Appendix

27 January 2023

Code to reproduce the map of the main text.

# 1 Theoretical example: homogeneous controls

Analyses rely on the *dbmss* (Marcon et al. 2015) package for R (R Core Team 2022).

## 1.1 Dataset simulation

We build a point pattern made of cases (the points of interest) and controls (the background distribution of points).

Cases are a Matérn (Matérn 1960) point pattern with (expected) clusters of (expected) points in a circle of radius *scale*. Controls are a Poisson point pattern (i.e. complete spatial randomness) of (expected) points.

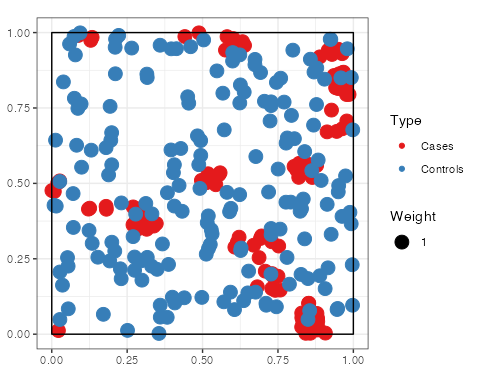
library(dplyr)  
library(dbmss)  
# Simulation of cases (clusters)  
rMatClust(kappa = 10, scale = 0.05, mu = 10) %>%  
 as.wmppp -> CASES  
CASES$marks$PointType <- "Cases"  
# Number of points  
CASES$n

## [1] 113

# Simulation of controls (random distribution)  
rpoispp(lambda = 200) %>%  
 as.wmppp -> CONTROLS  
CONTROLS$marks$PointType <- "Controls"  
# Number of points  
CONTROLS$n

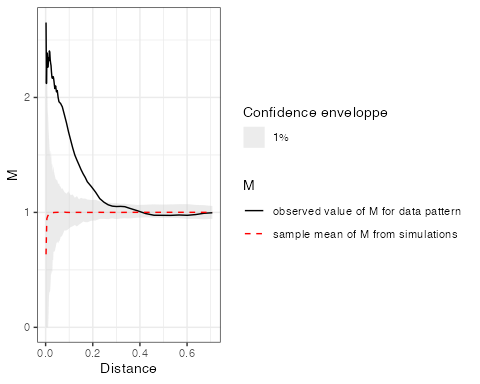
## [1] 185

# Mixed patterns (cases and controls)  
ALL <- superimpose(CASES, CONTROLS)  
autoplot(ALL)



## 1.2 Calculate and plot M Cases

# Fix the number of simulations and the level of  
# risk  
NumberOfSimulations <- 1000  
Alpha <- 0.01  
# Calculate and plot M Cases  
ALL %>%  
 MEnvelope(ReferenceType = "Cases", SimulationType = "RandomLocation",  
 NumberOfSimulations = NumberOfSimulations,  
 Alpha = Alpha, Global = TRUE) -> M\_env\_cases  
autoplot(M\_env\_cases)



The plot shows a clear relative concentration of cases.

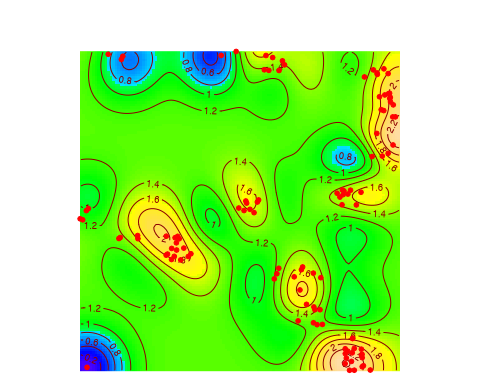
## 1.3 Map M results

To plot the individual values of *M* around each case, a distance must be chosen. Then, the function must be computed at this distance with individual values. Finally, a kriged weighted, marked, planar point patterns (kwmppp) object is produced and plotted.

