



Design and analysis of prevalence surveys for NTDs: how to increase model uptake?

Claudio Fronterre

Lecturer in Biostatistics (Global Health)

CHICAS, Lancaster Medical School, Lancaster University

22 June 2023, N8CIR Digital Health Community Day

Neglected tropical diseases

"A diverse group of communicable diseases that prevail in tropical and subtropical conditions in 149 countries – affect more than one billion people and cost developing economies billions of dollars every year."

Major attributes of the NTDS

- most prevalent among poor people
- associated with high disease burden but low mortality
- endemic in rural areas of low-income countries
- chronic
- disabling (growth delays, blindness, or disfigurement)
- stigmatizing

Burden of NTDs in the world

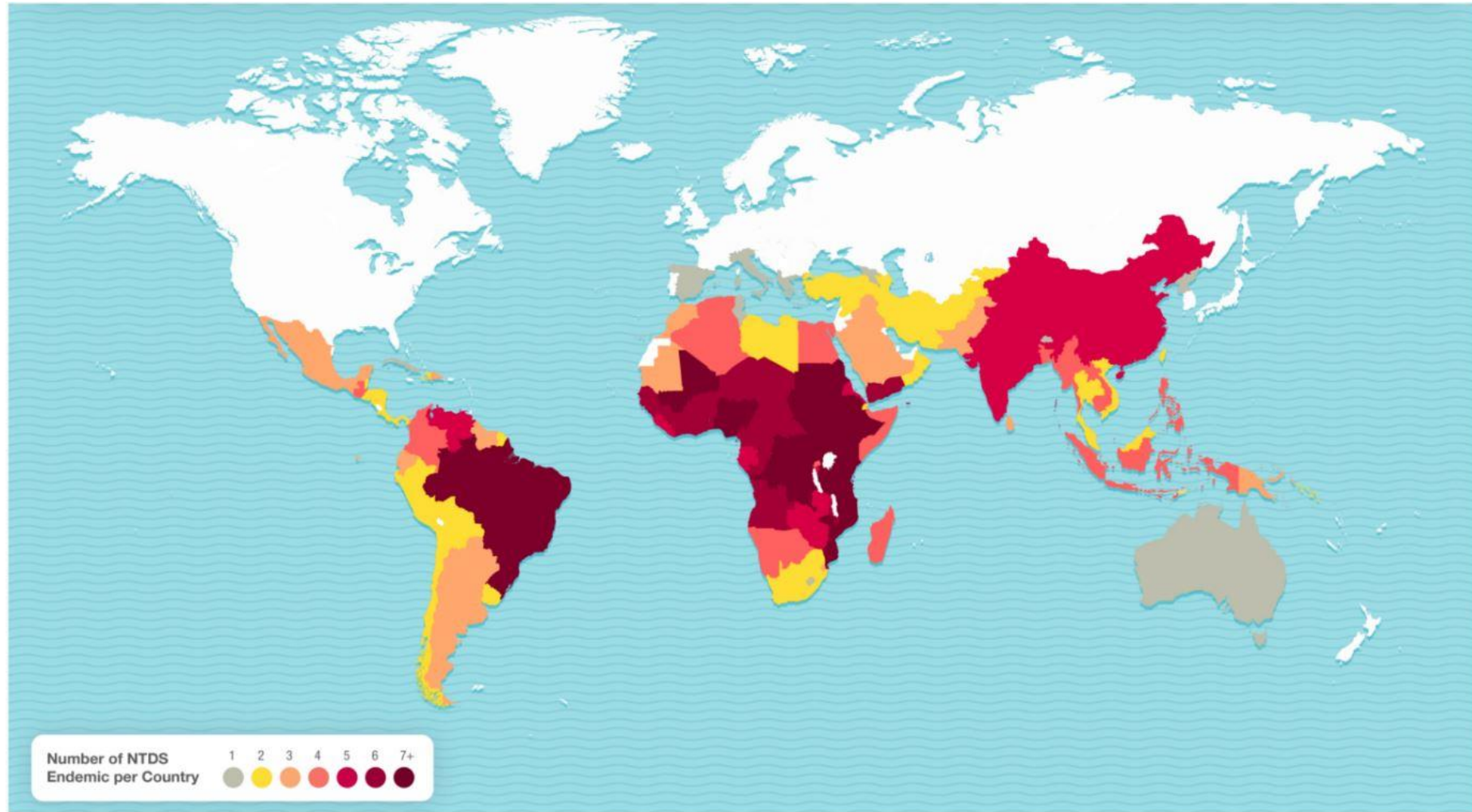
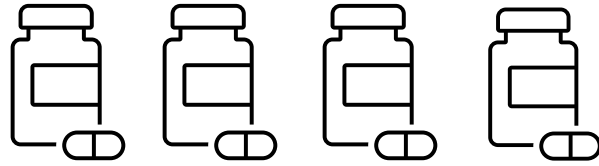


Fig. 1: NTDs endemicity based on 2009-2010 data. Source: <https://unitingtocombatntds.org/>.

