZs are divided into elevation bands called Seed Planning Units (SPUs), which form this project's provenances. At least 13 of the seed orchards are first generation orchards and should faithfully represent their provenances. These first generation orchards represent 6 provenances at 5 sites. Second generation orchards have been selectively crossed and this may skew the mean or variance of phenology for a provenance if pollination phenology varies by provenance.

Most provenances are represented at 2 to 3 sites and have at least three years of data at a given site spanning 1997-2012 (Figure 4). The Prince George Tree Improvement Station (PGTIS) provides a continuous 15-year record of its three orchards' phenology.

Provenances and sites included in my data exclude the coldest and wettest parts of the range. Figure 5 shows the 1961-1990 climate normal mean annual temperature (MAT) and mean annual precipitation (MAP) for data from all lodgepole pine provenances in BC, provenances included in my data set, and sites where provenances were grown. Orchard sites are generally much warmer and drier than many locations where

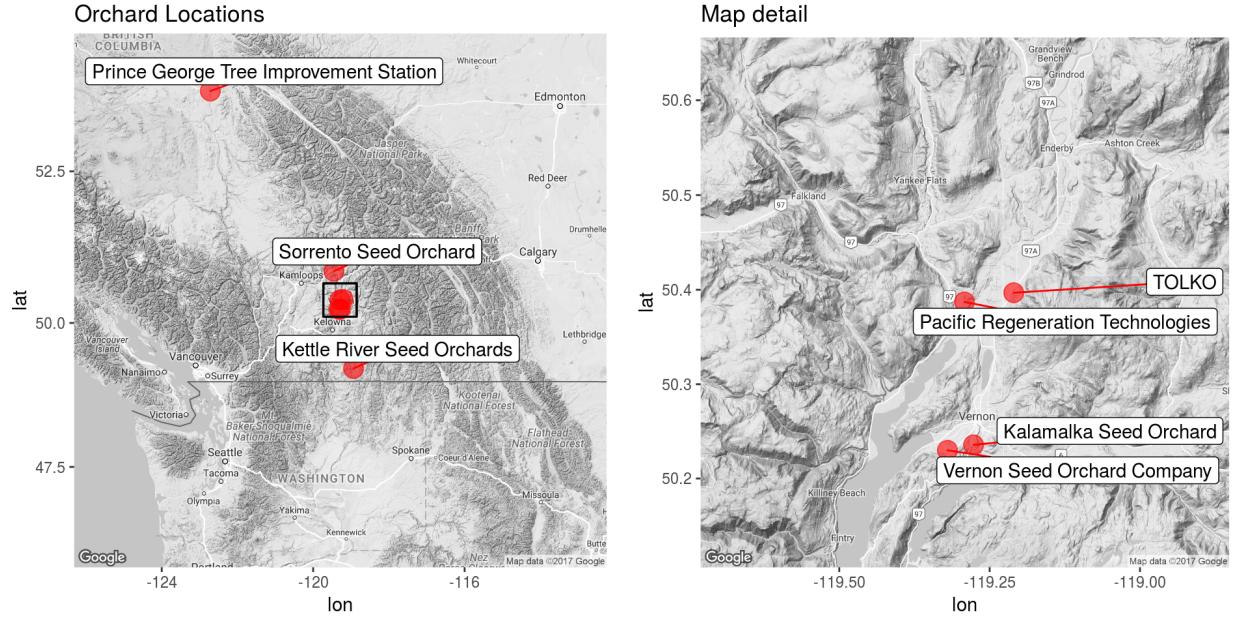


Figure 2: Map of seed orchard locations. Boxed area in map on left is shown in greater detail on the right.

lodgepole pine grow. I may have difficulty extrapolating into the coolest and wettest parts of the lodgepole pine range, depending on actual yearly conditions at sites for which I have data. However, this may be advantageous when it comes to predicting phenology under climate change.