# Choose the distance to plot  
Distance <- 0.1  
# Calculate the M values to plot  
ALL %>%  
 Mhat(r = c(0, Distance), ReferenceType = "Cases",  
 NeighborType = "Cases", Individual = TRUE) ->  
 M\_TheoEx  
# Map resolution  
resolution <- 512  
# Create a kriged weighted marked planar point  
# pattern (kwmppp)  
M\_TheoEx\_map <- kwmppp(ALL, fvind = M\_TheoEx, ReferenceType = "Cases",  
 distance = Distance)

## [using ordinary kriging]

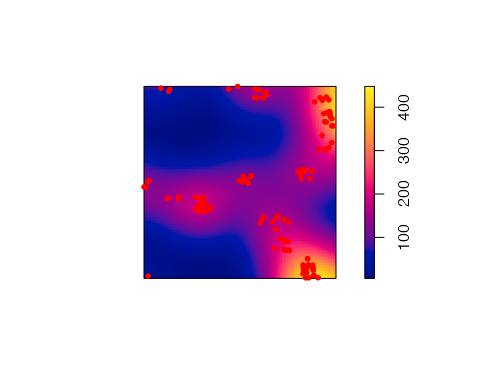
# Plot the point pattern with values of  
# M(Distance)  
plot(M\_TheoEx\_map)  
# Add the cases to the map  
points(ALL[ALL$marks$PointType == "Cases"], pch = 20,  
 col = "red")



## 1.4 Compare with the density of cases

The density of cases is plotted. High densities are similar to high relative concentrations in this example because the control points are homogeneously distributed.

plot(density(CASES), main = "")  
points(ALL[ALL$marks$PointType == "Cases"], pch = 20,  
 col = "red")



# 2 Theoretical example: inhimogeneous controls

Analyses rely on the *dbmss* (Marcon et al. 2015) package for R (R Core Team 2022).

## 2.1 Dataset simulation

We build a point pattern made of cases (the points of interest) and controls (the background distribution of points).

Cases are a Matérn (Matérn 1960) point pattern with (expected) clusters of (expected) points in a circle of radius *scale*. Controls are a Poisson point pattern whose density decreases exponentially along the y-axis (we will call “north” the higher y values).

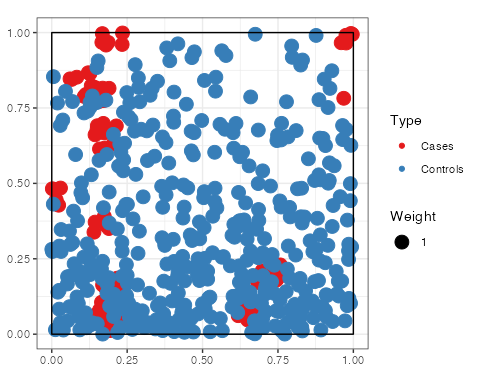
library(dbmss)  
# Simulation of cases (clusters)  
rMatClust(kappa = 10, scale = 0.05, mu = 10) %>%  
 as.wmppp -> CASES  
CASES$marks$PointType <- "Cases"  
# Number of points  
CASES$n

## [1] 94

# Simulation of controls (random distribution)  
rpoispp(function(x, y) {  
 1000 \* exp(-2 \* y)  
}) %>%  
 as.wmppp -> CONTROLS  
CONTROLS$marks$PointType <- "Controls"  
# Number of points  
CONTROLS$n

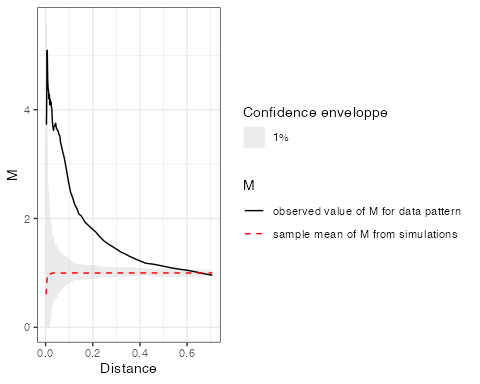
## [1] 428

# Mixed patterns (cases and controls)  
ALL <- superimpose(CASES, CONTROLS)  
autoplot(ALL)



## 2.2 Calculate and plot M Cases

# Fix the number of simulations and the level of  
# risk  
NumberOfSimulations <- 1000  
Alpha <- 0.01  
# Calculate and plot M Cases  
ALL %>%  
 MEnvelope(ReferenceType = "Cases", SimulationType = "RandomLocation",  
 NumberOfSimulations = NumberOfSimulations,  
 Alpha = Alpha, Global = TRUE) -> M\_env\_cases  
autoplot(M\_env\_cases)



The plot shows a clear relative concentration of cases.

