CPSC 444 Capstone Project

Kavya Puranam

2024-12-18

Introduction

Climate change and varying environmental conditions endanger crop stability. Understanding the factors that influence crop growth and being able to predict these changes are important for farmers and breeders alike.

Rice growth occurs in distinct phases such as vegetative, reproductive, and ripening. Vegetative is the initial growth stage from germination to panicle initiation, reproductive is Each phase is also sensitive to different environmental factors such as temperature and precipitation.

This project aims to analyze spatial autocorrelation of environmental values and growth phase durations, investigate the relationship between environmental covariates and rice growth phases using spatial models, and predict the impact of environmental changes on rice growth in different locations using geographically weighted regression (GWR) and kriging. Spatial autocorrelation is important due to building the foundational understanding of how the data values show there may be something of interest in certain locations that requires further analysis. GWR is a spatial regression method used to predict a variable by fitting a regression equation to every variable in the dataset. By modeling the effects of environmental covariates such as precipitation, temperature, potential evapotranspiration on the growing phases such as vegetative, reproductive, and ripening we can develop predictive models to forecast phase durations under different environmental scenarios.

We aim to bridge the gap between environmental science, agronomy, and spatial analytics through understanding rice growth and the impacts the environment can have in order to improve cultivation strategies and improve global food security initiatives.

Datasets

The field data encompasses multiple locations 15 separate locations spread across Africa and Asia with varying climates, soil types, and agricultural practices reflecting the diverse growing conditions faced by rice crop globally. This diversity ensures the robustness of the models developed and their applicability to different regions.

Environmental data such as weather data PP The total precipitation (mm), DPT Dew-point temperature at two meters (°C · d-1), PET Potential evapotranspiration (mm · d-1), VPD Vapour pressure deficit (kPa · d-1), TM Mean temperature at two meters (°C · d-1), TR Temperature range (0C · d-1), APAR All-sky surface photosynthetically active radiation total (W · m-2), and CPAR Clear sky surface photosynthetically active radiation total (W · m-2) are used to create the analysis. Growing season dates are also used to calculate the growing phase periods.

I used table S4 of the Multi-environment Genomic Selection in Rice Elite Breeding Lines study from the Supplementary Information of this study. This table shows the environmental covariates (ECs) of each environment throughout the whole growing season for the different phases: vegetative (VE), reproductive (RE), and ripening (RI).

For the model the key variables used were:

Response variable: Growth Phase Duration (either Vegetative, Reproductive, or Ripen) is the primary focus serving as the dependent variable in predictive modeling

Predictor Variables: Environmental Variables Precipitation (PP), temperature (TM), vapor pressure deficit (VPD), potential evapotranspiration (PET), all sky solar radiation (APAR) and clear sky solar radiation (CPAR).

Using these variables enables the development of robust predictive models to forecast optimal growth periods under varying environmental conditions and optimize planting stratefies for high-yield rice varieties.

Methods

Data preprocessing * Converting Easting and Northing values into Latitude * Calculating growing phase durations from dates * Normalizing all environmental covariates for interpretability

Use semi-variograms to analyze spatial autocorrelation of the environmental covariates and the growth phase durations.

Apply Geographically Weighted Regression to model spatially varying relationships between environmental covariates and growth durations and allows us to account for heterogeneity.

Kriging is used to predict rice growth phase durations in unmeasured locations to estimate theses values in regions where data is unavailable.

The expected results for this is that there is autocorrelation between each growth period especially with the vegetative growth phase as environmental factors would typically effect the length of germination periods. I predict that there would be less autocorrelation in the ripening phase as that may take up less resources or environmental inputs.

Results and Impact

According to the Semi-Variogram of the Vegetative Phase, the range is large as it is hard to tell where the sill starts. We can say that spatial correlation extends over a large distance since there is autocorrelation.

The Semi-Variogram of the Reproductive Phase is the most steep where it goes from less than 100 semi-variance to 200. This could mean that there is spatial dependence over the entire distance of increase which shows autocorrelation for these values as well.

The Semi-Variogram of the Ripening Phase has the least steep increase which means that it nearby points could be less correlated than the rest of the phases.

Overall, there seems to be correlation in location of the Vegetative Phase and Reproductive Phase for the rice plants .

The map for the effects of precipitation on the Vegetation phase shows significant clusters of areas where less precipitation effected reduced the duration of the Vegetative phase. Similarly for the other phases as well.

The map for the effects of mean temperature on the phases shows that certain clusters are effected negatively or positively by temperatures, which in turn increase or decrease the duration of the phase. (Values of estimated coefficients for temperature changed due to the selection of the optimal bandwidth).

Overall, the visualizations show there is significant autocorrelation with certain environmental variables which effect the duration of the growth phases of Oryza sativa. With this work, researchers can gain better insights into the principles of spatial statistics and it's effects on genomics. Using predictive analytics breeders can consider a changing environment in their understanding of genetics and implement new strategies for future breeding.

Future extensions of this can be used in research for Genotype-Environment Interaction Analysis where we can see how genetic traits change or interact with environmental factors which can enhance precision in breeding techniques for more climate and disease resistant crops.

Future studies can be also be done on the effects of environmental factors on depensation of rice crops by testing the growth rates or reproduction rates with varying environmental factors to see if a self-pollinating plant such as rice is effected significantly by the environment that could lead to depensation.

References

Gollini, I., Lu, B., Charlton, M., Brunsdon, C., & Harris, P. (2015). gwmodel: An r package for exploring spatial heterogeneity using geographically weighted models. Journal of Statistical Software, 63(17). https://doi.org/10.18637/jss.v063.i17

Nguyen, V.H., Morantte, R.I.Z., Lopena, V. et al. Multi-environment Genomic Selection in Rice Elite Breeding Lines. Rice 16, 7 (2023). https://doi.org/10.1186/s12284-023-00623-6

Spatial autocorrelation—An overview | sciencedirect topics. (n.d.). Retrieved December 18, 2024, from https://www.sciencedirect.com/topics/computer-science/spatial-autocorrelation#:~:text=The%20term%20spatial%20autocorrela

Getting the Dataset

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.3

## Warning: package 'ggplot2' was built under R version 4.3.2

## Warning: package 'readr' was built under R version 4.3.3

## Warning: package 'dplyr' was built under R version 4.3.3

## Warning: package 'forcats' was built under R version 4.3.3

## Warning: package 'lubridate' was built under R version 4.3.3

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
```

```
library(sp)
```

```
## Warning: package 'sp' was built under R version 4.3.3
```

```r
library(gstat)
```

```
## Warning: package 'gstat' was built under R version 4.3.3
```

```r
library(spdep)
```

```
## Warning: package 'spdep' was built under R version 4.3.3
```

```
## Loading required package: spData
```

```
## Warning: package 'spData' was built under R version 4.3.3
```

```
## Loading required package: sf
```

```
## Warning: package 'sf' was built under R version 4.3.3
```

```
## Linking to GEOS 3.11.2, GDAL 3.8.2, PROJ 9.3.1; sf_use_s2() is TRUE
```

```r
library(spatialreg)
```

```
## Warning: package 'spatialreg' was built under R version 4.3.3
```

```
## Loading required package: Matrix
##
## Attaching package: 'Matrix'
##
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
##
##
## Attaching package: 'spatialreg'
##
## The following objects are masked from 'package:spdep':
##
##     get.ClusterOption, get.coresOption, get.mcOption,
##     get.VerboseOption, get.ZeroPolicyOption, set.ClusterOption,
##     set.coresOption, set.mcOption, set.VerboseOption,
##     set.ZeroPolicyOption
```

```r
library(sf)

data <- read.csv("Assignment 8 Dataset.csv")
head(data)
```

```
##   Environment N_latitude E_longitude Transplating_date
## 1   BD-GZ-19W     239905      904019        2019-07-25
## 2   BD-NM-19W     242548      903018        2019-08-04
## 3   IN-HY-18W     175128      782744        2018-08-13
```

```
## 4   IN-CU-19W     204475      859399        2019-08-09
## 5   IN-HY-19D     175128      782744        2019-02-12
## 6   IN-HY-19W     175124      782754        2019-08-16
##   First_lines_flowered_of_50. Last_lines_flowered_of._50. Harvesting_date   PP
## 1                  2019-09-20                  2019-10-21      2019-11-05 1121
## 2                  2019-09-24                  2019-10-24      2019-11-10 1054
## 3                  2018-10-20                  2018-11-11      2018-11-27  198
## 4                  2019-10-01                  2019-10-25      2019-11-19  813
## 5                  2019-04-10                  2019-04-27      2019-05-19   26
## 6                  2019-10-08                  2019-10-28      2019-12-07  353
##     DPT   PET  VPD    TM    TR  APAR  CPAR PP_VE DPT_VE PET_VE VPD_VE TM_VE
## 1 25.15  7.35 0.56 27.45  5.74  9280 12462   690   26.5    7.7    0.5  28.5
## 2 24.76  7.21 0.58 27.22  5.66  8426 11228   542   26.4    7.9    0.6  28.8
## 3 17.05  8.12 1.53 25.20 11.96  9971 12181   197   19.4    8.0    1.3  25.5
## 4 24.27  6.91 0.57 26.68  5.91  8388 11830   497   25.8    6.7    0.5  27.8
## 5  9.57 10.83 3.76 31.02 16.24 11570 12252    11    8.9   10.4    3.4  29.3
## 6 20.25  7.38 0.79 23.91  8.72  9809 12587   284   21.7    7.6    0.8  25.2
##   TR_VE APAR_VE CPAR_VE PP_RE DPT_RE PET_RE VPD_RE TM_RE TR_RE APAR_RE CPAR_RE
## 1   5.2    5498    7640   286   24.2    7.1    0.6  26.8   6.1    2735    3535
## 2   5.1    4898    6625   321   23.6    6.7    0.6  26.3   5.7    2444    3189
## 3  10.5    6395    8312     0   14.3    8.6    1.9  24.9  14.3    2217    2410
## 4   4.9    4324    6957   238   24.1    7.7    0.6  26.6   6.0    2267    2814
## 5  16.8    6654    6948    13   11.3   11.5    4.1  32.9  16.0    2299    2460
## 6   7.4    4837    6677    58   21.4    7.0    0.5  23.8   7.0    1725    2284
##   PP_RI DPT_RI PET_RI VPD_RI TM_RI TR_RI APAR_RI CPAR_RI
## 1   150   22.2    6.5    0.6  24.8   6.9    1240    1501
## 2   270   21.7    5.7    0.5  24.0   6.8    1193    1617
## 3     0   10.3    8.2    2.1  24.2  15.2    1571    1673
## 4   121   21.4    6.6    0.6  24.4   7.8    1957    2281
## 5     3   10.2   11.3    4.5  34.2  14.8    2857    3109
## 6    12   17.8    7.3    0.9  22.3  11.2    3426    3840
```

