

Exercise 3

Malaria mapping in Tanzania

In this exercise, you will use the model from Exercise 2, where Enhanced Vegetation Index (EVI) and Temperature are predictors for malaria prevalence. The linear predictor of the Binomial geostatistical model is given by:

$$\log \left\{ \frac{p(x_i)}{1 - p(x_i)} \right\} = \beta_0 + \beta_1 e(x_i) + \beta_2 t(x_i) + \beta_3 \max\{t(x_i) - 33, 0\} + S(x_i) + Z_i,$$

where:

- $e(x_i)$ is EVI at location x_i ,
- $t(x_i)$ is the temperature at location x_i ,
- $S(x_i)$ is a stationary and isotropic Gaussian process with mean zero and an exponential correlation function,
- Z_i are independent and identically distributed zero-mean Gaussian variables.

1. Generate Predictive Maps

- Using a regular grid with a spatial resolution of 10×10 km, generate:
 - A predictive map for malaria prevalence across Tanzania.
 - A probability map showing areas where prevalence is below 0.1.
 - A probability map showing areas where prevalence is between 0.1 and 0.2.
 - A classification map where each grid pixel is assigned to one of three prevalence classes: (0, 0.1), (0.1, 0.2), or (0.2, 1) based on the highest predictive probability from the fitted model.

Question: Where are the main malaria hotspots?

2. Regional-Level Estimates

- Using administrative level 1 boundaries:
 - Repeat the previous analysis, but consider the regional average prevalence (without weighting) as the predictive target.
 - Compare the regional-level maps with the high-resolution maps from step 1.

Question: What differences do you observe between the two approaches?

3. Estimating Malaria Cases

- Estimate the total number of malaria cases in Tanzania.
- Compute the 95% confidence interval for this estimate.
- Discuss whether the estimate is likely to be an overestimate or underestimate and explain why.

Counting *Anopheles gambiae* Mosquitoes in Southern Cameroon

Consider the outcome `An.gambiae` from the `anopheles` dataset in `RiskMap`. To define the study area, you can create a shapefile using the *convex hull* of the data locations. The convex hull is the smallest convex boundary that encloses all the points in a dataset. You can read more about it [here](#).

Use the following lines of code to compute the convex hull:

```
anopheles_sf <- st_as_sf(anopheles, coords = c("web_x", "web_y"), crs = 3857)
shp_ch <- convex_hull_sf(anopheles_sf)
```

Next, using a Poisson geostatistical model with elevation as a predictor:

- Generate a map of the predicted number of mosquitoes.
 - Estimate the total number of mosquitoes in the area between 400 meters and 830 meters in elevation.
-