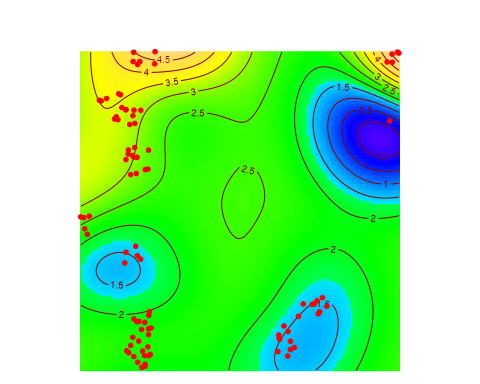
## 2.3 Map M results

To plot the individual values of *M* around each case, a distance must be chosen. Then, the function must be computed at this distance with individual values. Finally, a kriged weighted, marked, planar point patterns (kwmppp) object is produced and plotted.

# Choose the distance to plot  
Distance <- 0.1  
# Calculate the M values to plot  
ALL %>%  
 Mhat(r = c(0, Distance), ReferenceType = "Cases",  
 NeighborType = "Cases", Individual = TRUE) ->  
 M\_TheoEx  
# Map resolution  
resolution <- 512  
# Create a kriged weighted marked planar point  
# pattern (kwmppp)  
M\_TheoEx\_map <- kwmppp(ALL, fvind = M\_TheoEx, ReferenceType = "Cases",  
 distance = Distance)

## [using ordinary kriging]

# Plot the point pattern with values of  
# M(Distance)  
plot(M\_TheoEx\_map)  
# Add the cases to the map  
points(ALL[ALL$marks$PointType == "Cases"], pch = 20,  
 col = "red")



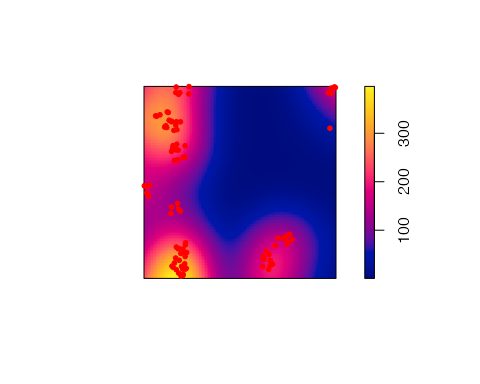
We can see that cases are concentrated everywhere (local M value above 1) because we chose a Matérn point pattern.

The areas with the higher relative concentration are located in the north of the map because the controls are less dense there. The southern-most cluster illustrates that the relative concentration of cases, although higher than 1, is clearly lower than that of the northern clusters, which have the same characteristics but are in a less-dense control neighborhood.

## 2.4 Compare with the density of cases

The density of cases is plotted. High densities are not similar to high relative concentrations in this example because the control points are not homogeneously distributed.

plot(density(CASES), main = "")  
points(ALL[ALL$marks$PointType == "Cases"], pch = 20,  
 col = "red")



# 3 Suzanne Lenglen Park

## 3.1 Data

Our data is extracted from “Paris open data” [[1]](#footnote-54).

### 3.1.1 Data wrangling

Data are stored in trees\_2021.zip which contains two GeoJSON files:

* trees\_2021 stores all trees of the city of Paris in 2021.
* trees\_logged contains all trees logged

They must be read. Data are projected into the Lambert 93 datum so that coordinates are in meters.

unzip("data/trees\_2021.zip", exdir = "data")  
library("sf")  
read\_sf("data/trees\_2021.geojson") %>%  
 st\_transform(crs = 2154) -> trees\_all\_raw  
read\_sf("data/trees\_logged.geojson") %>%  
 st\_transform(crs = 2154) -> trees\_logged\_raw

#### 3.1.1.1 All trees

The first dataset contains all trees in Paris in 2021, including those to be cut.

