

Use Growth Stage Model to Visualize Corn Growth

Plant development depends on temperature. Plants require a specific amount of heat to develop from one point in their lifecycle to another, such as from seeding to the four-leaf stage. Traditionally a calendar is used to predict plant growth, however it is possible for temperature to vary greatly from year to year, and even in locations where temperature is relatively stable, insight can still be gained by analyzing growth stages. Growing degree days (GDD) is an agronomic value that assigns a heat value to each day. Based on GDD, agronomic scientists have developed growth stage models to describe a plant's physiological response to temperature. Most models are developed for crops in the developed world, where large commercial farms have an incentive to use models for insight over a large land area. In this case study, we will be using data from a trials study in Uganda of the Situka variety of maize to create a growth stage model that can represent the development of a Ugandan maize crop. Situka 1 and Situka-M1 are two types of drought tolerant and natural-calamity resistant maize variety that were introduced in 2011 in Tanzania by the Selian Agricultural Research Institutes (SARI). Situka 1 can be ready for harvest in 75 days, and Situka-M1 can be ready for harvest in between 75 and 105 days. To make the model, we have assumed the following: - The Situka maize varieties require fewer GDDs to mature compared to other Ugandan maize varieties, and we can use the Corn2300ISUAbendroth model as a reference - The majority of farmers in this trial harvested maize immediately after the crop reached physiological maturity. - The Growth stage model is scalable.

The trials data we will use was compiled by AGRA. According to MS Seth et al., for the Situka variety of maize we can expect 99 days to anthesis and an additional 6 days interval from anthesis to silking (ASI), for a total of 105 days to silking.

<http://www.ajol.info/index.php/ajb/article/view/101751> The international research project Water Efficient Maize for Africa (WEMA) noted that Situka can be grown in arid conditions like those in Tanzania and be ready for harvest in just 75 days. Most maize varieties require at least 90 days to mature. <http://news.trust.org/item/20120514103400-g0an4/?source=spotlight>. Additional documentation suggests about Situka requires about 57-59 days to tasselling. http://www.subaagro.com/index_files/OPV.htm Based on this evidence our modeling assumptions will be that Situka requires fewer GDDs to maturity than traditional maize, and the maize will be harvested once it has reached physical maturity.

In [1]:

```
httr::set_config(httr::config(ssl_verifypeer = 0L))
# devtools::install_github("yizhexu/aWhere-R-Library")
```

In [2]:

```
suppressWarnings(suppressMessages(library(aWhereAPI)))
library(magrittr)
library(openxlsx)
library(httr)
library(dplyr)

awhere_endpoints <- httr::oauth_endpoint(base_url = "https://api.awhere.com/v2",
                                       access = "https://api.awhere.com/oauth/token",
                                       authorize = NULL,
                                       fields = "fields",
                                       plantings = "agronomics/plantings",
                                       weather = "weather/fields",
                                       imagery = "imagery/weathermaps",
                                       agronomics = "agronomics/fields",
                                       crops = "agronomics/crops",
                                       models = "agronomics/models",
                                       batch = "jobs")

# get gdd data for all farmers
api_key = "yizhexu@awhere.com"
api_secret = "181225tiancai@X"

get_token(api_key, api_secret)
```

```
Warning message:
: package 'magrittr' was built under R version 3.2.4
Warning message:
: package 'openxlsx' was built under R version 3.2.4
```

In [3]:

```
data <- read.xlsx("./example_trials.xlsx")
data <- data[data$Variety == "Situka", ]
head(data, n=6)
```

Out[3]:

	FarmerId	Year	Season	PlantingDate	HarvestingDate	Latitude	Longitude	Altitude	Plot	Subplot	Treat	RepNo
19	Farmer-2	2011	Short rain	40819	40944	1.377222	31.76889	1283	1	1	1	1
20	Farmer-2	2011	Short rain	40819	40944	1.377222	31.76889	1283	2	2	2	1
21	Farmer-2	2011	Short rain	40819	40944	1.377222	31.76889	1283	3	3	3	1
22	Farmer-2	2011	Short rain	40819	40944	1.377222	31.76889	1283	4	4	4	1
23	Farmer-2	2011	Short rain	40819	40944	1.377222	31.76889	1283	5	5	5	1
24	Farmer-2	2011	Short rain	40819	40944	1.377222	31.76889	1283	6	6	6	1