```
#summary(data)
#ggplot(data, aes(x=PP)) + geom_histogram(bins=15) + theme_minimal() + ggtitle("Distribution of Precipi
```

First I need to convert the coordinates from Easting and Northing to Latitude and Longitude

```
utm_coords <- data.frame( data["N_latitude"], data["E_longitude"])
utm_coords
```

```
##    N_latitude E_longitude
## 1      239905      904019
## 2      242548      903018
## 3      175128      782744
## 4      204475      859399
## 5      175128      782744
## 6      175124      782754
## 7      166302      817444
## 8      212514      816296
## 9       -1388      349379
```

```
## 10      -6489      372841
## 11    -245261      330083
## 12    -264694      326372
## 13     141687     1212554
## 14     141687     1212554
## 15     -64348       37538
```

```r
#guess for zone
utm_crs <- "+proj=utm +zone=33 +datum=WGS84 +units=m +no_defs"
utm_sf <- st_as_sf(utm_coords, coords = c("E_longitude", "N_latitude"), crs = utm_crs)
lat_lon_sf <- st_transform(utm_sf, crs = 4326)
print(lat_lon_sf)
```

```
## Simple feature collection with 15 features and 0 fields
## Geometry type: POINT
## Dimension:     XY
## Bounding box:  xmin: 10.84742 ymin: -2.393861 xmax: 21.39174 ymax: 2.189989
## Geodetic CRS:  WGS 84
## First 10 features:
##                         geometry
## 1     POINT (18.63095 2.166103)
## 2     POINT (18.62203 2.189989)
## 3     POINT (17.54107 1.582866)
## 4     POINT (18.22978 1.846987)
## 5     POINT (17.54107 1.582866)
## 6      POINT (17.54116 1.58283)
## 7     POINT (17.85258 1.502707)
## 8     POINT (17.84289 1.920293)
## 9  POINT (13.64653 -0.01255414)
## 10 POINT (13.85733 -0.05869626)
```

```r
coords <- st_coordinates(lat_lon_sf)
data$N_latitude <- coords[, "Y"]
data$E_longitude <- coords[, "X"]
#check if N_lat and E_lon were changed into lat and lon values
head(data)
```

```
##   Environment N_latitude E_longitude Transplating_date
## 1   BD-GZ-19W   2.166103    18.63095        2019-07-25
## 2   BD-NM-19W   2.189989    18.62203        2019-08-04
## 3   IN-HY-18W   1.582866    17.54107        2018-08-13
## 4   IN-CU-19W   1.846987    18.22978        2019-08-09
## 5   IN-HY-19D   1.582866    17.54107        2019-02-12
## 6   IN-HY-19W   1.582830    17.54116        2019-08-16
##   First_lines_flowered_of_50. Last_lines_flowered_of._50. Harvesting_date   PP
## 1                  2019-09-20                  2019-10-21      2019-11-05 1121
## 2                  2019-09-24                  2019-10-24      2019-11-10 1054
## 3                  2018-10-20                  2018-11-11      2018-11-27  198
## 4                  2019-10-01                  2019-10-25      2019-11-19  813
## 5                  2019-04-10                  2019-04-27      2019-05-19   26
## 6                  2019-10-08                  2019-10-28      2019-12-07  353
##     DPT   PET  VPD    TM    TR  APAR  CPAR PP_VE DPT_VE PET_VE VPD_VE TM_VE
## 1 25.15  7.35 0.56 27.45  5.74  9280 12462   690   26.5    7.7    0.5  28.5
```

```
## 2 24.76  7.21 0.58 27.22  5.66  8426 11228   542   26.4    7.9    0.6  28.8
## 3 17.05  8.12 1.53 25.20 11.96  9971 12181   197   19.4    8.0    1.3  25.5
## 4 24.27  6.91 0.57 26.68  5.91  8388 11830   497   25.8    6.7    0.5  27.8
## 5  9.57 10.83 3.76 31.02 16.24 11570 12252    11    8.9   10.4    3.4  29.3
## 6 20.25  7.38 0.79 23.91  8.72  9809 12587   284   21.7    7.6    0.8  25.2
##   TR_VE APAR_VE CPAR_VE PP_RE DPT_RE PET_RE VPD_RE TM_RE TR_RE APAR_RE CPAR_RE
## 1   5.2    5498    7640   286   24.2    7.1    0.6  26.8   6.1    2735    3535
## 2   5.1    4898    6625   321   23.6    6.7    0.6  26.3   5.7    2444    3189
## 3  10.5    6395    8312     0   14.3    8.6    1.9  24.9  14.3    2217    2410
## 4   4.9    4324    6957   238   24.1    7.7    0.6  26.6   6.0    2267    2814
## 5  16.8    6654    6948    13   11.3   11.5    4.1  32.9  16.0    2299    2460
## 6   7.4    4837    6677    58   21.4    7.0    0.5  23.8   7.0    1725    2284
##   PP_RI DPT_RI PET_RI VPD_RI TM_RI TR_RI APAR_RI CPAR_RI
## 1   150   22.2    6.5    0.6  24.8   6.9    1240    1501
## 2   270   21.7    5.7    0.5  24.0   6.8    1193    1617
## 3     0   10.3    8.2    2.1  24.2  15.2    1571    1673
## 4   121   21.4    6.6    0.6  24.4   7.8    1957    2281
## 5     3   10.2   11.3    4.5  34.2  14.8    2857    3109
## 6    12   17.8    7.3    0.9  22.3  11.2    3426    3840
```

Calculate phase durations for transplating, flowering, and harvesting dates

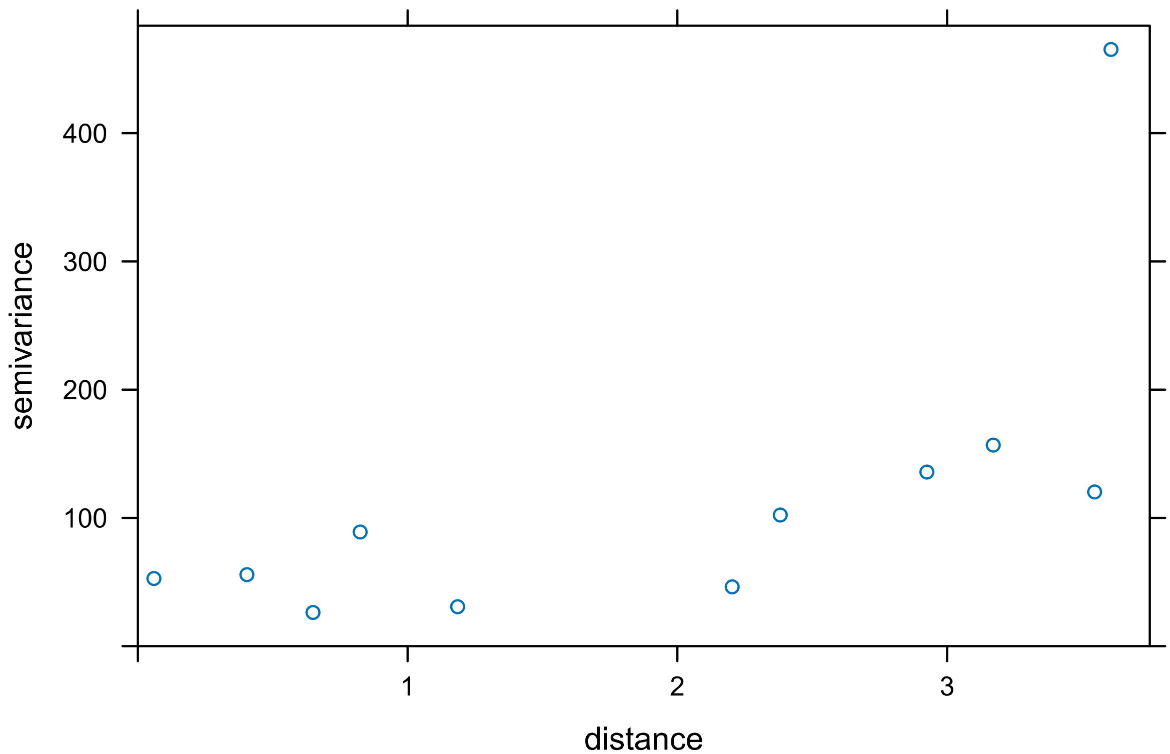
```r
# Convert dates to Date objects
data$Transplating_date <- as.Date(data$Transplating_date, format = "%Y-%m-%d")
data$Harvesting_date <- as.Date(data$Harvesting_date, format = "%Y-%m-%d")
data$First_lines_flowered_of_50. <-
as.Date(data$First_lines_flowered_of_50., format = "%Y-%m-%d")
data$Last_lines_flowered_of._50. <-
as.Date(data$Last_lines_flowered_of._50., format = "%Y-%m-%d")


# Calculate phase durations
data$VE_Duration <- as.numeric(data$First_lines_flowered_of_50. - data$Transplating_date)
data$RE_Duration <- as.numeric(data$Last_lines_flowered_of._50. - data$First_lines_flowered_of_50.)
data$RI_Duration <- as.numeric(data$Harvesting_date - data$Last_lines_flowered_of._50.)
data$Total_Duration <- as.numeric(data$Harvesting_date - data$Transplating_date)

#normalize environmental covariates
#do this because we want features to be on a similar scale
normalize <- function(x) { (x - min(x)) / (max(x) - min(x)) }
data <- data %>%
  mutate(across(c(PP, DPT, PET, VPD, TM, TR, APAR, CPAR), normalize))
```