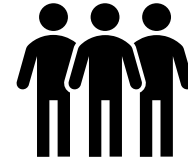
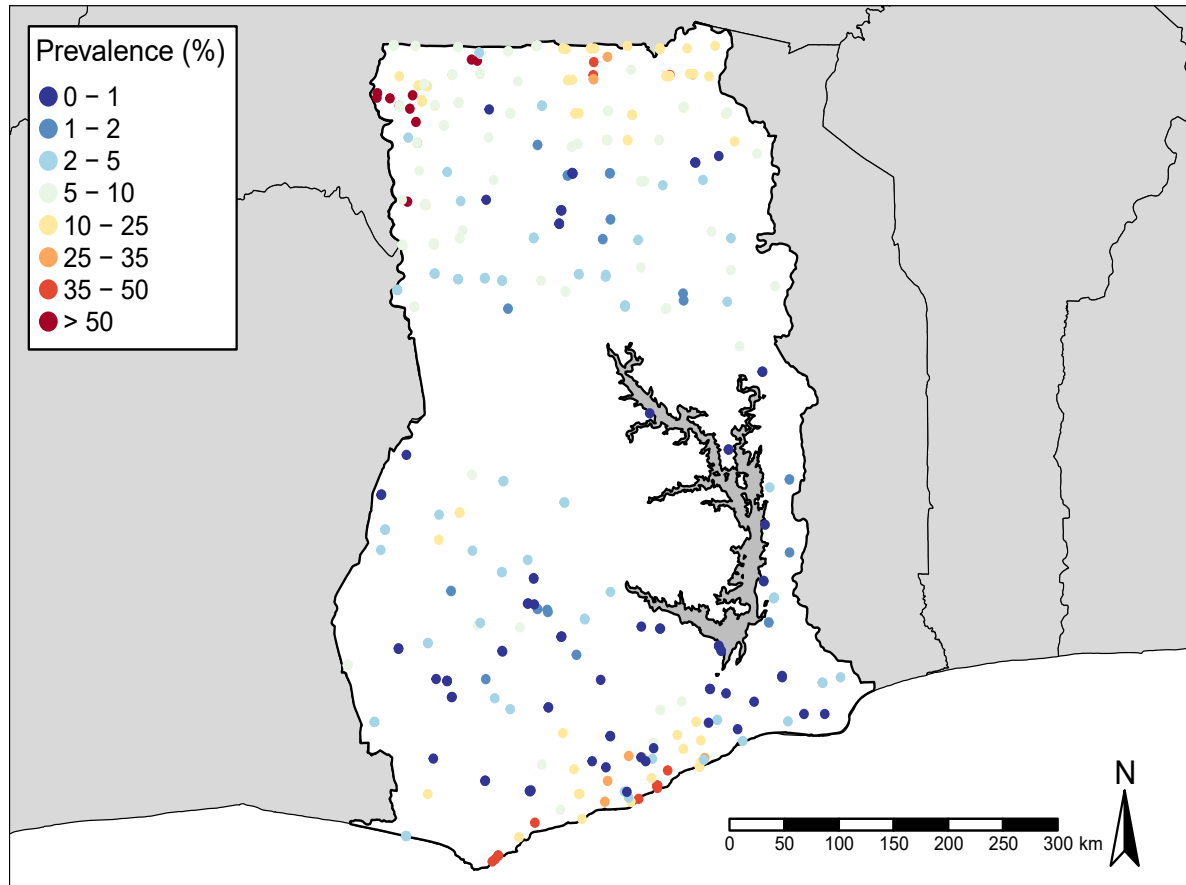
NTD control programme stages

MDA distribution



Prevalence surveys

Lymphatic filariaris (LF) mapping surveys in Ghana



Limited resources allow to sample only a small portion of the population.



Imperfect diagnostic tests.

Mapping surveys

Collect prevalence data: (x_i, m_i, Y_i) $i=1, \dots, n$

- asses the endemicity status
- decide the appropriate treatment

Understanding the **spatial** distribution and burden of the disease is essential to inform disease control policies.