Phenophases

There are three phenological stages of interest each for a tree's male and female cones

- 1. The cones have not yet flowered
- 2. The cones are flowering
- 3. The cones have finished flowering

Phenophases in the field were recorded using different symbol sets and resolutions. I assigned each symbol to one of the phenophases above. Trees that did not produce cones are assigned phenological stage 0.

```
no_flowers <- '0'
#before_flowering <- c('1', '2.5', '-') #stage 1
before_flowering <- c("1", 2.5, "-")
#flowering <- c('3', '3.5', '4', '4.5', '5', 'pollenshed20', 'receptive20') #stage 2
flowering <- c("3", 3.5, 4, 4.5, 5, pollenshed20, receptive20")
#after_flowering <- c('-', 'receptive80', 'pollenshed80') #stage 3
after_flowering <- c("-", receptive80, pollenshed80")
derived_phenophase <- c(0:3)
male_desc <- c("none produced", "not yet shedding", "shedding", "finished shedding")
female_desc <- c("none produced", "not yet receptive", "receptive", "no longer receptive")
recorded_as <- c(no_flowers, before_flowering, flowering, af=after_flowering)

phenophase_legend <- data.frame(Phenophase = derived_phenophase, Symbols = recorded_as, Male.Cones = ma
```

Lodgepole Pine Seed Planning Units

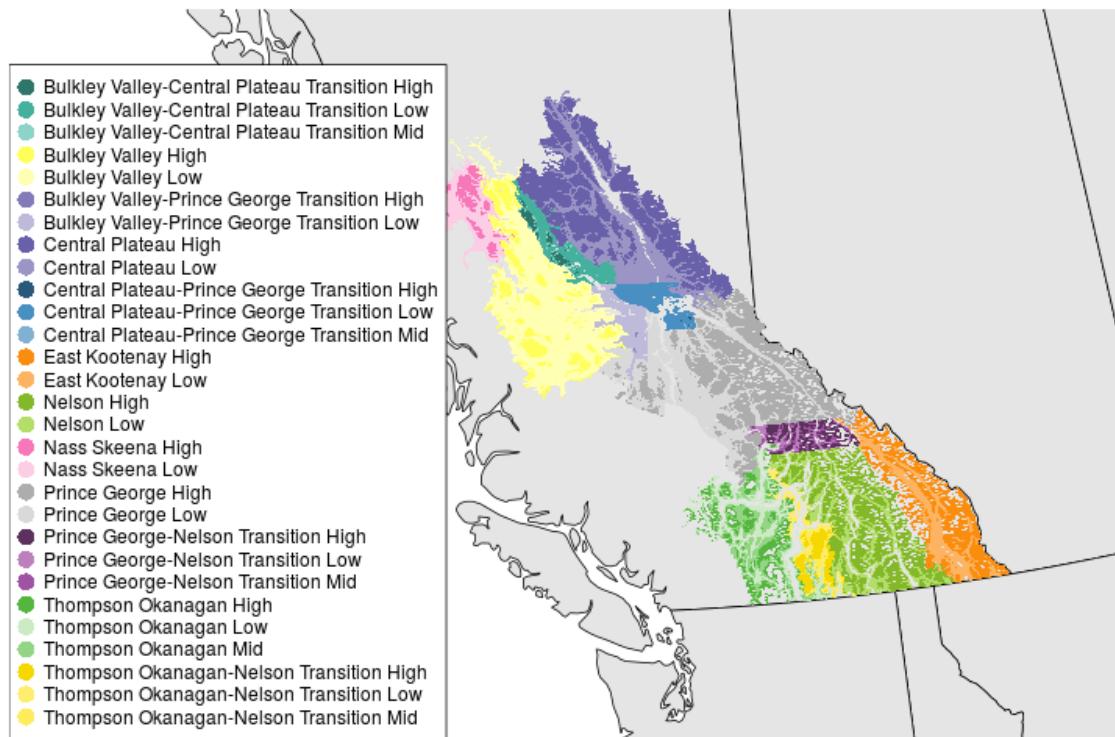


Figure 3: Map of Seed Planning Units (SPUs). Seed planning units are biogeoclimatic and political units used for seed planting purposes by British Columbia. Seed planning units form this project's provenances. High, Low, and Mid refer to elevational bands. Data is also available for East Kootenay Low, but will likely not be included in any analysis as it includes only one year at one site.