```r
coordinates(data) <- ~E_longitude + N_latitude
```

Semi Variogram Analysis

Now I will generate a semi-variogram on an environmental covariate

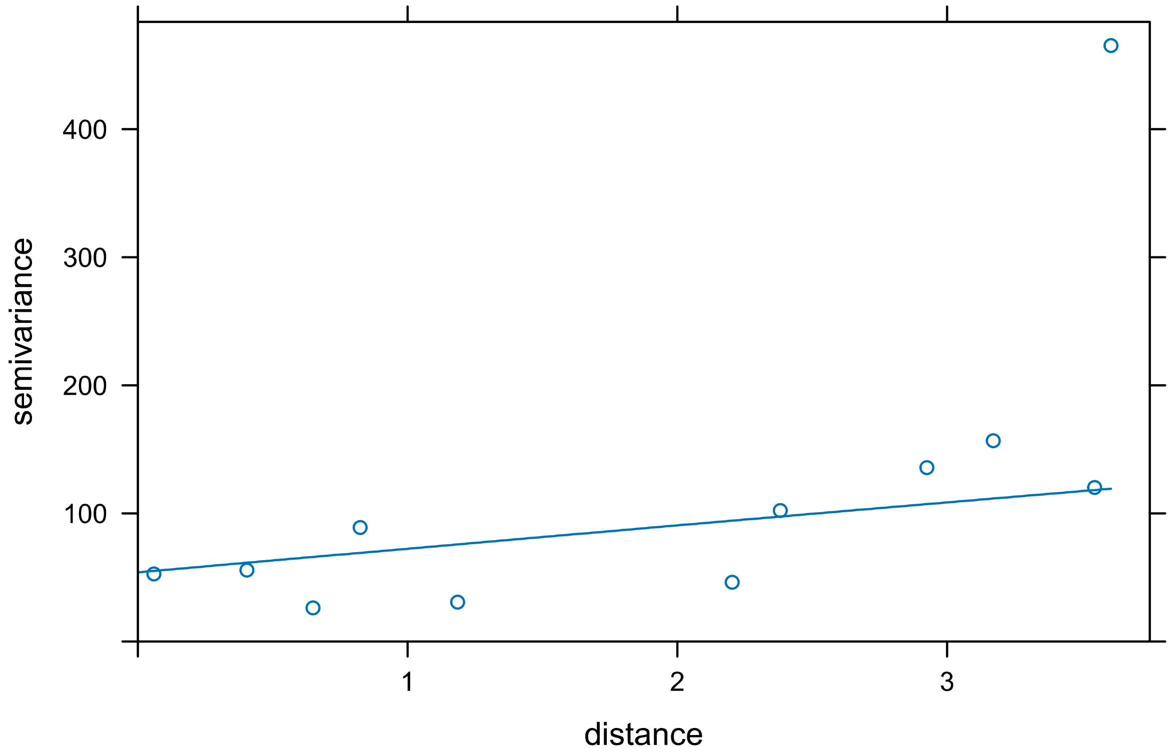
```r
#generating a semi-variogram on a selected environmental covariate
variogram_pp <- variogram(VE_Duration ~ 1, data)
plot(variogram_pp)
```



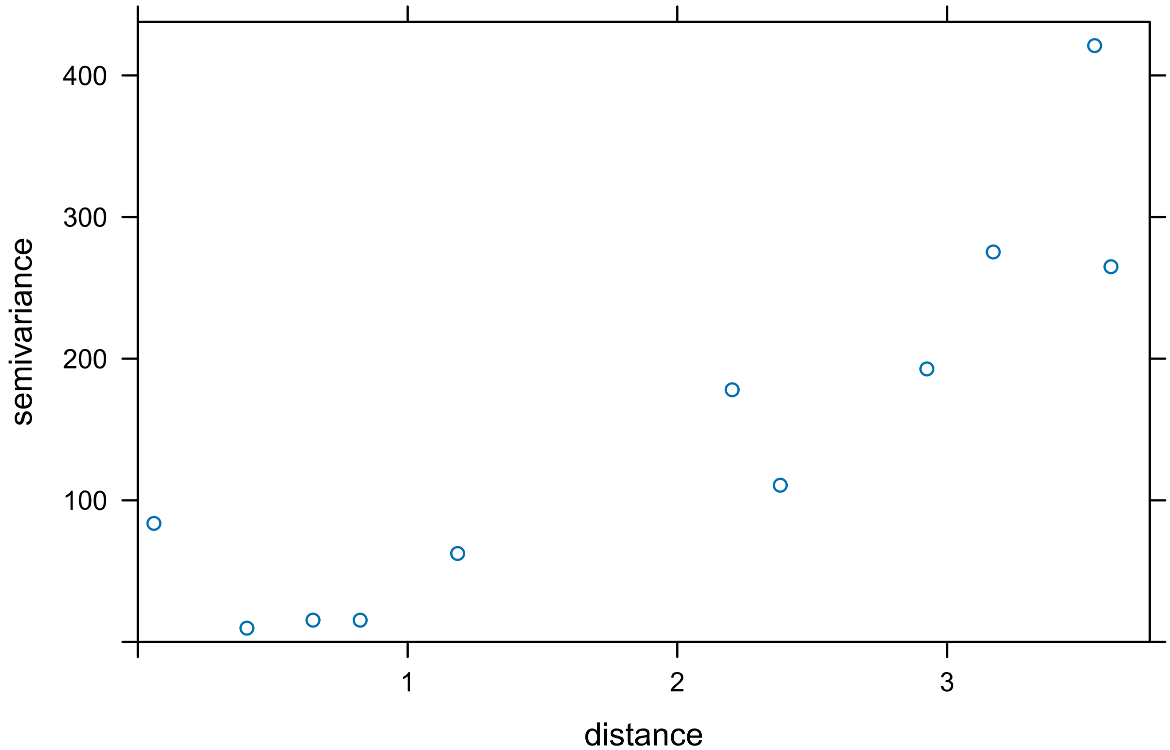

```r
variogram_model <- fit.variogram(variogram_pp, model=vgm("Sph"))
```

```
## Warning in fit.variogram(variogram_pp, model = vgm("Sph")): No convergence
## after 200 iterations: try different initial values?
```

```r
plot(variogram_pp, model=variogram_model)
```



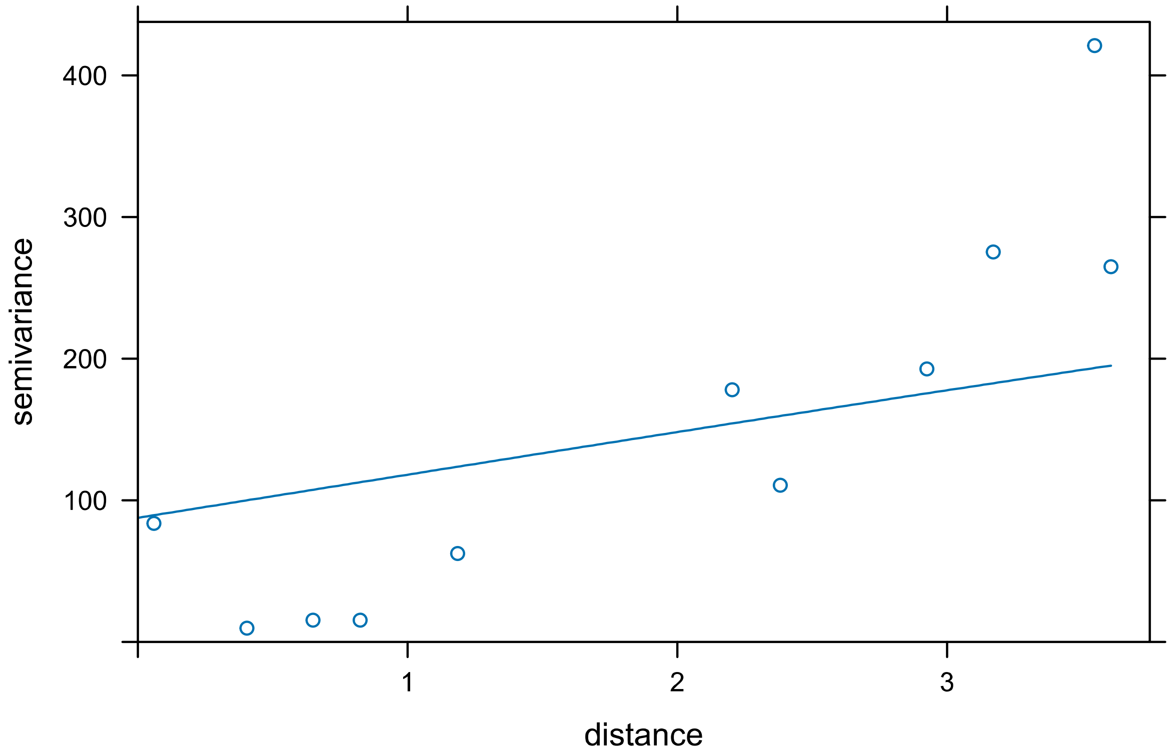