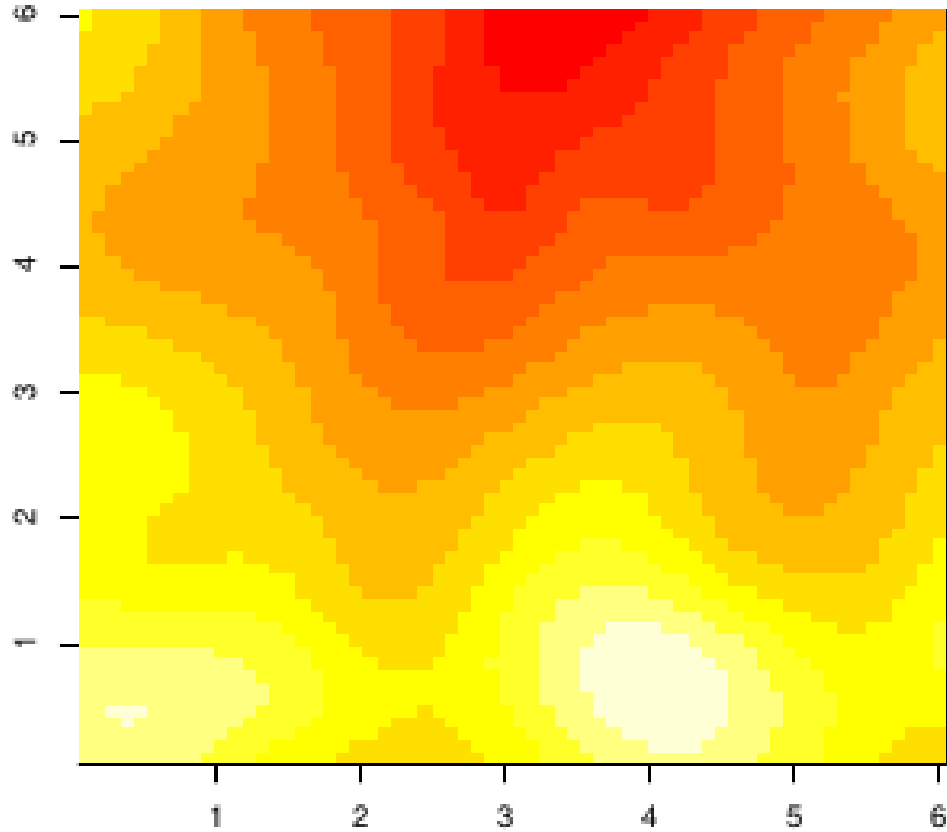
The role of model-based approaches

Provide a set of tools that can:

- improve the **design** and **analysis** of prevalence surveys
- support decision making under **uncertainty**.

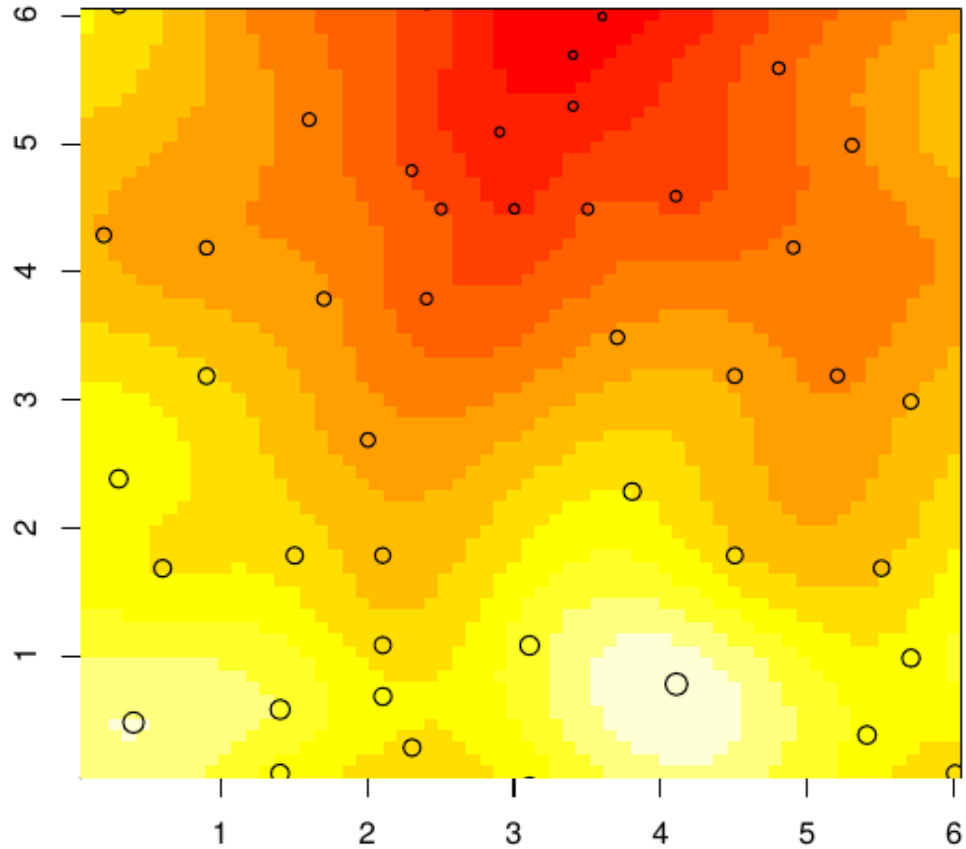
Why use model-based geostatistics?

Area-wide prevalence = 41.7%



Why use model-based geostatistics?