	Kalamalka	Kettle River	PGTIS	PRT	Sorrento	TOLKO	Vernon	Total
Bulkley Valley Low	5	0	17	0	2	0	3	27
Central Plateau Low	0	0	15	0	0	0	4	19
Nelson Low	5	0	0	3	0	0	0	8
Prince George Low	0	2	18	0	0	0	4	24
Thompson Okanagan High	0	0	0	0	0	3	0	3
Thompson Okanagan Low	0	0	0	6	0	0	0	6
Total	10	2	50	9	2	3	11	87

Figure 4: Contingency table of Seed Planning Units (rows) and Seed Orchard Sites (columns). Seed Planning Zones, used as provenances in this project, are usually represented at multiple years and multiple sites. There is particularly good representation at PGTIS.

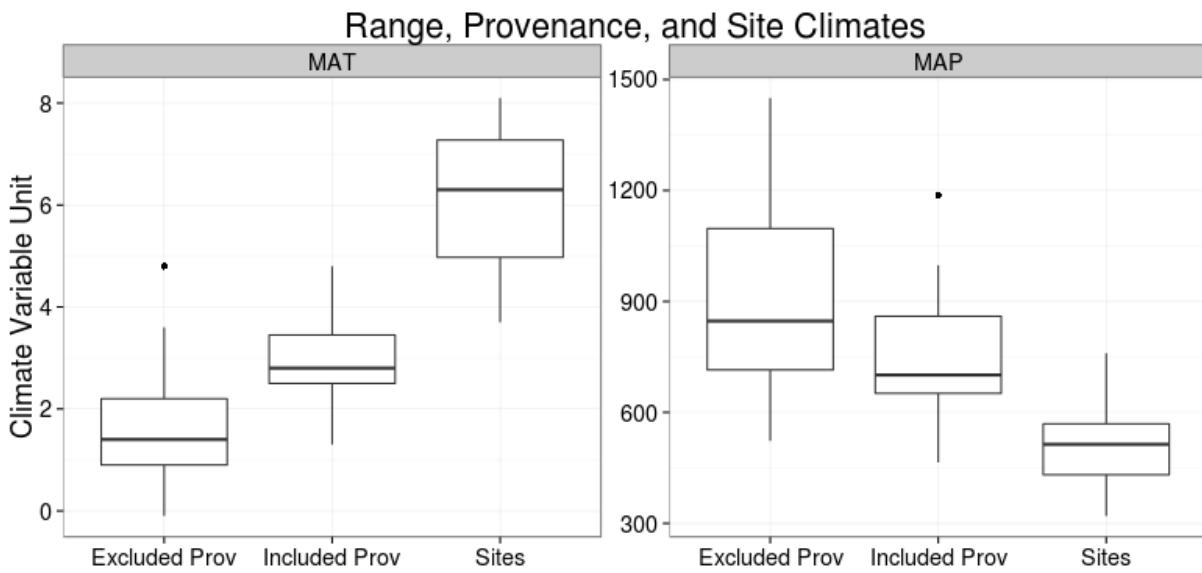


Figure 5: Provenance (SPU) climates of data included in this proposal, provenances not included in this proposal, and sites where provenances were grown. MAT is mean annual temperature and MAP is mean annual precipitation.

Phenophase	Symbols	Male.Cones	Female.Cones
0	0	none produced	none produced
1	1, 2.5, -	not yet shedding	not yet receptive
2	3, 3.5, 4, 4.5, 5, pollensheded20, receptive20	shedding	receptive
3	-, receptive80, pollensheded80	finished shedding	no longer receptive

My aim was to model pollination phenology in lodgepole pine and calculate synchrony across the distribution. To determine the timing of pollen shed and cone receptivity, I modeled the forcing requirements for pollination phenology in lodgepole pine and investigated differences in forcing requirements between males and females and among provenances. I then applied the model to weather data across the lodgepole pine distribution to determine the distance between populations with synchronous pollen shed and cone receptivity.