Trees from the Suzanne Lenglen park are selected. Columns of interest are:

* ID: a numeric unique identifier for each tree.
* Species\_name: the scientific name of the tree species, i.e. Genus species.
* Status: Alive.
* Genus.
* Species.
* French\_species\_name: vernacular name.
* Circumference: in cm.

library("dplyr")  
trees\_all\_raw %>%  
 # Filter Suzanne Lenglen park  
filter(adresse == "PARC OMNISPORT SUZANNE LENGLEN / 7 BOULEVARD DES FRERES VOISIN") %>%  
 # Create a field with the species name  
mutate(Species\_name = as.factor(paste(genre, espece))) %>%  
 # Create a field with the status  
mutate(Status = "Alive") %>%  
 # Genus and Species fields  
mutate(Genus = as.factor(genre)) %>%  
 mutate(Species = as.factor(espece)) %>%  
 # Rename and finally select columns  
rename(ID = idbase, French\_species\_name = libellefrancais,  
 Circumference = circonferenceencm) %>%  
 select(ID, Species\_name, Status, Genus, Species,  
 French\_species\_name, Circumference) -> trees\_all  
# Number of trees  
trees\_all %>%  
 nrow()

## [1] 1472

We have 1472 trees in the park.

#### 3.1.1.2 Logged trees

Logged trees are in the second dataset.

Their status is “Logged”. An extra field, Logging\_reason contains the motivation to cut them off (in French). Circumfernce is absent.

# Tree description  
trees\_logged\_raw %>%  
 # Filter Suzanne Lenglen park  
filter(adresse == "PARC OMNISPORT SUZANNE LENGLEN / 7 BOULEVARD DES FRERES VOISIN") %>%  
 # Exclude unidentified trees  
filter(!is.na(especearbreprecedent), !is.na(libellefrancaisarbreprecedent),  
 !is.na(genrearbreprecedent)) %>%  
 filter(libellefrancaisarbreprecedent != "Non spécifié") %>%  
 filter(especearbreprecedent != "n. sp.") %>%  
 # Create a field with the species name  
mutate(Species\_name = as.factor(paste(genrearbreprecedent,  
 especearbreprecedent))) %>%  
 # Create a numeric ID  
mutate(ID = as.integer(idbase)) %>%  
 # Create a field with the status  
mutate(Status = "Logged") %>%  
 # Genus and Species fields  
mutate(Genus = as.factor(genrearbreprecedent)) %>%  
 mutate(Species = as.factor(especearbreprecedent)) %>%  
 # Reason for logging (in French)  
mutate(Logging\_reason = motifabattagearbreprecedent) %>%  
 # Rename and finally select columns  
rename(French\_species\_name = libellefrancaisarbreprecedent) %>%  
 select(ID, Species\_name, Status, Genus, Species,  
 Logging\_reason, French\_species\_name) ->  
 trees\_logged  
# Number of trees  
trees\_logged %>%  
 nrow()

## [1] 48

48 among the 1472 trees of the park were logged.

#### 3.1.1.3 Merge

The two datasets are merged here.

The logged trees must be removed from the first one. Circumference is removed because it is missing from the logged trees dataset.

# All trees  
trees\_all %>%  
 # Delete the logged trees  
filter(!(ID %in% trees\_logged$ID)) %>%  
 # Delete the circumference that is absent in  
 # trees\_logged  
mutate(Circumference = NULL) %>%  
 # Bind the logged trees  
bind\_rows(trees\_logged) -> trees\_no\_circumference

Circumferences of all trees, including logged ones, are in tree\_all from where they can be recovered.