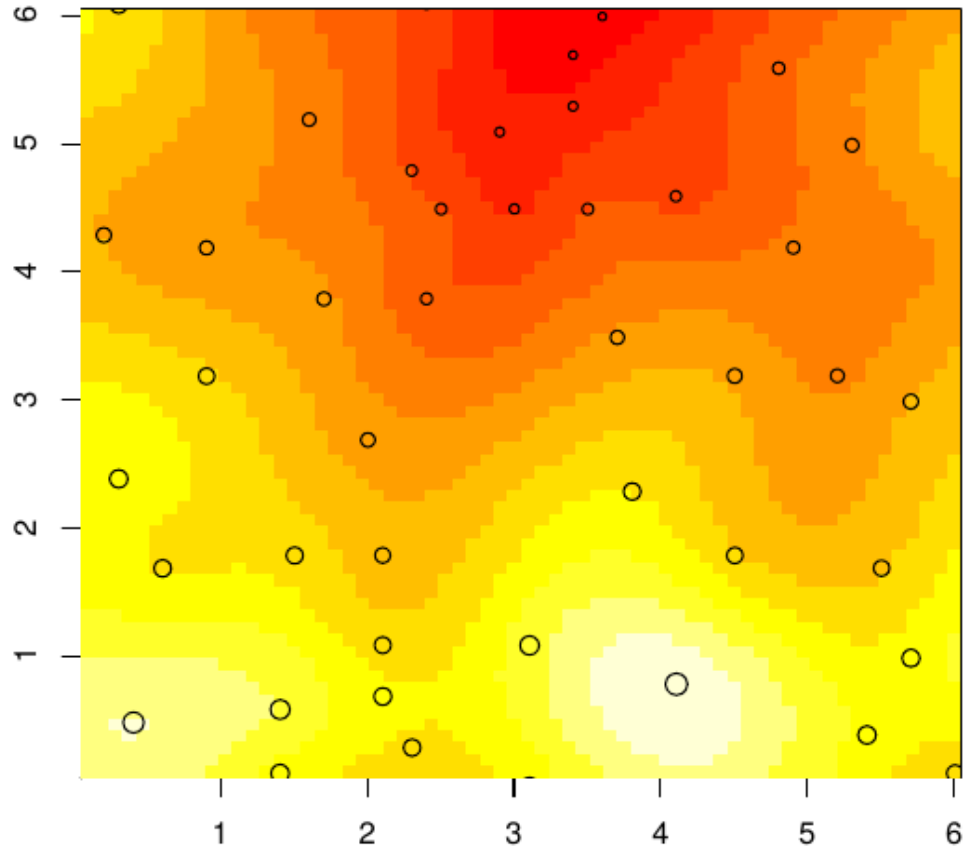
Area-wide prevalence = 41.7%



Measure prevalence at sampling locations

Why use model-based geostatistics?

Area-wide prevalence = 41.7%



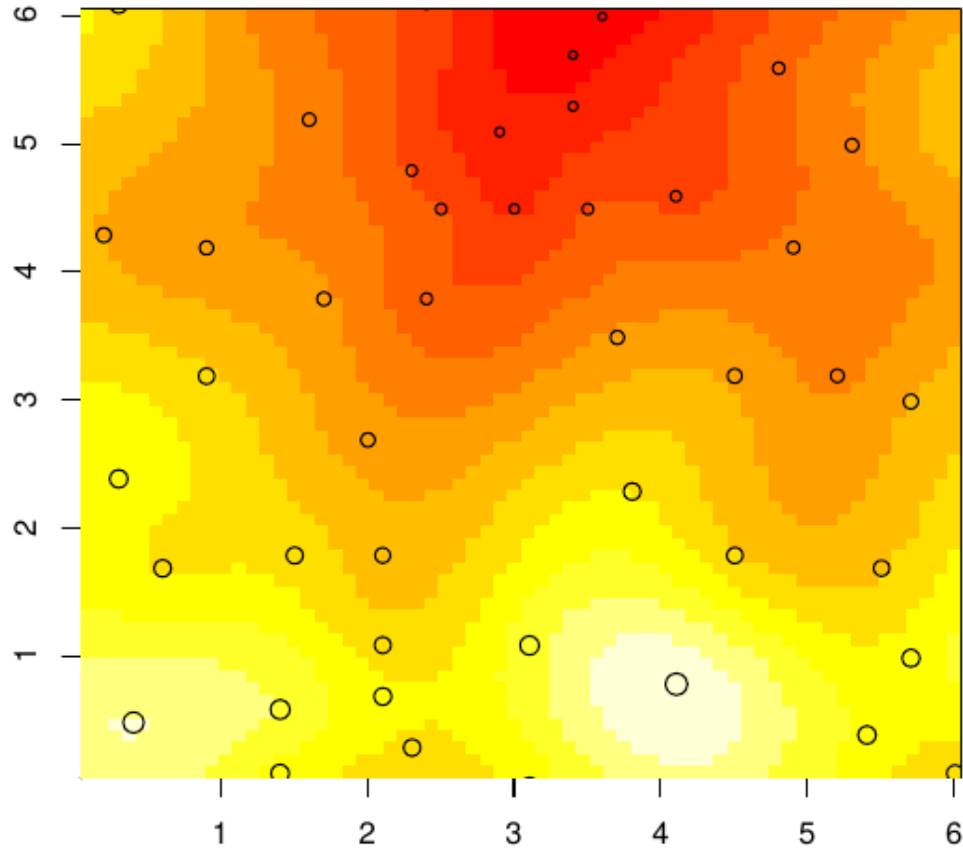
Measure prevalence at sampling locations

Survey sampling analysis

41.3 ± 8.6

Why use model-based geostatistics?