Materials and methods

Between 1997 and 2011, flowering phenology of lodgepole pine was recorded at 8 seed orchard sites in the interior of British Columbia as part of a project to identify causes of low seed set [Cite 0722]. [define seed orchard] Prince George Tree Improvement Station has the most complete record of phenology. 4 of the sites are clustered near to one another, but sites span about 5 degrees of latitude and are at elevations from x to y. [map] Trees grown at the seed orchard sites were sourced from 6 biogeoclimate regions known as Seed Planning Units. [describe selection and generations]. Trees with the same SPU provenance are grown together in an orchard at a given site. Genotypes (labelled with a Clone number in the data) in the orchards are represented by multiple ramets.

[table with Site Columns and SPU rows with years of data as values][table of number of trees/clones in a given year for a given site/spu combo]

Protocol C in [@Woods1996] was used as the basis for collecting pollen shed and cone receptivity data, though operational constraints led to some modifications. Workers monitored seed orchards for the beginning of pollen shed and cone receptivity. ~15 clones, usually represented by 2 trees each, were selected for specific observations. When the active period seemed to be starting, workers went into the orchard every few days to make observations on the selected trees.

Stages of pollen and seed cone development are described by [Owens & Molder 1984 and updated in Owens2006] and were used as a general guide for determining the phenological state of pollen and seed cones. Pollen cones are flowering when tapping causes pollen to be released and seed cones are flowering when there are gaps between the bract-scale complexes. Pollination drops are also produced during flowering, though they recede midday if pollen is not present [@Owens2006].

“Flowering” states were recorded for both pollen shed and cone receptivity at the level of each tree. Protocol C recommends marking the dates when 20% of the cones on a tree have begun flowering and when 80% of the cones on a tree have finished flowering. Operationally, there was some subjectivity and tree-level states for each cone type should be interpreted as “starting flowering” and “finished flowering.” There is some subjectivity here.

There were some differences in how data was recorded at the Prince George Tree Improvement Station versus the other sites. At Prince George, trees were marked as flowering or not flowering on each day of observation. At other sites, only the first day observed flowering and the first day observed finished flowering were recorded. I inferred that trees were not yet flowering on observation days prior to their first recorded flowering date. I cleaned and harmonized the data for analysis in a single model using R scripts provided so that, for pollen cones and seed cones, stage 1 = not yet flowering, stage 2 = flowering, and stage 3 finished flowering.

Interval censoring and truncation

Observations were not made every day and survey periods varied in length. At Prince George, not all trees have complete phenological records, *e.g.* if a tree is not flowering on the first day of observation, the start date is unknown.

Not all trees have complete phenological records. At Prince George, `# # Modeling`

Phenology Modeling

[Probably most of this isn't necessary] Scientists have modeled plant phenology for centuries with an overwhelming variety of model structures. Predictive phenology models are usually mechanistic, but the parameters are rarely directly measured since many stages of development are unseen. Instead, models rely on dates of a few visible stages of development, such as budburst or flowering. Modeling frameworks have different assumptions about how sensitive development is to chilling and warming (or “forcing”) and how sensitivity changes throughout the winter and spring. These assumptions are very difficult to test because of the difficulty of observing internal development. While numerous assumptions spawned even more numerous models, Isabelle Chuine noticed that all of the major budburst models could be represented by different parameterizations of the Gompertz function, a special case of the generalized logistic function [@Chuine2000].

Modeling pollination phenology in conifers is not new, but it is uncommon. Simple heat accumulation thresholds (*Pinus taeda*) [@Boyer1978] or elevation (*Pinus flexilis*) [@Schuster1989] were used previously to explain or predict pollen shed in limited spacial and temporal contexts. @Owens2006 reports that lodgepole pine pollen shed and cone receptivity occur when degree days reach about 500 at a threshold of 5 °C, but this is the only report of pollination phenology modeling for lodgepole pine and no details are provided. (Degree days are used to measure heating or cooling from a base temperature over time.) Models of lodgepole pine vegetative phenology, on the other hand, are better represented in the literature (*e.g.* @Chuine2001), and pollen shed and cone receptivity are not expected to have additional or more complex triggers or model forms than budburst [@Chuine2013a].