```
#generating a semi-variogram on a selected environmental covariate
variogram_pp <- variogram(RE_Duration ~ 1, data)
plot(variogram_pp)
```



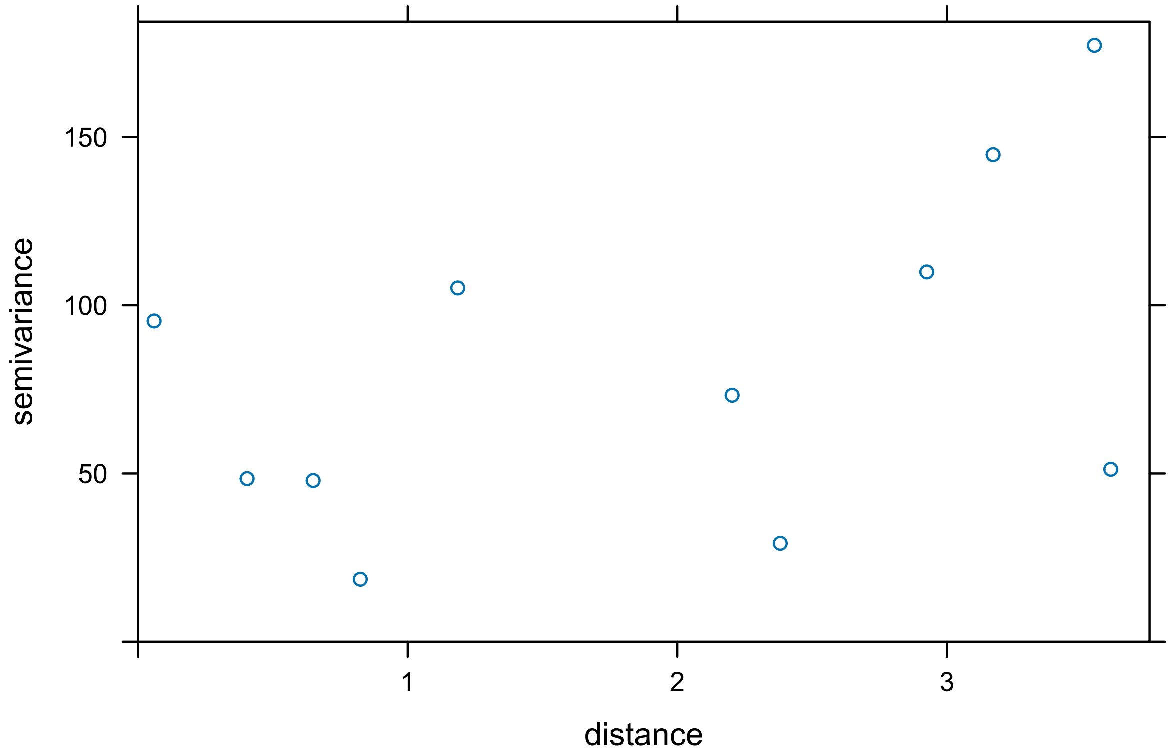

```r
variogram_model <- fit.variogram(variogram_pp, model=vgm("Sph"))
```

```
## Warning in fit.variogram(variogram_pp, model = vgm("Sph")): No convergence
## after 200 iterations: try different initial values?
```

```r
plot(variogram_pp, model=variogram_model)
```



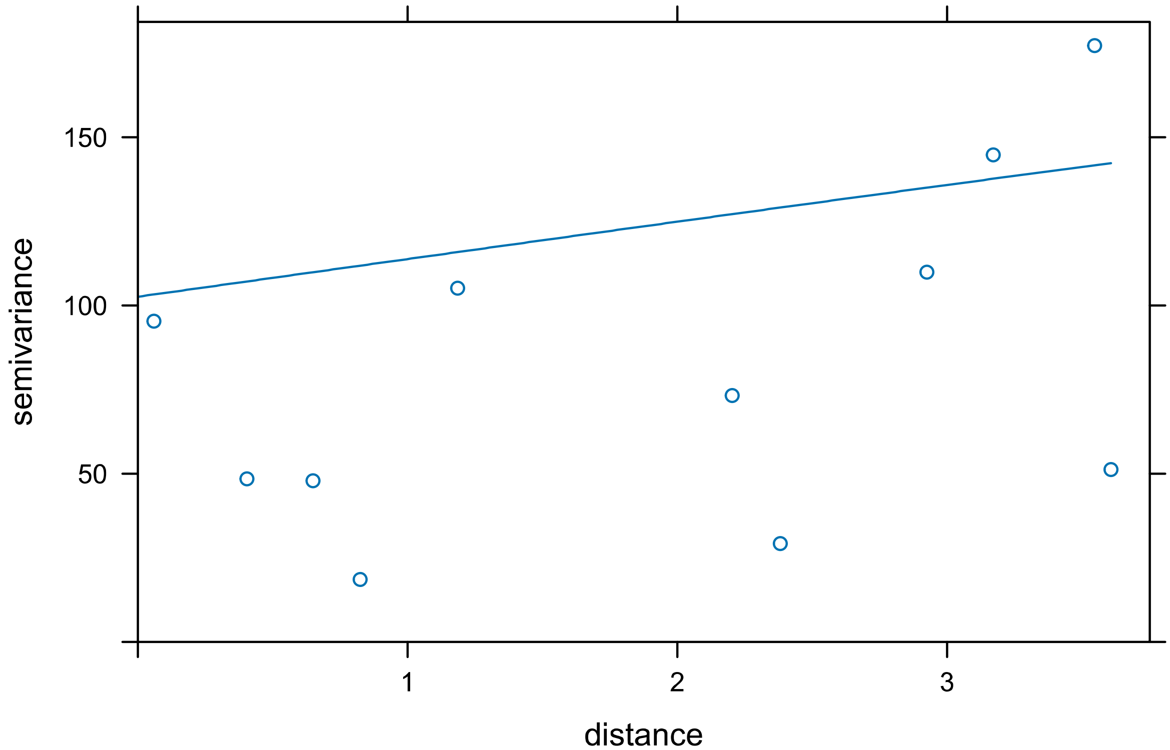

```r
#generating a semi-variogram on a selected environmental covariate
variogram_pp <- variogram(RI_Duration ~ 1, data)
plot(variogram_pp)
```




```
variogram_model <- fit.variogram(variogram_pp, model=vgm("Sph"))
```

```
## Warning in fit.variogram(variogram_pp, model = vgm("Sph")): No convergence
## after 200 iterations: try different initial values?
```

```
plot(variogram_pp, model=variogram_model)
```



Geographically Weighted Regression

Use GWR model to understand spatially varying effects of Precipitation (PP) on Vegetative phase duration (VE)

```
library(spgwr)
```

```
## Warning: package 'spgwr' was built under R version 4.3.3
```

```
## NOTE: This package does not constitute approval of GWR
## as a method of spatial analysis; see example(gwr)
```

```
gwr_bandwidth <- gwr.sel(VE_Duration ~ PP + DPT + PET + VPD + TM + TR + APAR + CPAR,
                         data = data, adapt = TRUE)
```

```
## Adaptive q: 0.381966 CV score: 824.5573
## Adaptive q: 0.618034 CV score: 1453.295
## Adaptive q: 0.236068 CV score: NA
```

```
## Warning in optimize(gwr.cv.adapt.f, lower = beta1, upper = beta2, maximum =
## FALSE, : NA/Inf replaced by maximum positive value
```

```
## Adaptive q: 0.5 CV score: 833.4677
## Adaptive q: 0.4392614 CV score: 832.617
## Adaptive q: 0.3262379 CV score: 840.0758
## Adaptive q: 0.4038509 CV score: 831.3833
## Adaptive q: 0.3606798 CV score: 820.4553
## Adaptive q: 0.3364241 CV score: 814.4616
## Adaptive q: 0.345689 CV score: 816.499
## Adaptive q: 0.3325333 CV score: 816.6414
## Adaptive q: 0.3392024 CV score: 815.009
## Adaptive q: 0.334938 CV score: 814.1961
## Adaptive q: 0.3340195 CV score: 814.0424
## Adaptive q: 0.3334518 CV score: 813.9514
## Adaptive q: 0.333101 CV score: 814.7137
## Adaptive q: 0.3336687 CV score: 813.9858
## Adaptive q: 0.3333178 CV score: 813.9848
## Adaptive q: 0.3334925 CV score: 813.9579
## Adaptive q: 0.3334111 CV score: 813.945
## Adaptive q: 0.3333704 CV score: 813.9387
## Adaptive q: 0.3333704 CV score: 813.9387
```

```r
ve_gwr_model <- gwr(VE_Duration ~ PP + DPT + PET + VPD + TM + TR + APAR + CPAR, data=data, adapt = gwr_

summary(ve_gwr_model)
```

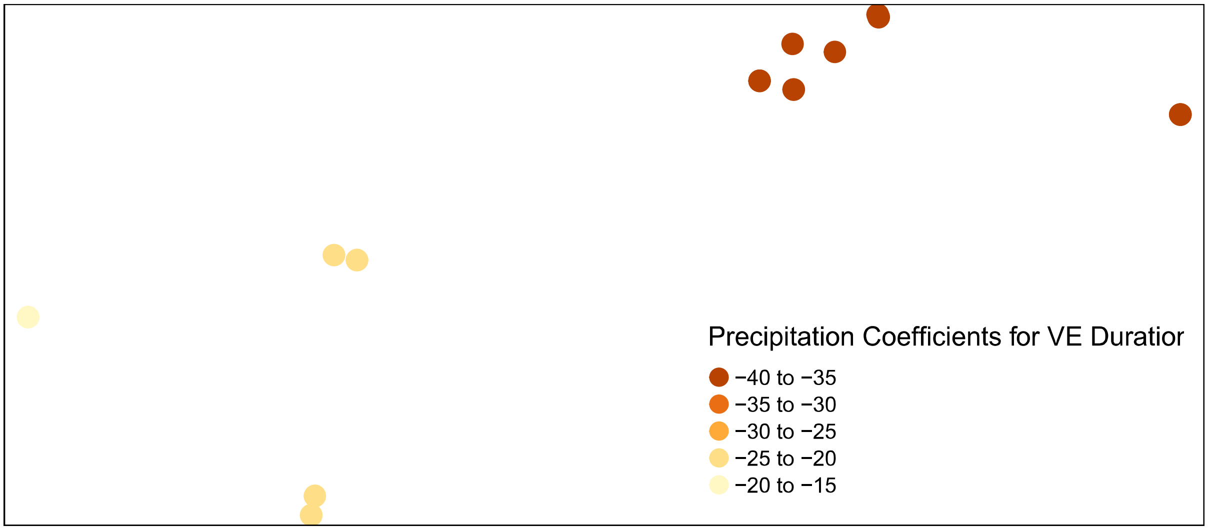
```
##           Length Class                  Mode
## SDF       15     SpatialPointsDataFrame S4
## lhat       1     -none-                 logical
## lm        11     -none-                 list
## results    0     -none-                 NULL
## bandwidth 15     -none-                 numeric
## adapt      1     -none-                 numeric
## hatmatrix  1     -none-                 logical
## gweight    1     -none-                 character
## gTSS       1     -none-                 numeric
## this.call  4     -none-                 call
## fp.given   1     -none-                 logical
## timings    8     -none-                 numeric
```