# Prepare a tibble with circumferences  
trees\_all %>%  
 select(ID, Circumference) %>%  
 # inner\_join.sf refuses sf objects  
st\_set\_geometry(NULL) -> Circumferences  
# Add the Circumference of trees  
trees\_no\_circumference %>%  
 inner\_join(Circumferences, by = "ID") -> trees

#### 3.1.1.4 Simpler logging reasons

Logging reasons can be:

* Decaying: the tree’s condition is not healthy enough to keep it safely in a public park.
* Infected: the tree is a maple affected by the (contagious) sooty bark disease, caused by the fungus *Cryptostroma corticale*.

library("stringr")  
trees$Logging\_reason[is.na(trees$Logging\_reason)] <- ""  
trees$Logging\_reason %>%  
 str\_replace("Arbre.\*", "Decaying") %>%  
 str\_replace("Foyer.\*", "Infected") -> trees$Logging\_reason

#### 3.1.1.5 Factors

Several fields are converted to factors for efficiency.

trees$Logging\_reason <- as.factor(trees$Logging\_reason)  
trees$Status <- as.factor(trees$Status)  
trees$French\_species\_name <- as.factor(trees$French\_species\_name)

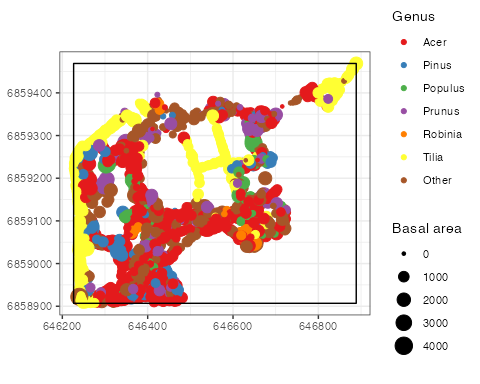
### 3.1.2 Point patterns

dbmms uses weighted, marked, planar point patterns (wmppp). A wmppp named trees\_infected is built. Point marks are their basal area (as weight) and either their logging reason or their genus if they are alive.

library("dbmss")  
trees %>%  
 # Weight is the basal area  
mutate(PointWeight = Circumference^2/4/pi) %>%  
 mutate(PointType = ifelse(Logging\_reason == "",  
 as.character(Genus), as.character(Logging\_reason))) %>%  
 # Add X and Y  
bind\_cols(st\_coordinates(trees)) %>%  
 wmppp(window = as.owin(st\_bbox(trees)), unitname = c("meter",  
 "meters")) -> trees\_infected

We also need a point pattern to describe the park before logging, as a reference.

trees\_all %>%  
 # Weight is the basal area  
mutate(PointWeight = Circumference^2/4/pi) %>%  
 # Genus is the point type  
rename(PointType = Genus) %>%  
 # Add X and Y  
bind\_cols(st\_coordinates(trees\_all)) %>%  
 wmppp(window = as.owin(st\_bbox(trees\_all)), unitname = c("meter",  
 "meters")) -> trees\_2021  
autoplot(trees\_2021) + ggplot2::labs(size = "Basal area",  
 color = "Genus")



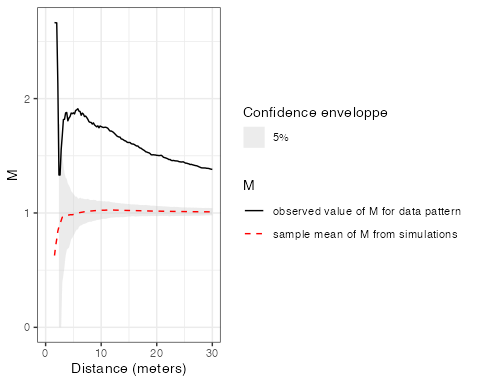
The map shows the tree genera. Maple (*Acer sp.*) are the most abundant trees in the park.

## 3.2 Spatial analyses

### 3.2.1 Spatial concentration of maple trees

The M statistic is computed to detect the spatial concentration of maple trees before logging.

Distance <- 15  
NumberOfSimulations <- 1000  
trees\_2021 %>%  
 MEnvelope(r = 0:(10 \* Distance)/5, ReferenceType = "Acer",  
 NeighborType = "Acer", NumberOfSimulations = NumberOfSimulations) ->  
 M\_Acer  
autoplot(M\_Acer)



To map it, individual values must be calculated at the chosen distance, that is 15 meters.

trees\_2021 %>%  
 Mhat(r = c(0, Distance), ReferenceType = "Acer",  
 NeighborType = "Acer", Individual = TRUE) ->  
 M\_ind\_Acer

The map requires krigeing the individual values on a grid of points. To build the grid, the size ratio of the spatial window is calculated. The number of rows and columns of the grid will respect this ratio so that its points are equally spaced.