In [4]:

```
# separate long rain farmers and short rain farmers
long_rain <- data[data$Season == "Long rain", c("FarmerId", "PlantingDate", "HarvestingDate", "Latitude", "Longitude")]
short_rain <- data[data$Season == "Short rain", c("FarmerId", "PlantingDate", "HarvestingDate", "Latitude", "Longitude")]
```

In [5]:

```
long_rain <- long_rain %>% unique()
short_rain <- short_rain %>% unique()

nrow(long_rain)
nrow(short_rain)
```

Out[5]:

6

Out[5]:

17

Evaluate total accumulative GDD for each long rain trial

The goal of this exercise is to scale up a GDD model for use in Uganda. To do this we must use the trials data to determine the optimal GDD load for Situka in the Ugandan climate. Africa typically has two crop seasons, corresponding to two distinct periods of rains. Because the seasonal characteristics are different, we will evaluate the trials for each season separately.

Below we construct an API call that uses the trial planting date, harvest date, and location of trials during the long rain season to calculate the total GDDs recorded at each field during their individual seasons. The result is graphed.

In [6]:

```
query <- paste0(
  file.path("https://api.awhere.com/v2/agronomics/locations", paste0(long_rain$Latitude, ",", long_rain$Longitude),
    "agronomicvalues", paste0(as.character(as.Date(long_rain$PlantingDate[1], origin = "1899-12-30")),
    ",", as.character(as.Date(long_rain$HarvestingDate[1], origin = "1899-12-30"))),
    "?properties=accumulatedGdd&gddMethod=standard&gddMaxBoundary=35")

all_result <- lapply(1:length(query), function(i) {
  GET(query[i], add_headers(Authorization = paste0("Bearer ", awhereEnv75247$token))) %>% content()
})
```

In [7]:

```
all_gdd <- unlist(lapply(1:length(all_result), function(i) {
  all_result[[i]]$accumulations$gdd
})))
```

In [8]:

```
sort(all_gdd)
```

Out[8]:

```
1616.97500896454 1641.48000049591 1692.84499549866 1707.61500358582 1707.61500358582
1948.84000301361
```

In [9]:

```
library(ggplot2)
ggplot(data.frame(x=all_gdd), aes(x)) + geom_density()
```

From the GDD distribution density chart above, we can see the data clusters around 1700 GDDs from planting to harvest for each trial, but that a small number are observed clustering closer to 2000 GDDs. There are a few potential explanations for the skew, including misrecorded dates of planting or harvest. Most notably, however, since the trials data records harvest date but not physical maturity date, it is very possible that on some trial sites harvest did not occur immediately after the crop reached maturity, whether due to lack of available labor or adverse conditions.

Because the majority of trials fell into a relatively narrow GDD range, and the academic research suggests Situka maize should theoretically require fewer GDDs overall, we will subset the trials that fell within the accumulative GDD range of 1640 to 1710, and assume that the remaining trials are outliers.

In [10]:

```
# all_gdd <- all_gdd %>% unique()

selected_farmers <- (1:length(all_gdd))[all_gdd >= 1640 & all_gdd <= 1710]
selected_farmers_accgdd <- all_gdd[all_gdd >= 1640 & all_gdd <= 1710]
```

In [11]:

```
# Find the daily accumulative gdds for all of these selected farmers
selected_gdd <- unlist(lapply(1:length(selected_farmers), function(i) {
  farmer = selected_farmers[i]
  all_result[[farmer]]$dailyValues
}))
```

As we've seen, Situka is a drought-tolerant variety that requires relatively fewer days than normal to mature in Ugandan climate. Therefore theoretically when matching it up to a comparable US variety, it should fit more into a lower GDD model. Using aWhere's Models API, we can see the lowest GDD crop stage model for corn is named "corn2300ISUAbendroth".

The code below retrieves the detailed for Corn2300ISUAbendroth.