```r
# Visualize the geographically varying coefficients
library(tmap)
```

```
## Warning: package 'tmap' was built under R version 4.3.3

## Breaking News: tmap 3.x is retiring. Please test v4, e.g. with
## remotes::install_github('r-tmap/tmap')
```

```r
tm_shape(ve_gwr_model$SDF) +
  tm_dots(col = "PP", size = 0.5, title = "Precipitation Coefficients for VE Duration")
```

```
## Warning: Currect projection of shape ve_gwr_model$SDF unknown. Long-lat (WGS84)
## is assumed.
```



Use GWR model to understand spatially varying effects of Precipitation (PP) on Reproductive phase duration (RE)

```
library(spgwr)
gwr_bandwidth <- gwr.sel(RE_Duration ~ PP + DPT + PET + VPD + TM + TR + APAR + CPAR,
                         data = data, adapt = TRUE)
```

```
## Adaptive q: 0.381966 CV score: 5994.182
## Adaptive q: 0.618034 CV score: 5152.561
## Adaptive q: 0.763932 CV score: 4829.267
## Adaptive q: 0.854102 CV score: 4535.362
## Adaptive q: 0.9098301 CV score: 4377.136
## Adaptive q: 0.9442719 CV score: 4316.758
## Adaptive q: 0.9655581 CV score: 4291.064
## Adaptive q: 0.9787138 CV score: 4278.685
## Adaptive q: 0.9868444 CV score: 4272.062
## Adaptive q: 0.9918694 CV score: 4268.305
## Adaptive q: 0.994975 CV score: 4266.101
## Adaptive q: 0.9968944 CV score: 4264.781
## Adaptive q: 0.9980806 CV score: 4263.981
## Adaptive q: 0.9988138 CV score: 4263.493
## Adaptive q: 0.9992669 CV score: 4263.193
## Adaptive q: 0.9995469 CV score: 4263.009
```

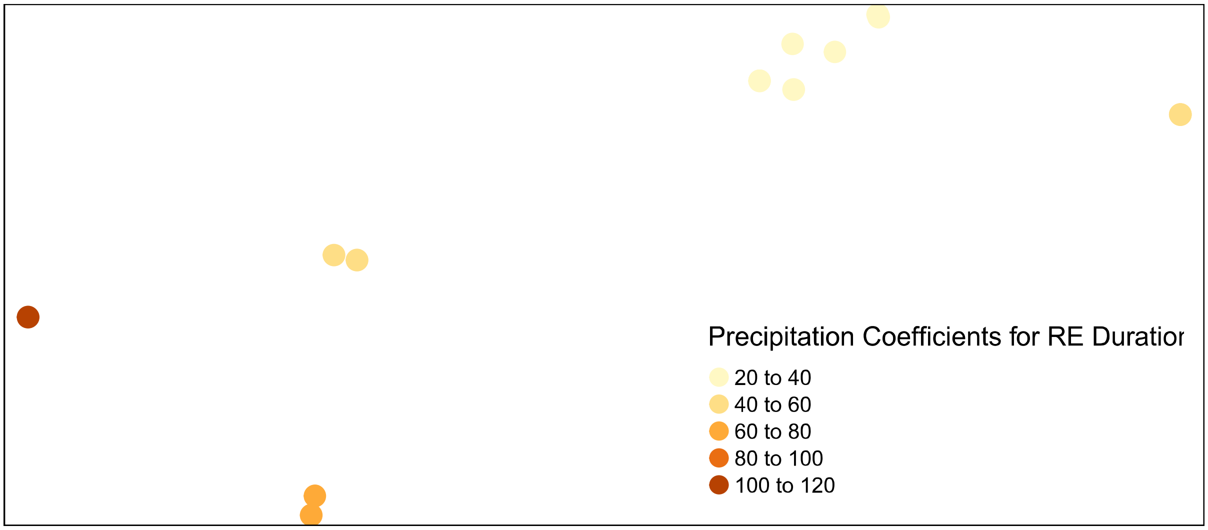
```
## Adaptive q: 0.99972 CV score: 4262.895
## Adaptive q: 0.9998269 CV score: 4262.825
## Adaptive q: 0.999893 CV score: 4262.782
## Adaptive q: 0.9999339 CV score: 4262.755
## Adaptive q: 0.9999339 CV score: 4262.755

## Warning in gwr.sel(RE_Duration ~ PP + DPT + PET + VPD + TM + TR + APAR + :
## Bandwidth converged to upper bound:1
```

```
gwr_model <- gwr(RE_Duration ~ PP + DPT + PET + VPD + TM + TR + APAR + CPAR, data=data, bandwidth = gwr

# Print the results of the GWR model
summary(gwr_model)
```

```
##            Length Class                  Mode
## SDF        15     SpatialPointsDataFrame S4
## lhat        1     -none-                 logical
## lm         11     -none-                 list
## results     0     -none-                 NULL
## bandwidth  1      -none-                 numeric
## adapt      0      -none-                 NULL
## hatmatrix  1      -none-                 logical
## gweight    1      -none-                 character
## gTSS       1      -none-                 numeric
## this.call  4      -none-                 call
## fp.given   1      -none-                 logical
## timings    8      -none-                 numeric
```

```
# Visualize the geographically varying coefficients
library(tmap)
tm_shape(gwr_model$SDF) +
  tm_dots(col = "PP", size = 0.5, title = "Precipitation Coefficients for RE Duration")
```

```
## Warning: Currect projection of shape gwr_model$SDF unknown. Long-lat (WGS84) is
## assumed.
```



USe GWR model to understand spatially varying effects of Precipitation (PP) on Ripen phase duration (RI)

```
library(spgwr)

gwr_bandwidth <- gwr.sel(RI_Duration ~ PP + DPT + PET + VPD + TM + TR + APAR + CPAR,
                         data = data, adapt = TRUE)
```

```
## Adaptive q: 0.381966 CV score: 12704.28
## Adaptive q: 0.618034 CV score: 6215.218
## Adaptive q: 0.763932 CV score: 6018.045
## Adaptive q: 0.7008581 CV score: 6066.192
## Adaptive q: 0.7861467 CV score: 6017.307
## Adaptive q: 0.776979 CV score: 6017.546
## Adaptive q: 0.8678314 CV score: 6076.592
## Adaptive q: 0.8055156 CV score: 6018.282
## Adaptive q: 0.7861874 CV score: 6017.306
## Adaptive q: 0.7935701 CV score: 6017.178
## Adaptive q: 0.7981329 CV score: 6017.127
## Adaptive q: 0.8009529 CV score: 6017.199
## Adaptive q: 0.7969733 CV score: 6017.138
## Adaptive q: 0.79921 CV score: 6017.118
## Adaptive q: 0.7998757 CV score: 6017.113
```

```
## Adaptive q: 0.8002872 CV score: 6017.133
## Adaptive q: 0.7996215 CV score: 6017.115
## Adaptive q: 0.8000329 CV score: 6017.114
## Adaptive q: 0.799835 CV score: 6017.113
## Adaptive q: 0.7999164 CV score: 6017.113
## Adaptive q: 0.7999609 CV score: 6017.112
## Adaptive q: 0.7999609 CV score: 6017.112
```

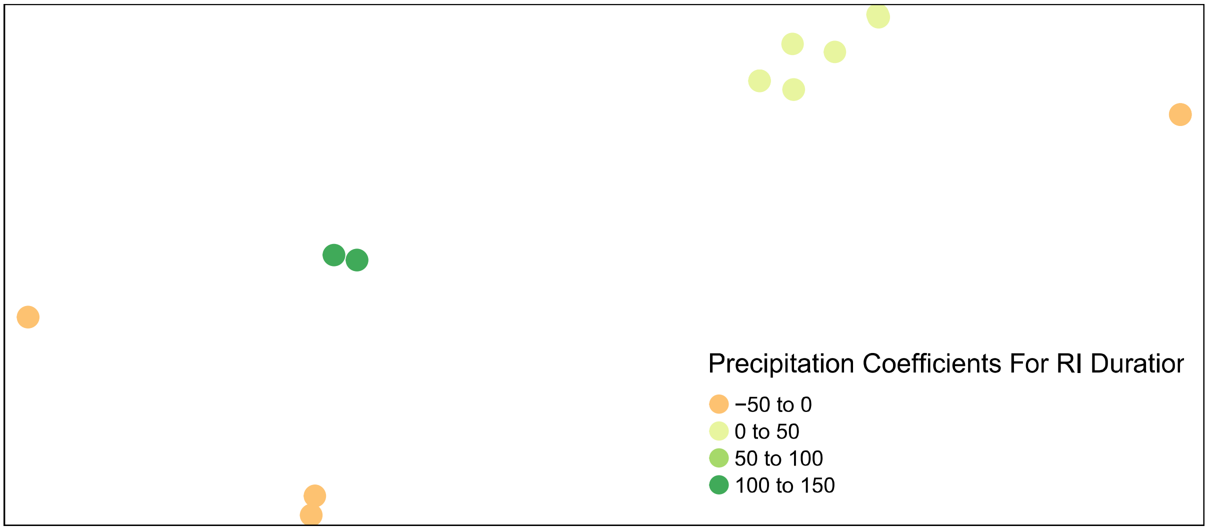
```
gwr_model <- gwr(RI_Duration ~ PP + DPT + PET + VPD + TM + TR + APAR + CPAR, data=data, bandwidth = gwr

# Print the results of the GWR model
summary(gwr_model)
```

```
##           Length Class                  Mode
## SDF       15     SpatialPointsDataFrame S4
## lhat       1     -none-                 logical
## lm        11     -none-                 list
## results    0     -none-                 NULL
## bandwidth  1     -none-                 numeric
## adapt      0     -none-                 NULL
## hatmatrix  1     -none-                 logical
## gweight    1     -none-                 character
## gTSS       1     -none-                 numeric
## this.call  4     -none-                 call
## fp.given   1     -none-                 logical
## timings    8     -none-                 numeric
```

```
# Visualize the geographically varying coefficients
library(tmap)
tm_shape(gwr_model$SDF) +
  tm_dots(col = "PP", size = 0.5, title = "Precipitation Coefficients For RI Duration")
```