Area-wide prevalence = 41.7%



Measure prevalence at sampling locations

Survey sampling analysis

41.3 ± 8.6

Geostatistical analysis

41.6 ± 1.4

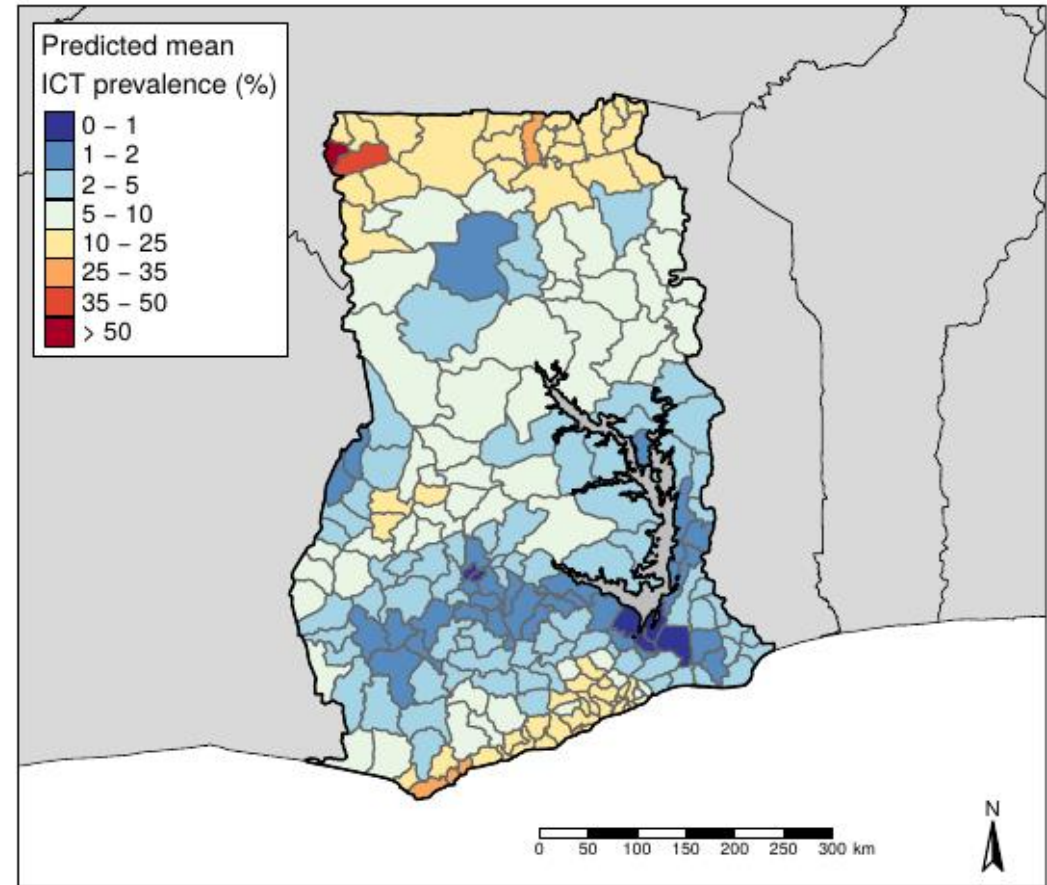
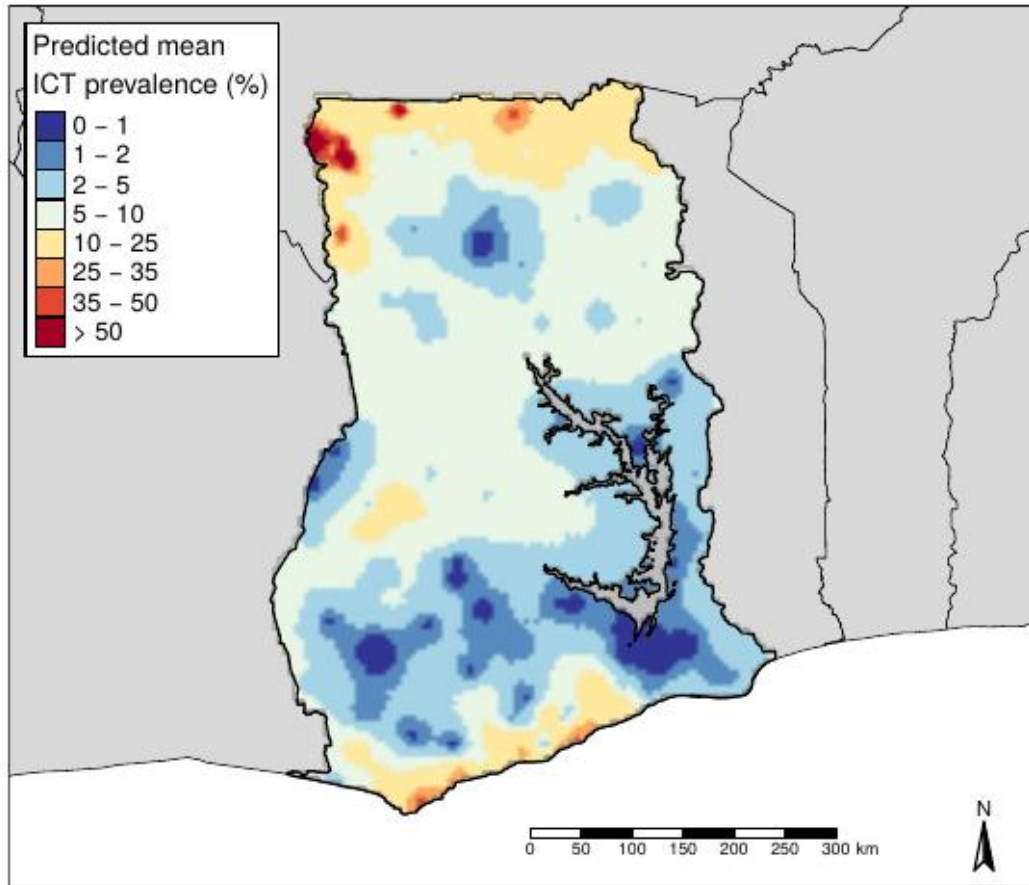
Where does the extra efficiency come from?