In [12]:

```
type <- "details"
corns <- "Corn2300ISUAbendroth"

query <- file.path(awhere_endpoints$models, corns, type)
request <- GET(query, add_headers(Authorization = paste0("Bearer ",awhereEnv75247$token))) %>% content()

growth_stage_us <- request$stages %>% jsonlite::toJSON(pretty = TRUE) %>% jsonlite::fromJSON(flatten = TRUE)
)
growth_stage_us <- data.frame(lapply(1:ncol(growth_stage_us), function(i) {
  unlist(growth_stage_us[[i]])
})) %>% setNames(colnames(growth_stage_us))
```

In [13]:

```
growth_stage_us
```

Out [13]:

	gddUnits	id	stage	description	gddThreshold
1	C	stage1	VE	Emergence	45
2	C	stage2	V2	Two leaves with collar visible	106
3	C	stage3	V4	Four leaves with collar visible	175
4	C	stage4	V6	Six leaves	236
5	C	stage5	V8	Eight leaves	304
6	C	stage6	V10	Ten leaves	381
7	C	stage7	V12	Twelve leaves	441
8	C	stage8	V14	Fourteen leaves	502
9	C	stage9	V16	Sixteen leaves	551
10	C	stage10	V18	Eighteen leaves	568
11	C	stage11	VT	Tasseling	609
12	C	stage12	R1	Silking	679
13	C	stage13	R2	Blister	771
14	C	stage14	R3	Milk	856
15	C	stage15	R4	dough stage	952
16	C	stage16	R5	Dent	1088
17	C	stage17	R6	Physiological maturity	1278

In [14]:

```
# scale this growth stage model to fit selected farmers accgdd
long_scaled_models <- do.call(cbind, lapply(1:length(selected_farmers), function(i) {
  multiplier <- selected_farmers_accgdd[i] / growth_stage_us$gddThreshold[[grep("R6", growth_stage_us$stage)]]
  growth_stage_us$gddThreshold*multiplier
}))
```

In [15]:

```
# average each row to find average threshold gdd for selected farmer
# long_agra_farmer_model <- cbind(growth_stage_us[, 1:4], gddThreshold = rowMeans(scaled_models))
# long_agra_farmer_model
```

Evaluate total accumulative GDD for each short rain trial

The above process evaluated the trials conducted during the long rain season, and we will now follow the same process to find a comparable growth stage model for trials conducted during the short rain season.

In [16]:

```
query <- paste0(
  file.path("https://api.awhere.com/v2/agronomics/locations", paste0(short_rain$Latitude, ",", short_rain$
Longitude),
  "agronomicvalues", paste0(as.character(as.Date(short_rain$PlantingDate[1], origin = "1899-12-3
0")), ",", as.character(as.Date(short_rain$HarvestingDate[1], origin = "1899-12-30")))),
  "?properties=accumulatedGdd&gddMethod=standard&gddMaxBoundary=35")

all_result <- lapply(1:length(query), function(i) {
  GET(query[i], add_headers(Authorization = paste0("Bearer ", awhereEnv75247$token))) %>% content()
})

all_gdd <- unlist(lapply(1:length(all_result), function(i) {
  all_result[[i]]$accumulations$gdd
}))

library(ggplot2)
ggplot(data.frame(x=all_gdd), aes(x)) + geom_density()

sort(all_gdd)
```

Out[16]:

```
1607.43999624252 1649.15000247955 1649.15000247955 1649.15000247955 1649.15000247955
1655.04500150681 1656.50499820709 1661.17000341415 1721.78999853134 1727.92000055313
1741.43499469757 1745.31000423431 1746.10999965668 1824.41999435425 1824.41999435425
1824.41999435425 1982.41499233246
```

In [17]:

```
# remove the
# all_gdd <- all_gdd %>% unique()

selected_farmers <- (1:length(all_gdd))[all_gdd >= 1640 & all_gdd <= 1710]
selected_farmers_accgdd <- all_gdd[all_gdd >= 1640 & all_gdd <= 1710]

# Find the daily accumulative gdds for all of these selected farmers
selected_gdd <- unlist(lapply(1:length(selected_farmers), function(i) {
  farmer = selected_farmers[i]
  all_result[[farmer]]$dailyValues
}))

short_scaled_models <- do.call(cbind, lapply(1:length(selected_farmers), function(i) {
  multiplier <- selected_farmers_accgdd[i] / growth_stage_us$gddThreshold[[grep("R6", growth_stage_us$sta
ge)]]
  growth_stage_us$gddThreshold*multiplier
}))
```

In [18]:

```
selected_farmers
```

Out[18]:

```
4 5 9 10 11 13 15
```

Does the Situka-scaled model perform differently in short rain vs long rain seasons?