```
## Warning: Currect projection of shape gwr_model$SDF unknown. Long-lat (WGS84) is
## assumed.
```

```
## Variable(s) "PP" contains positive and negative values, so midpoint is set to 0. Set midpoint = NA t
```



Use GWR model to understand spatially varying effects of Temperature Mean (TM) on Vegetative phase duration (VE)

```r
library(spgwr)

gwr_bandwidth <- gwr.sel(VE_Duration ~ PP + DPT + PET + VPD + TM + TR + APAR + CPAR,
                         data = data, adapt = TRUE)
```

```
## Adaptive q: 0.381966 CV score: 824.5573
## Adaptive q: 0.618034 CV score: 1453.295
## Adaptive q: 0.236068 CV score: NA

## Warning in optimize(gwr.cv.adapt.f, lower = beta1, upper = beta2, maximum =
## FALSE, : NA/Inf replaced by maximum positive value

## Adaptive q: 0.5 CV score: 833.4677
## Adaptive q: 0.4392614 CV score: 832.617
## Adaptive q: 0.3262379 CV score: 840.0758
## Adaptive q: 0.4038509 CV score: 831.3833
## Adaptive q: 0.3606798 CV score: 820.4553
## Adaptive q: 0.3364241 CV score: 814.4616
## Adaptive q: 0.345689 CV score: 816.499
## Adaptive q: 0.3325333 CV score: 816.6414
```

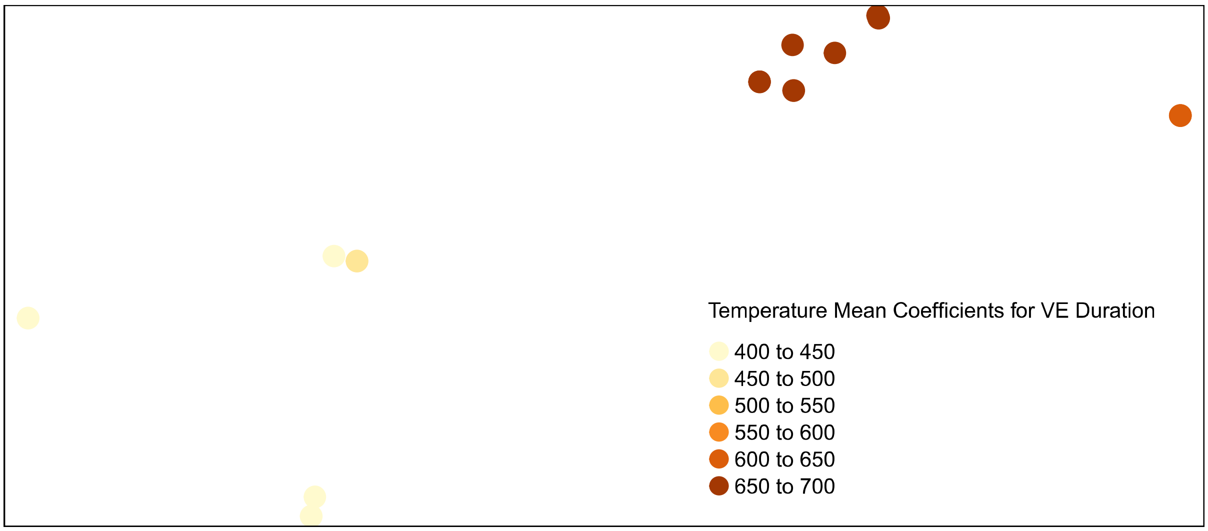
```
## Adaptive q: 0.3392024 CV score: 815.009
## Adaptive q: 0.334938 CV score: 814.1961
## Adaptive q: 0.3340195 CV score: 814.0424
## Adaptive q: 0.3334518 CV score: 813.9514
## Adaptive q: 0.333101 CV score: 814.7137
## Adaptive q: 0.3336687 CV score: 813.9858
## Adaptive q: 0.3333178 CV score: 813.9848
## Adaptive q: 0.3334925 CV score: 813.9579
## Adaptive q: 0.3334111 CV score: 813.945
## Adaptive q: 0.3333704 CV score: 813.9387
## Adaptive q: 0.3333704 CV score: 813.9387
```

```
gwr_model <- gwr(VE_Duration ~ PP + DPT + PET + VPD + TM + TR + APAR + CPAR, data=data, adapt = gwr_ban

# Print the results of the GWR model
summary(gwr_model)
```

```
##           Length Class                  Mode
## SDF       15     SpatialPointsDataFrame S4
## lhat       1     -none-                 logical
## lm        11     -none-                 list
## results    0     -none-                 NULL
## bandwidth 15     -none-                 numeric
## adapt      1     -none-                 numeric
## hatmatrix  1     -none-                 logical
## gweight    1     -none-                 character
## gTSS       1     -none-                 numeric
## this.call  4     -none-                 call
## fp.given   1     -none-                 logical
## timings    8     -none-                 numeric
```

```
# Visualize the geographically varying coefficients
#plot(gwr_model$SDF)
library(tmap)
tm_shape(gwr_model$SDF) +
  tm_dots(col = "TM", size = 0.5, title = "Temperature Mean Coefficients for VE Duration")
```

```
## Warning: Currect projection of shape gwr_model$SDF unknown. Long-lat (WGS84) is
## assumed.
```



Use GWR model to understand spatially varying effects of Temperature Mean (TM) on Reproductive phase duration (RE)

```r
library(spgwr)

gwr_bandwidth <- gwr.sel(RE_Duration ~ PP + DPT + PET + VPD + TM + TR + APAR + CPAR,
                         data = data, adapt = TRUE)
```

```
## Adaptive q: 0.381966 CV score: 5994.182
## Adaptive q: 0.618034 CV score: 5152.561
## Adaptive q: 0.763932 CV score: 4829.267
## Adaptive q: 0.854102 CV score: 4535.362
## Adaptive q: 0.9098301 CV score: 4377.136
## Adaptive q: 0.9442719 CV score: 4316.758
## Adaptive q: 0.9655581 CV score: 4291.064
## Adaptive q: 0.9787138 CV score: 4278.685
## Adaptive q: 0.9868444 CV score: 4272.062
## Adaptive q: 0.9918694 CV score: 4268.305
## Adaptive q: 0.994975 CV score: 4266.101
## Adaptive q: 0.9968944 CV score: 4264.781
## Adaptive q: 0.9980806 CV score: 4263.981
## Adaptive q: 0.9988138 CV score: 4263.493
## Adaptive q: 0.9992669 CV score: 4263.193
```

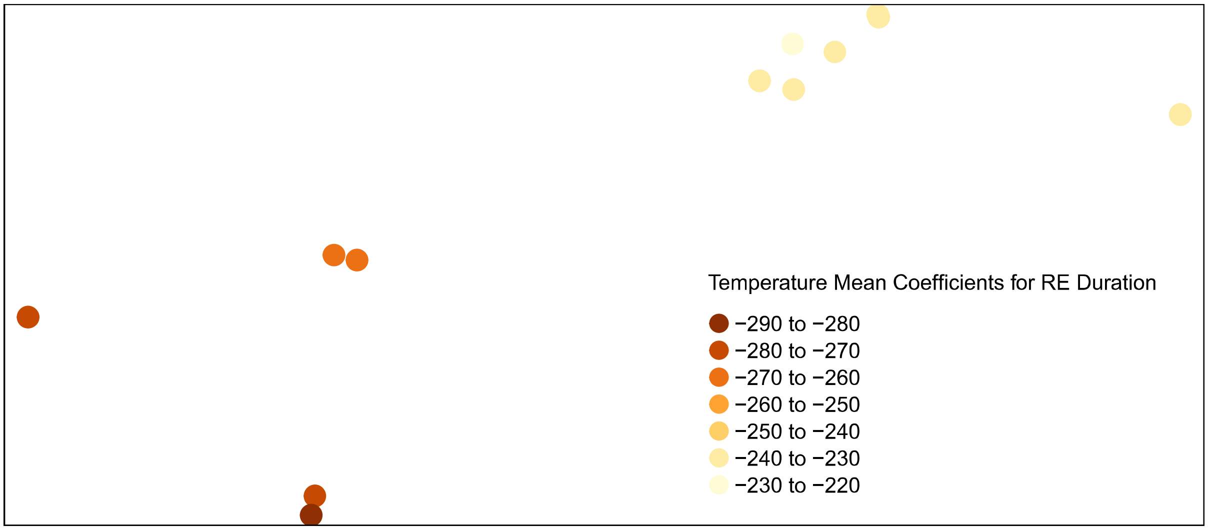
```
## Adaptive q: 0.9995469 CV score: 4263.009
## Adaptive q: 0.99972 CV score: 4262.895
## Adaptive q: 0.9998269 CV score: 4262.825
## Adaptive q: 0.999893 CV score: 4262.782
## Adaptive q: 0.9999339 CV score: 4262.755
## Adaptive q: 0.9999339 CV score: 4262.755

## Warning in gwr.sel(RE_Duration ~ PP + DPT + PET + VPD + TM + TR + APAR + :
## Bandwidth converged to upper bound:1
```

```
gwr_model <- gwr(RE_Duration ~ PP + DPT + PET + VPD + TM + TR + APAR + CPAR, data=data, adapt = gwr_ban

# Print the results of the GWR model
summary(gwr_model)
```

```
##           Length Class                  Mode
## SDF       15     SpatialPointsDataFrame S4
## lhat       1     -none-                 logical
## lm        11     -none-                 list
## results    0     -none-                 NULL
## bandwidth 15     -none-                 numeric
## adapt      1     -none-                 numeric
## hatmatrix  1     -none-                 logical
## gweight    1     -none-                 character
## gTSS       1     -none-                 numeric
## this.call  4     -none-                 call
## fp.given   1     -none-                 logical
## timings    8     -none-                 numeric
```

```
# Visualize the geographically varying coefficients
#plot(gwr_model$SDF)
library(tmap)
tm_shape(gwr_model$SDF) +
  tm_dots(col = "TM", size = 0.5, title = "Temperature Mean Coefficients for RE Duration")
```