Exploiting spatial correlation

- Data at location x are informative of prevalence at “neighbouring” locations x'
- Model-fitting process allows data to choose optimal definition of “neighbouring”

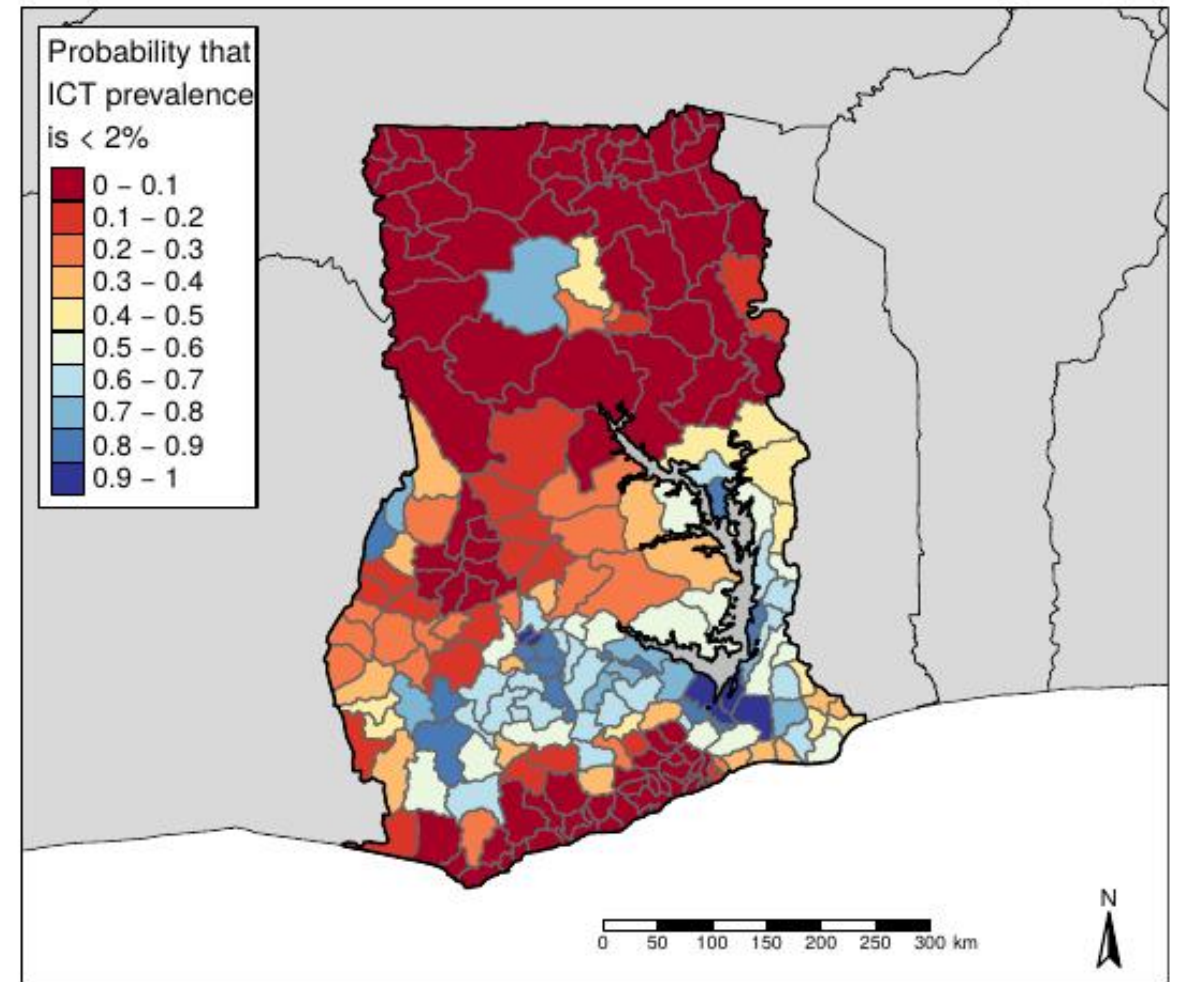
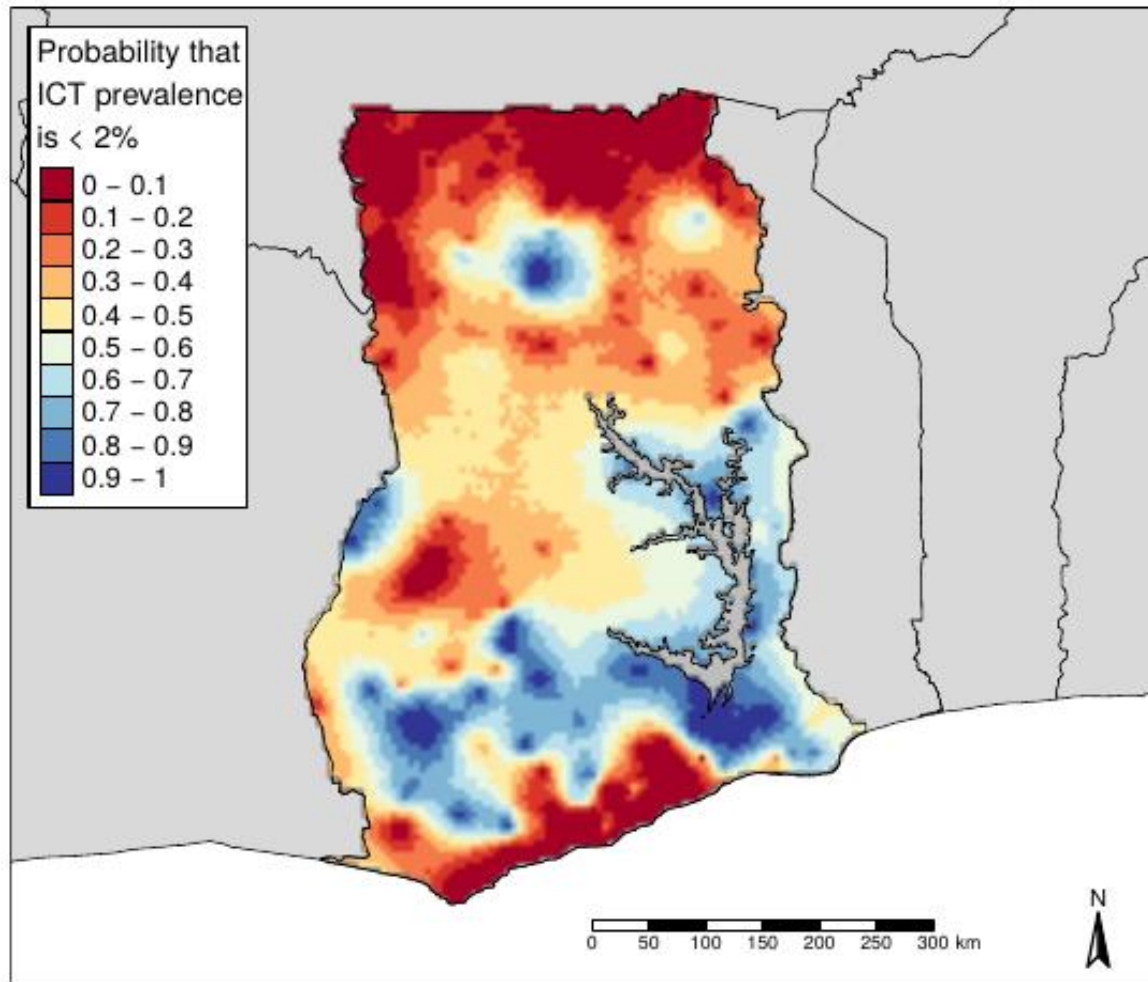
Model-based geostatistics

Exploits **spatial correlation** present in the data to predict the summary of interest at any **unsampled location** and any **desired spatial scale**.



Model-based geostatistics

Honest reporting of the **uncertainty** around the prediction target of interest.



How about model-uptake?

Statistical and technical skills required to implement these approaches.

```
# MODEL-BASED DECISION WITH GEOSTATISTICS -----  
for (i in 1:nrow(scenarios)) {  
  
  message("Doing scenario ", scenarios$ID[i])  
  
  # Assign number of children to test according to scenario  
  schools$examined <- scenarios$children[i]  
  
  # Simulate number of children tested positive according to  
  # underlying true prevalence  
  schools$positive <- rbinom(n = nrow(schools),  
                           size = schools$examined,  
                           prob = schools$true_prev)  
  schools <- as.data.frame(schools)  
  
  # Sample schools accordingly to the sampling strategy  
  # (either random or regular)  
  if(scenarios$sampling[i] == "random") {  
    sample_id <- sample(nrow(schools), scenarios$nschools[i], size = )  
    schools_sampled <- schools[sample_id, ]  
  } else {  
    delta <- 10 # minumun distance allowed between pairs (in km)  
    sample_id <- geo_sample(xy.all = schools[c("utm_x", "utm_y")],  
                           n = scenarios$nschools[i],  
                           delta = delta, k = 0)  
    schools_sampled <- schools[sample_id, ]  
  }  
  
  # Fit geostatistical model to simulated data  
  schools_sampled$logitp <- with(schools_sampled, elogit(positive, examined))  
  schools_sampled$prev <- schools_sampled$positive / schools_sampled$examined  
}
```