Previous models for lodgepole pine have fitted spring growth phenology models without considering chilling [@Chuine2001] and, in general, budburst models for boreal tree species have not required chilling to give accurate predictions [@Linkosalo2000]. For lodgepole pine growing in natural environments, chilling is always met and should continue to be met under the next century of climate change; in the AdapTree project, lodgepole pine seedlings grown for three to four weeks at 4 °C met their chilling requirements. Thus, I expect to be able to disregard the state of chilling and fit only Equations ?? and ??). Once the model is parameterized I will be able to calculate the timing of pollen shed and cone receptivity across the lodgepole pine species range using location specific climate data from seed orchards and ClimateWNA [@Wang2012]. I will validate the model by comparing predictions to phenological measurements from the National Phenology Network.

Actual methods

I modeled the derived phenological states with a Bayesian multilevel ordered logistic model fit in Stan [@standevelopmentteamRStanInterfaceStan2018].

For phenological states $k \in \{1, 2, 3\}$, the probability of

The ordered logistic probability mass function is

$$\text{OrderedLogistic}(k \mid \eta, c) = \begin{cases} 1 - \text{logistic}(\eta - c_1) & \text{if } k = 1, \\ \text{logistic}(\eta - c_1) - \text{logistic}(\eta - c_2) & \text{if } k = 2, \text{ and} \\ \text{logistic}(\eta - c_3) - 0 & \text{if } k = 3. \end{cases}$$

S_i	$\sim \text{OrderedLogistic}(\eta, \kappa)$	probability of data
η_i	$= \beta_{\text{sex}} h_i + \alpha_{\text{provenance}} + \alpha_{\text{clone}}$	linear model
κ_k	$\sim \text{normal}(5, 10)$	common prior for cutpoints
β_{clone}	$\sim \text{beta}(1, 1)$	prior for transition speed
α_{clone}	$\sim \text{normal}(0, 1)$	prior for clone threshold effects
$\alpha_{\text{provenance}}$	$\sim \text{normal}(0, 1)$	prior for provenance threshold effects

The overall threshold for phase transition can be calculated by dividing the slope by the intercept.

$\mathbf{H_1}$	$= \beta_{\text{sex}} / (c_1 + \alpha_{\text{clone}} + \alpha_{\text{provenance}})$	
$\mathbf{H_2}$	$= \beta_{\text{sex}} / (c_2 + \alpha_{\text{clone}} + \alpha_{\text{provenance}})$	
S_i	$\sim \text{Categorical}(\mathbf{p})$	probability of data
p_1	$= q_1$	probabilities of each value of k
p_2	$= q_2 - q_1$	
p_3	$= 1 - q_3$	
$\text{logit}(q_k)$	$= \kappa_k - \eta_i$	cumulative logit link
η_i	$= \beta_{\text{sex}} h_i + \alpha_{\text{provenance}} + \alpha_{\text{clone}}$	linear model
κ_k	$\sim \text{normal}(5, 10)$	common prior for cutpoints
β_{clone}	$\sim \text{beta}(1, 1)$	
α_{clone}	$\sim \text{normal}(0, 1)$	
$\alpha_{\text{provenance}}$	$\sim \text{normal}(0, 1)$	
$\mathbf{H_1}$	$= \beta_{\text{sex}} / (c_1 + \alpha_{\text{clone}} + \alpha_{\text{provenance}})$	
$\mathbf{H_2}$	$= \beta_{\text{sex}} / (c_2 + \alpha_{\text{clone}} + \alpha_{\text{provenance}})$	

To quantify accuracy of the model, I will calculate the difference in variance within and between populations. Only phenological differences larger than the within population variance will be considered. The maps and parameters of these models can be compared to answer the following questions

Results

- 1) How strong are the genetic vs. environmental effects on pollination phenology?