The long and short rain seasons occur at different times of the year, and we might be concerned that the different calendar will significantly impact growth stage thresholds. A statistical test can help us determine this - in our case, we cannot assume the two samples are taken from populations that follow a normal distribution. By using the Mann-Whitney U-test, we can test if the trial GDD thresholds had identical distribution at a $p < 0.05$ significance.

In [19]:

```
wilcox.test(long_scaled_models[1,], short_scaled_models[1, ], alternative = "two.sided")
Warning message:
In wilcox.test.default(long_scaled_models[1, ], short_scaled_models[1, : cannot compute exact p-value with t
ies
```

Out[19]:

```
Wilcoxon rank sum test with continuity correction

data: long_scaled_models[1, ] and short_scaled_models[1, ]
W = 21, p-value = 0.2076
alternative hypothesis: true location shift is not equal to 0
```

The U-test tried to determine if the GDD thresholds to emergence (using the scaled versions adapted from Corn2300ISUAbendroth) are significantly different for the long and short rain trials. The null hypothesis is that the GDD thresholds are identical. As the p-value turns out to be 0.2076, we cannot reject the null hypothesis. Repeating this test for all the other growth stages will similarly yield no significant differences. As a result, we will create only one Ugandan Situka Maize GDD model.

NOTE: Due to small sample sizes, data subsetting, and the theoretical assumptions made to scale the models, we should not rely too much on the statistical tests demonstrated here to prove model validity. This exercise is not intended to be scientifically reliable or used for decisions, rather it is intended as a demonstration of how real trials data can be used to estimate a GDD model based on thresholds already outlined in scientific literature.

In [20]:

```
scaled_models <- cbind(long_scaled_models, short_scaled_models)

# average each row to find average threshold gdd for selected farmer
uganda_situka_model <- cbind(growth_stage_us[, 1:4], gddThreshold = rowMeans(scaled_models))
uganda_situka_model
write.xlsx(uganda_situka_model, "uganda_situka_model.xlsx")
```

Out[20]:

	gddUnits	id	stage	description	gddThreshold
1	C	stage1	VE	Emergence	58.63916
2	C	stage2	V2	Two leaves with collar visible	138.1278
3	C	stage3	V4	Four leaves with collar visible	228.0412
4	C	stage4	V6	Six leaves	307.5298
5	C	stage5	V8	Eight leaves	396.1401
6	C	stage6	V10	Ten leaves	496.4783
7	C	stage7	V12	Twelve leaves	574.6638
8	C	stage8	V14	Fourteen leaves	654.1525
9	C	stage9	V16	Sixteen leaves	718.004
10	C	stage10	V18	Eighteen leaves	740.1566
11	C	stage11	VT	Tasseling	793.5834
12	C	stage12	R1	Silking	884.7998
13	C	stage13	R2	Blister	1004.684
14	C	stage14	R3	Milk	1115.447
15	C	stage15	R4	dough stage	1240.544
16	C	stage16	R5	Dent	1417.765
17	C	stage17	R6	Physiological maturity	1665.352

Scale a model for the Kilima maize variety

The above exercise used data specific to a drought-tolerant, low-GDD variety of maize in use in East Africa. The Kilima variety is not drought-tolerant, and should show different characteristics when similarly examined.