# Window ratio  
ratio <- with(trees\_infected$window, {  
 (yrange[2] - yrange[1])/(xrange[2] - xrange[1])  
})  
# Map resolution: number of columns of the grid.  
resolution <- 512

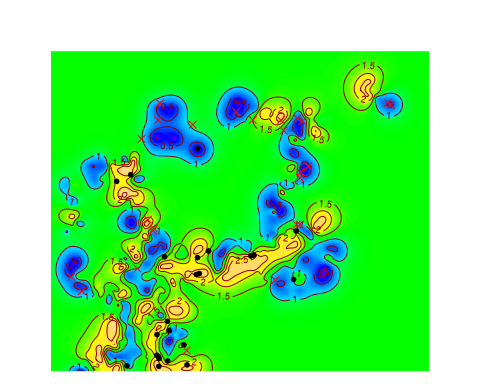
A kriged weighted, marked, planar point patterns (kwmppp) object is produced and plotted. Logged trees are added to the map:

* Infected trees are black points,
* Decaying trees are red crosses.

trees\_2021 %>%  
 kwmppp(fvind = M\_ind\_Acer, distance = Distance,  
 ReferenceType = "Acer", Nbx = resolution, Nby = resolution \*  
 ratio) -> map\_acer

## [using ordinary kriging]

plot(map\_acer)  
# Add infected trees  
points(trees\_infected[trees\_infected$marks$PointType ==  
 "Infected"], pch = 20)  
# And decaying trees  
points(trees\_infected[trees\_infected$marks$PointType ==  
 "Decaying"], pch = 4, col = "red")



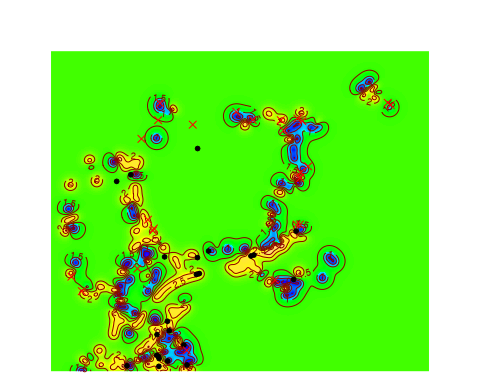
Infected trees are present in the areas where maples are concentrated.

The 15-meter distance has been chosen because it is a plausible distance of contagion between trees. A shorter distance, say 6 meters, would be justified as a peak of M(distance) but neighbors are scarse at this scale, leading to less statistical power. At greater distances, such as 30m, concentration is still significant but neighborhoods overlap more, leading to less clear results.

Both alternative distances are tested here.

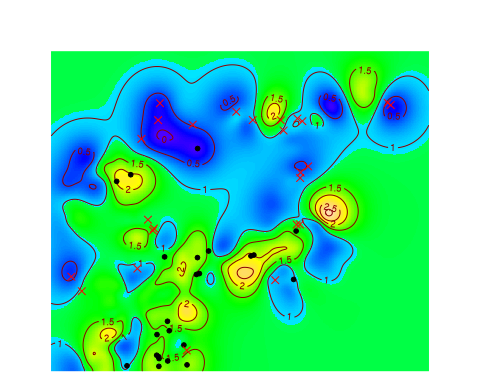
Considering 6-meter-radius neighborhoods, the map shows that some high and low values of local concentration are not detected:

## [using ordinary kriging]



30-meter-radius neighborhoods are shown below:

## [using ordinary kriging]

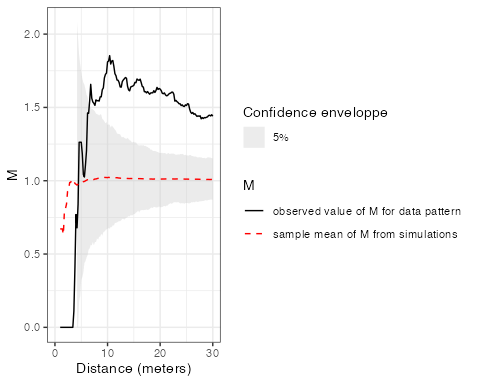


Increasing the distance that defines neighborhoods extends the areas where high or low concentration is detected around the most isolated trees, making them more visible, but blurs them in dense areas, where high and low values overlap and cancel out (e.g. in the south-eastern part of the map). So the appropriate distance should be chosen according to the knowledge of the studied process.

### 3.2.2 Concentration of maples around infected trees

To test the intertype concentration between sane and infected maple trees, the intertype M statistic is computed.

trees\_infected %>%  
 MEnvelope(r = 0:(10 \* Distance)/5, ReferenceType = "Infected",  
 NeighborType = "Acer", NumberOfSimulations = NumberOfSimulations) ->  
 M\_Infected\_Acer  
autoplot(M\_Infected\_Acer)

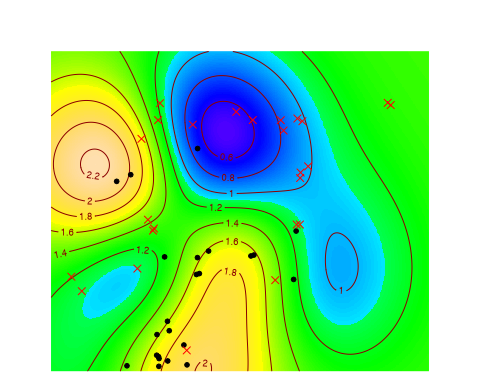


The map is produced.

trees\_infected %>%  
 Mhat(ReferenceType = "Infected", NeighborType = "Acer",  
 Individual = TRUE) -> M\_ind\_Infected\_Acer  
trees\_infected %>%  
 kwmppp(fvind = M\_ind\_Infected\_Acer, distance = Distance,  
 ReferenceType = "Infected", Nbx = resolution,  
 Nby = resolution \* ratio) -> map\_infected\_acer

## [using ordinary kriging]

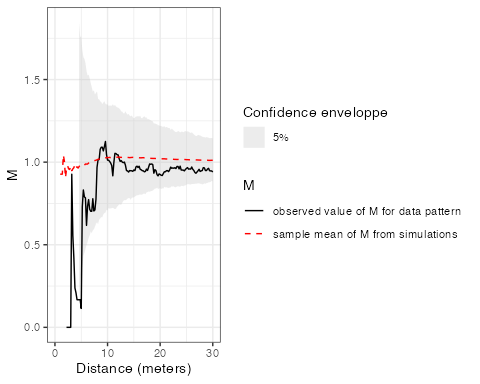
plot(map\_infected\_acer)  
# Add infected trees  
points(trees\_infected[trees\_infected$marks$PointType ==  
 "Infected"], pch = 20)  
# And decaying trees  
points(trees\_infected[trees\_infected$marks$PointType ==  
 "Decaying"], pch = 4, col = "red")



Infected and sane maple trees are significantly concentrated.

In contrast, decaying trees and mapples rather repulse each others.

trees\_infected %>%  
 MEnvelope(r = 0:(10 \* Distance)/5, ReferenceType = "Decaying",  
 NeighborType = "Acer", NumberOfSimulations = NumberOfSimulations) ->  
 M\_Decaying\_Acer  
autoplot(M\_Decaying\_Acer)



# References

Marcon, Eric, Stéphane Traissac, Florence Puech, and Gabriel Lang. 2015. “Tools to Characterize Point Patterns: Dbmss for R.” *Journal of Statistical Software* 67 (3): 1–15. <https://doi.org/10.18637/jss.v067.c03>.

Matérn, Bertil. 1960. “Spatial Variation.” *Meddelanden Från Statens Skogsforskningsinstitut* 49 (5): 1–144.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

1. <https://opendata.paris.fr>. [↑](#footnote-ref-54)