To answer this question I will compare the parameters of models fit by provenance. If different provenances have significantly different model parameters - that is, have different heat sum requirements for pollen shed and cone receptivity by provenance, then the rest of my project will be limited to the provenances I have data from. If not, I can use the same models across the entire range when predicting phenological events.

The budburst and spring pollination phenology of temperate tree species is determined by temperature, and there may be some regional variation in the temperatures required for pollen shed and cone receptivity; in lodgepole pine, the threshold for shoot elongation was 5.1 °C for coastal and interior provenances, but 6.5 °C for northern provenances [@Chuine2001].

However, seedling spring vegetative phenology data from four growth chamber temperature and moisture treatments in the AdapTree project suggests that budburst shows substantial clinal variation (> 10% of phenotypic variance explained) only in the coldest treatment tested (MAT of 1 °C), and much colder than the seed orchard locations (Figure 5, Figure 6, @Liepe2014 (submitted)). Analysis of phenotypic data from an outdoor common garden in Vancouver (*personal communication, Ian MacLachlan, AdapTree Graduate Student, January, 1 2015*) also shows that very little variation exists among provenances for spring phenology, regardless of whether seedlings originate from wild-stand or selectively bred seed orchard seedlots.

- 2) How does faster spring warm-up like we expect under climate change affect within population mating success and frequency of outcrossing?

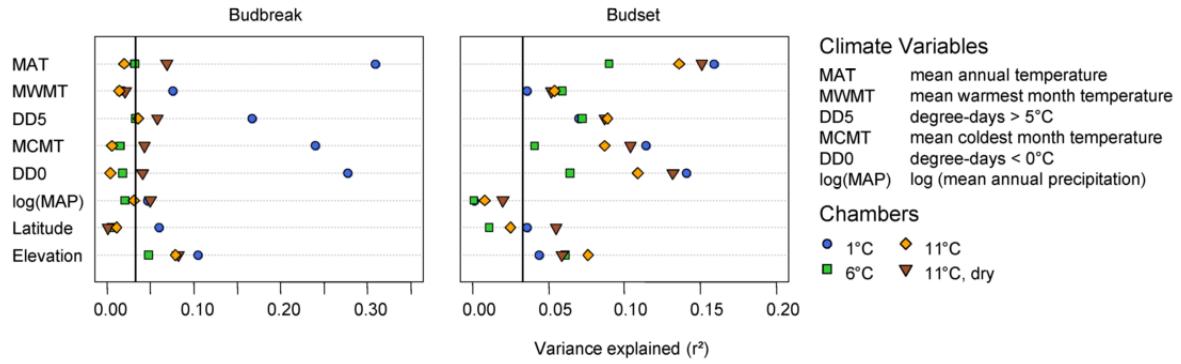


Figure 6: Variance in budbreak and budset for lodgepole pine explained by provenance climate. Seedlings from 281 locations across lodgepole's BC range were grown in growth chambers under four treatments: approximated seasonal and daily variation in temperature for geographic locations with a MAT of 1, 6, and 11 °C, with the warmest treatment having a dry and wet version. Pearson's correlation coefficient between phenotypic traits in lodgepole pine and provenance climates are presented. The vertical line represents the critical r^2 value after an adjustment for multiple inferences. From @Liepe2014 (submitted).

I expect that faster spring warm-up will condense as well as advance growth initiation dates and also shorten pollination phenological periods by speeding development prior to and during the phenological period. This should further synchronize populations in similar climates and act as a barrier to breeding with populations in more different climates. On the other hand, it could favor outcrossing over selfing by causing protandry - pollen shed in advance of cone receptivity. Based on anecdotal reports from seed orchards managers, I expect protandry (and thus outcrossing) to increase under climate change.

I will examine the variability of the start date across the range (t_1) to see if it shrinks in warmer years and as the climate changes. I will also calculate overlap between within population cone receptivity and pollen shed under current conditions and with climate warming and see if the heat sum requirements for cone receptivity and pollen shed are different.