```
## Warning: Currect projection of shape gwr_model$SDF unknown. Long-lat (WGS84) is
## assumed.
```



Use GWR model to understand spatially varying effects of Temperature Mean (TM) on Ripen phase duration (RI)

```
library(spgwr)

gwr_bandwidth <- gwr.sel(RI_Duration ~ PP + DPT + PET + VPD + TM + TR + APAR + CPAR,
                         data = data, adapt = TRUE)
```

```
## Adaptive q: 0.381966 CV score: 12704.28
## Adaptive q: 0.618034 CV score: 6215.218
## Adaptive q: 0.763932 CV score: 6018.045
## Adaptive q: 0.7008581 CV score: 6066.192
## Adaptive q: 0.7861467 CV score: 6017.307
## Adaptive q: 0.776979 CV score: 6017.546
## Adaptive q: 0.8678314 CV score: 6076.592
## Adaptive q: 0.8055156 CV score: 6018.282
## Adaptive q: 0.7861874 CV score: 6017.306
## Adaptive q: 0.7935701 CV score: 6017.178
## Adaptive q: 0.7981329 CV score: 6017.127
## Adaptive q: 0.8009529 CV score: 6017.199
## Adaptive q: 0.7969733 CV score: 6017.138
## Adaptive q: 0.79921 CV score: 6017.118
## Adaptive q: 0.7998757 CV score: 6017.113
```

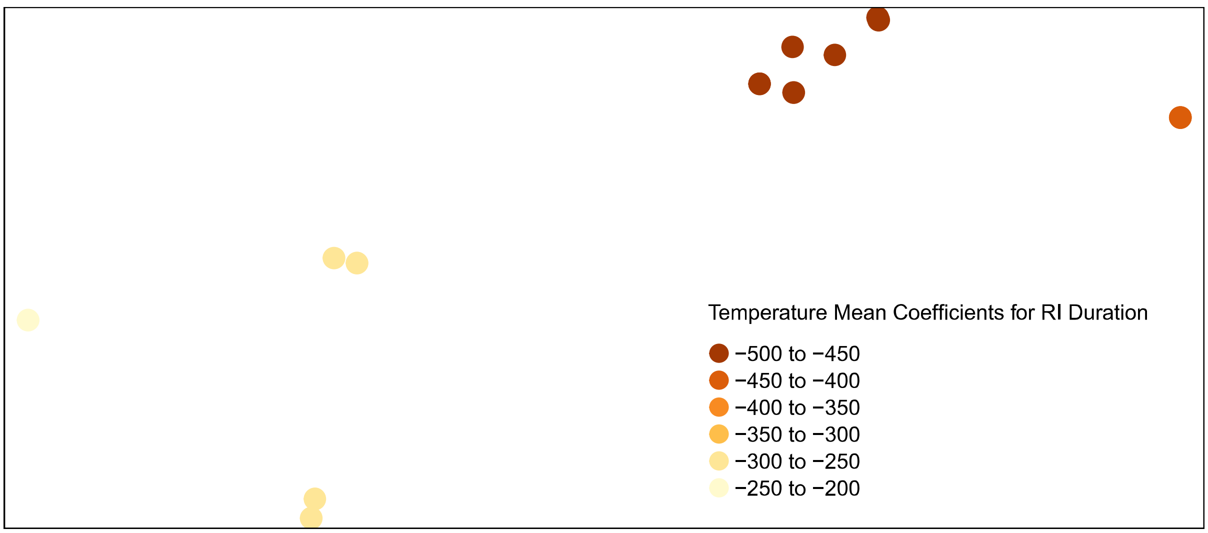
```
## Adaptive q: 0.8002872 CV score: 6017.133
## Adaptive q: 0.7996215 CV score: 6017.115
## Adaptive q: 0.8000329 CV score: 6017.114
## Adaptive q: 0.799835 CV score: 6017.113
## Adaptive q: 0.7999164 CV score: 6017.113
## Adaptive q: 0.7999609 CV score: 6017.112
## Adaptive q: 0.7999609 CV score: 6017.112
```

```
gwr_model <- gwr(RI_Duration ~ PP + DPT + PET + VPD + TM + TR + APAR + CPAR, data=data, adapt = gwr_ban

# Print the results of the GWR model
summary(gwr_model)
```

```
##           Length Class                  Mode
## SDF       15     SpatialPointsDataFrame S4
## lhat       1     -none-                 logical
## lm        11     -none-                 list
## results    0     -none-                 NULL
## bandwidth 15     -none-                 numeric
## adapt      1     -none-                 numeric
## hatmatrix  1     -none-                 logical
## gweight    1     -none-                 character
## gTSS       1     -none-                 numeric
## this.call  4     -none-                 call
## fp.given   1     -none-                 logical
## timings    8     -none-                 numeric
```

```
# Visualize the geographically varying coefficients
#plot(gwr_model$SDF)
library(tmap)
tm_shape(gwr_model$SDF) +
tm_dots(col = "TM", size = 0.5, title = "Temperature Mean Coefficients for RI Duration")
```

```
## Warning: Currect projection of shape gwr_model$SDF unknown. Long-lat (WGS84) is
## assumed.
```



Map the Mean Temperature on Growth Phase Duration Visually on a Global Map

```
data_df <- as.data.frame(data)
library(ggplot2)
library(maps)
```

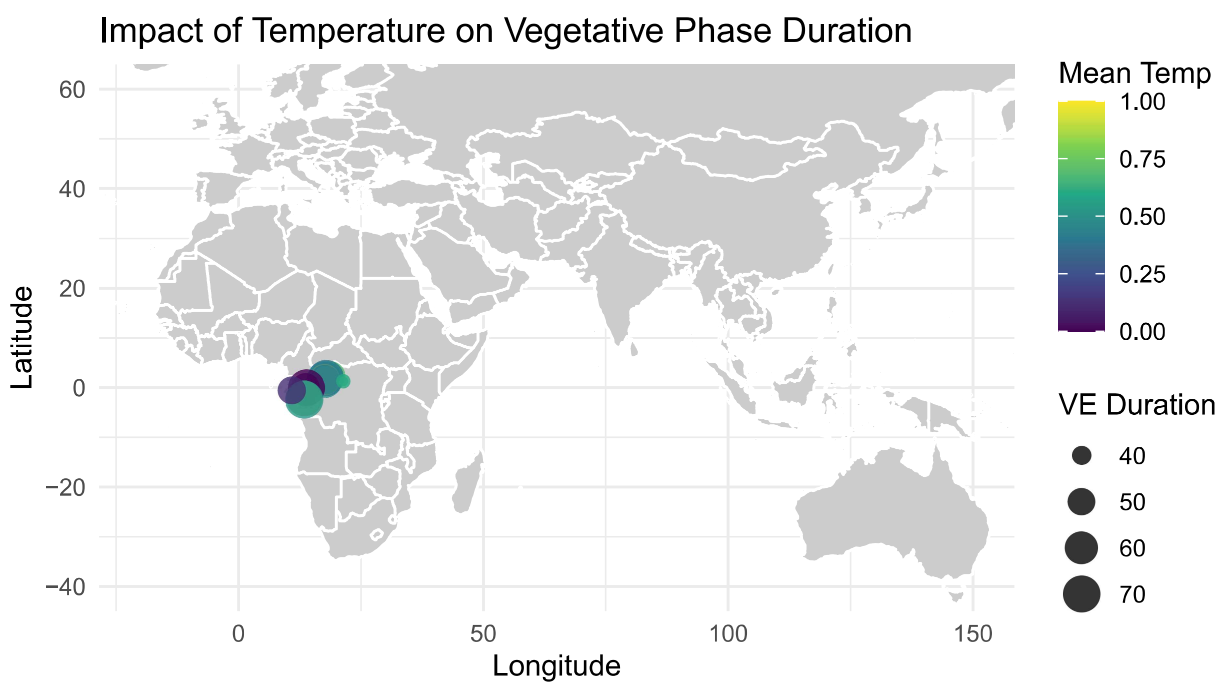
```
## Warning: package 'maps' was built under R version 4.3.3
```

```
##
## Attaching package: 'maps'
```

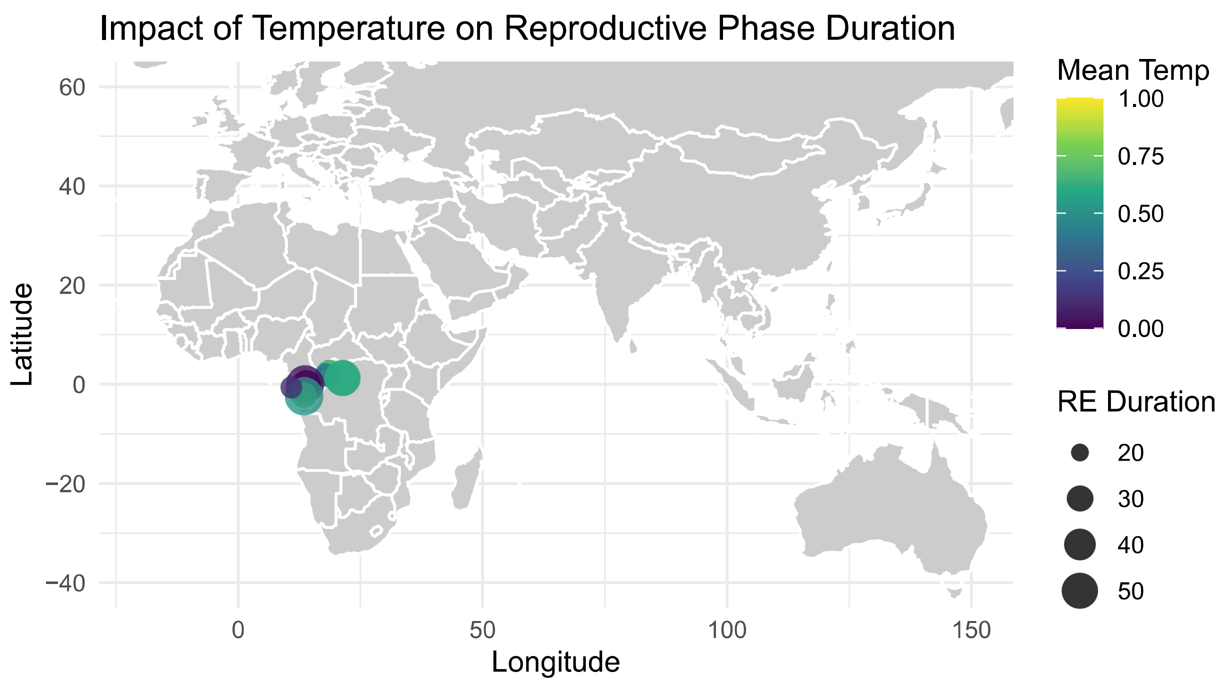
```
## The following object is masked from 'package:purrr':
##
##     map
```