In [21]:

```
data <- read.xlsx("./example_trials.xlsx")
data <- data[data$Variety == "Kilima", ]
head(data, n=6)
```

Out[21]:

	FarmerId	Year	Season	PlantingDate	HarvestingDate	Latitude	Longitude	Altitude	Plot	Subplot	Treat	RepNo
1	Farmer-1	2013	Long rain	41315	41450	1.09829	31.39215	1179	1	1	1	1
2	Farmer-1	2013	Long rain	41315	41450	1.09829	31.39215	1179	2	2	2	1
3	Farmer-1	2013	Long rain	41315	41450	1.09829	31.39215	1179	3	3	3	1
4	Farmer-1	2013	Long rain	41315	41450	1.09829	31.39215	1179	4	4	4	1
5	Farmer-1	2013	Long rain	41315	41450	1.09829	31.39215	1179	5	5	5	1
6	Farmer-1	2013	Long rain	41315	41450	1.09829	31.39215	1179	6	6	6	1

In [22]:

```
data <- data[data$Season == "Long rain", c("FarmerId", "PlantingDate", "HarvestingDate", "Latitude", "Longitude")] %>% unique()
query <- paste0(
  file.path("https://api.awhere.com/v2/agronomics/locations", paste0(data$Latitude, ",", data$Longitude),
    "agronomicvalues", paste0(as.character(as.Date(data$PlantingDate[1], origin = "1899-12-30")),
    ",", as.character(as.Date(data$HarvestingDate[1], origin = "1899-12-30")))),
  "?properties=accumulatedGdd&gddMethod=standard&gddMaxBoundary=35")

all_result <- lapply(1:length(query), function(i) {
  GET(query[i], add_headers(Authorization = paste0("Bearer ", anywhereEnv75247$token))) %>% content()
})

all_gdd <- unlist(lapply(1:length(all_result), function(i) {
  all_result[[i]]$accumulations$gdd
}))

sort(all_gdd)
ggplot(data.frame(x=all_gdd), aes(x)) + geom_density()
```

Out[22]:

```
1868.12999486923 1872.38499593735 1876.73499631882 1917.69499778748 1944.91999816895
1959.50000190735 1963.68999862671 1969.42500972748 1969.42500972748 1974.14000320435
2107.42000341415 2107.42000341415 2129.1350107193 2222.98000144958 2222.98000144958
```

In [23]:

```
selected_farmers <- (1:length(all_gdd))[all_gdd >= 1860 & all_gdd <= 1980]
selected_farmers_accgdd <- all_gdd[selected_farmers]

# Find the daily accumulative gdds for all of these selected farmers
selected_gdd <- unlist(lapply(1:length(selected_farmers), function(i) {
  farmer = selected_farmers[i]
  all_result[[farmer]]$dailyValues
}))

scaled_models <- do.call(cbind, lapply(1:length(selected_farmers), function(i) {
  multiplier <- selected_farmers_accgdd[i] / growth_stage_us$gddThreshold[[grep("R6", growth_stage_us$stage)]]
  growth_stage_us$gddThreshold*multiplier
}))

uganda_kilima_model <- cbind(growth_stage_us[, 1:4], gddThreshold = rowMeans(scaled_models))
uganda_kilima_model
write.xlsx(uganda_kilima_model, "uganda_kilima_model.xlsx")
```

Out [23] :

	gddUnits	id	stage	description	gddThreshold
1	C	stage1	VE	Emergence	68.01424
2	C	stage2	V2	Two leaves with collar visible	160.2113
3	C	stage3	V4	Four leaves with collar visible	264.4998
4	C	stage4	V6	Six leaves	356.6969
5	C	stage5	V8	Eight leaves	459.474
6	C	stage6	V10	Ten leaves	575.8539
7	C	stage7	V12	Twelve leaves	666.5396
8	C	stage8	V14	Fourteen leaves	758.7367
9	C	stage9	V16	Sixteen leaves	832.7966
10	C	stage10	V18	Eighteen leaves	858.4909
11	C	stage11	VT	Tasseling	920.4594
12	C	stage12	R1	Silking	1026.259
13	C	stage13	R2	Blister	1165.311
14	C	stage14	R3	Milk	1293.782
15	C	stage15	R4	dough stage	1438.879
16	C	stage16	R5	Dent	1644.433
17	C	stage17	R6	Physiological maturity	1931.605

source: <https://www.extension.purdue.edu/extmedia/nch/nch-18.html>