```
world_map <- map_data("world")

ggplot() +
  geom_polygon(data = world_map, aes(x = long, y = lat, group = group),
               fill = "gray80", color = "white") +
  geom_point(data = data_df, aes(x = E_longitude, y = N_latitude,
```

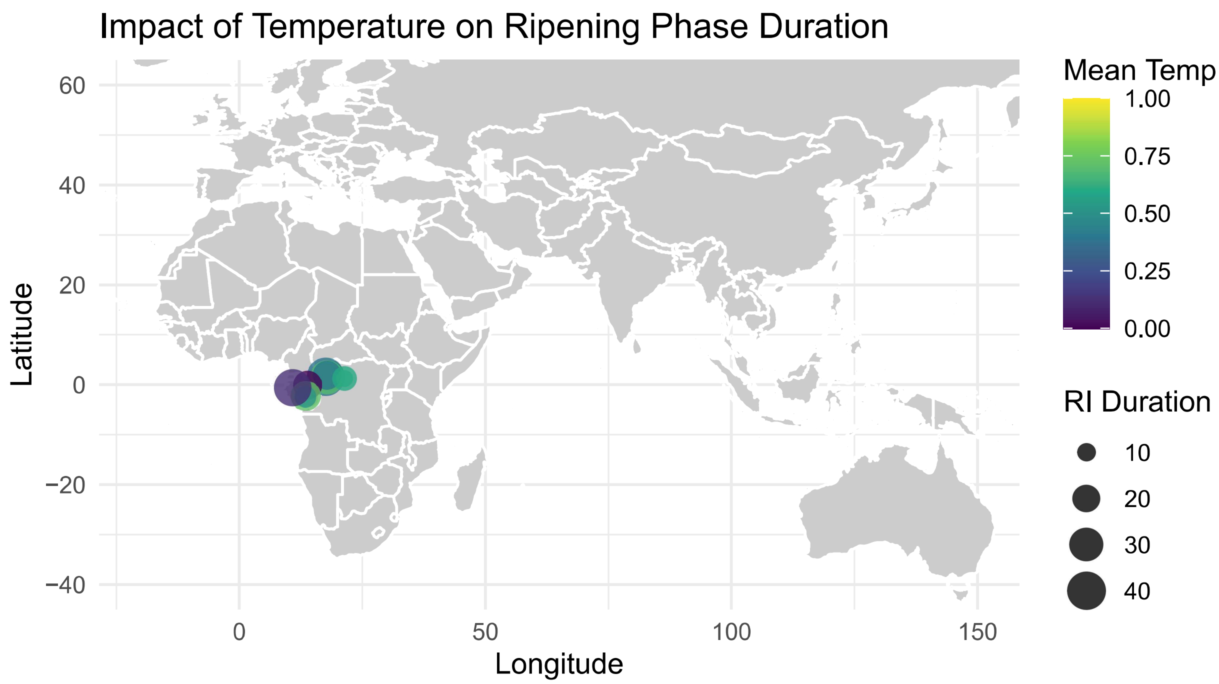
```
                                color = TM, size = VE_Duration), alpha = 0.8) +
  scale_color_viridis_c(name = "Mean Temp") +
  scale_size_continuous(name = "VE Duration") +
  labs(title = "Impact of Temperature on Vegetative Phase Duration",
       x = "Longitude", y = "Latitude") +
  coord_quickmap(xlim = c(-20, 150), ylim = c(-40, 60)) +  # Restrict to Africa and Asia
  theme_minimal()
```




```
ggplot() +
  geom_polygon(data = world_map, aes(x = long, y = lat, group = group),
             fill = "gray80", color = "white") +
  geom_point(data = data_df, aes(x = E_longitude, y = N_latitude,
                                color = TM, size = RE_Duration), alpha = 0.8) +
  scale_color_viridis_c(name = "Mean Temp") +
  scale_size_continuous(name = "RE Duration") +
  labs(title = "Impact of Temperature on Reproductive Phase Duration",
       x = "Longitude", y = "Latitude") +
  coord_quickmap(xlim = c(-20, 150), ylim = c(-40, 60)) +
  theme_minimal()
```




```r
ggplot() +
  geom_polygon(data = world_map, aes(x = long, y = lat, group = group),
               fill = "gray80", color = "white") +
  geom_point(data = data_df, aes(x = E_longitude, y = N_latitude,
                                 color = TM, size = RI_Duration), alpha = 0.8) +
  scale_color_viridis_c(name = "Mean Temp") +
  scale_size_continuous(name = "RI Duration") +
  labs(title = "Impact of Temperature on Ripening Phase Duration",
       x = "Longitude", y = "Latitude") +
  coord_quickmap(xlim = c(-20, 150), ylim = c(-40, 60)) +
  theme_minimal()
```

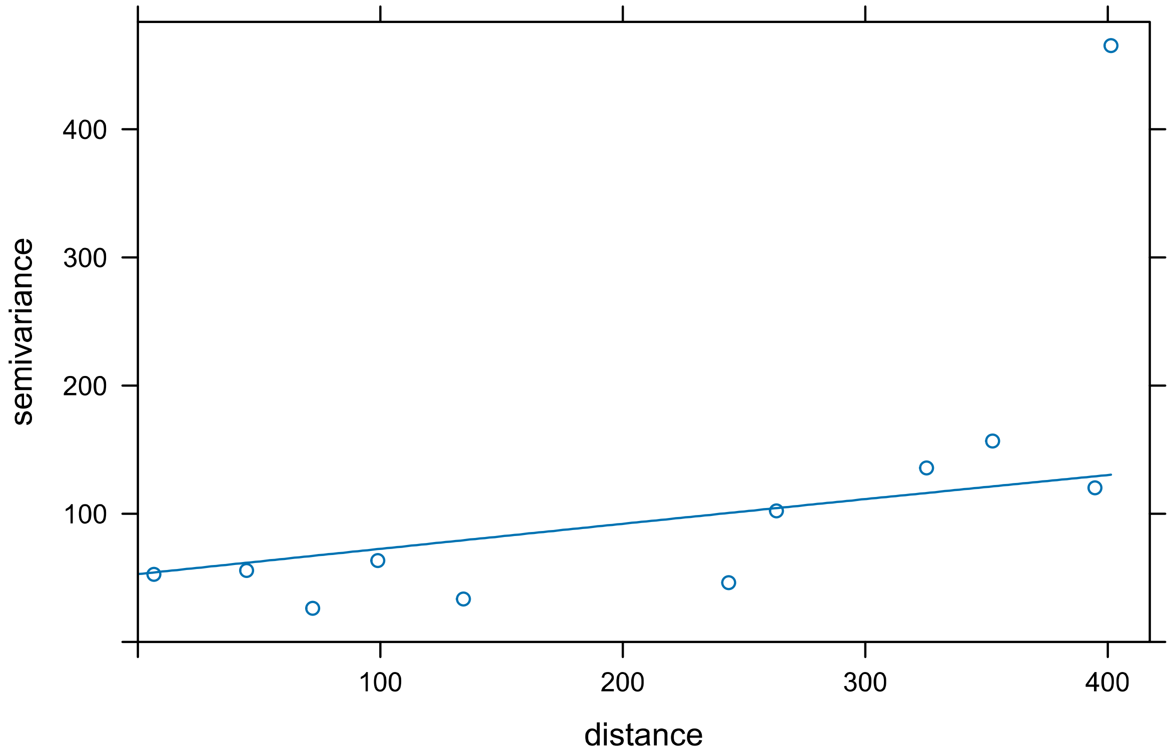


Use kriging to interpolate Vegetation Duration values at unknown locations

```r
library(sp)
library(gstat)

data_df <- as.data.frame(data)
proj4string(data) <- CRS("+proj=longlat +datum=WGS84")
data_utm <- spTransform(data, CRS(utm_crs))
# Create an empirical variogram
variogram_model <- variogram(VE_Duration ~ 1, data)

# Fit a theoretical variogram model
variogram_fit <- fit.variogram(variogram_model, model = vgm("Sph"))
```

```
## Warning in fit.variogram(variogram_model, model = vgm("Sph")): No convergence
## after 200 iterations: try different initial values?
```

```r
plot(variogram_model, variogram_fit)
```




```
grd <- expand.grid(
 longitude = seq(min(data$E_longitude), max(data$E_longitude), length.out = 100),
  latitude = seq(min(data$N_latitude), max(data$N_latitude), length.out = 100)
)
coordinates(grd) <- ~longitude + latitude
gridded(grd) <- TRUE
proj4string(grd) <- CRS("+proj=longlat +datum=WGS84")

# Ordinary Kriging
kriging_result <- krige(VE_Duration ~ 1, data, grd, variogram_fit)
```

```
## [using ordinary kriging]
```

```
## Warning in predict.gstat(g, newdata = newdata, block = block, nsim = nsim, :
## Covariance matrix singular at location [10.8474,-2.39386,0]: skipping...
```

```
## Warning in predict.gstat(g, newdata = newdata, block = block, nsim = nsim, :
## Covariance matrix singular at location [10.9539,-2.39386,0]: skipping...
```

```
## Warning in predict.gstat(g, newdata = newdata, block = block, nsim = nsim, :
## Covariance matrix singular at location [11.0604,-2.39386,0]: skipping...
```

```
## Warning in predict.gstat(g, newdata = newdata, block = block, nsim = nsim, :
## Covariance matrix singular at location [11.1669,-2.39386,0]: skipping...
```