User-friendly web applications

STH impact app

Import Data

EDA ▾

Analyse Data

Design Survey

Baseline Survey Data

Browse... No file selected

MDA Data

Browse... No file selected

Geographical boundaries

Browse... No file selected

Schools

Browse... No file selected

Instructions for uploading data

In this tab you need to upload four pieces of information:

- Baseline prevalence survey data
- MDA data that reports the treatment history for each implementation unit (IU) from baseline to impact.
- Geographic boundaries of the IUs
- Schools database

Please follow the instructions below.

Baseline survey data

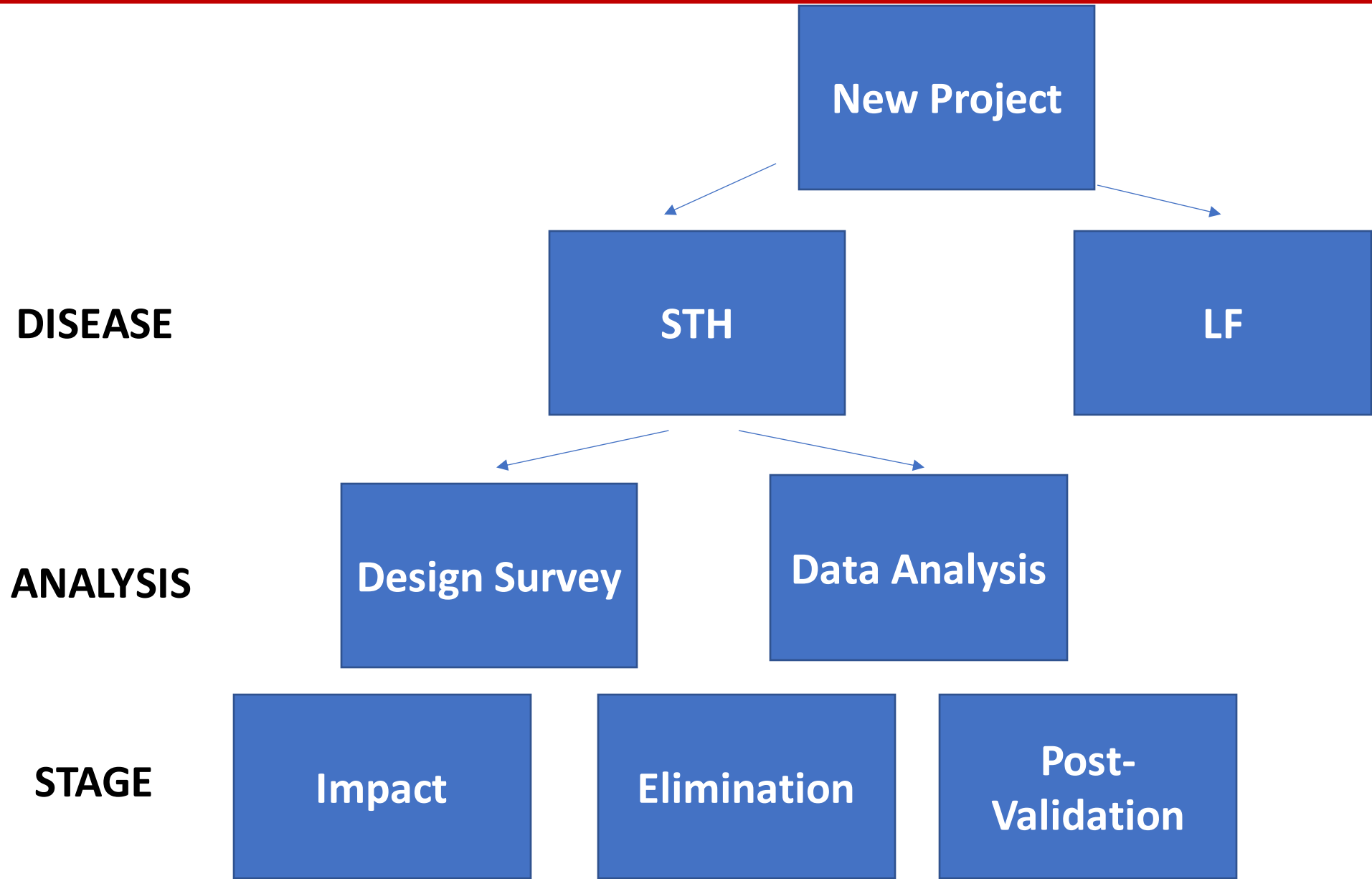
Pre-treatment prevalence data needs to be provided as a `.csv` file. A template that shows an extract of a sample dataset can be downloaded from [here](#).

Be careful to use the same column names provided in the template. The geographical coordinates for each school needs to be in longitude and latitude.

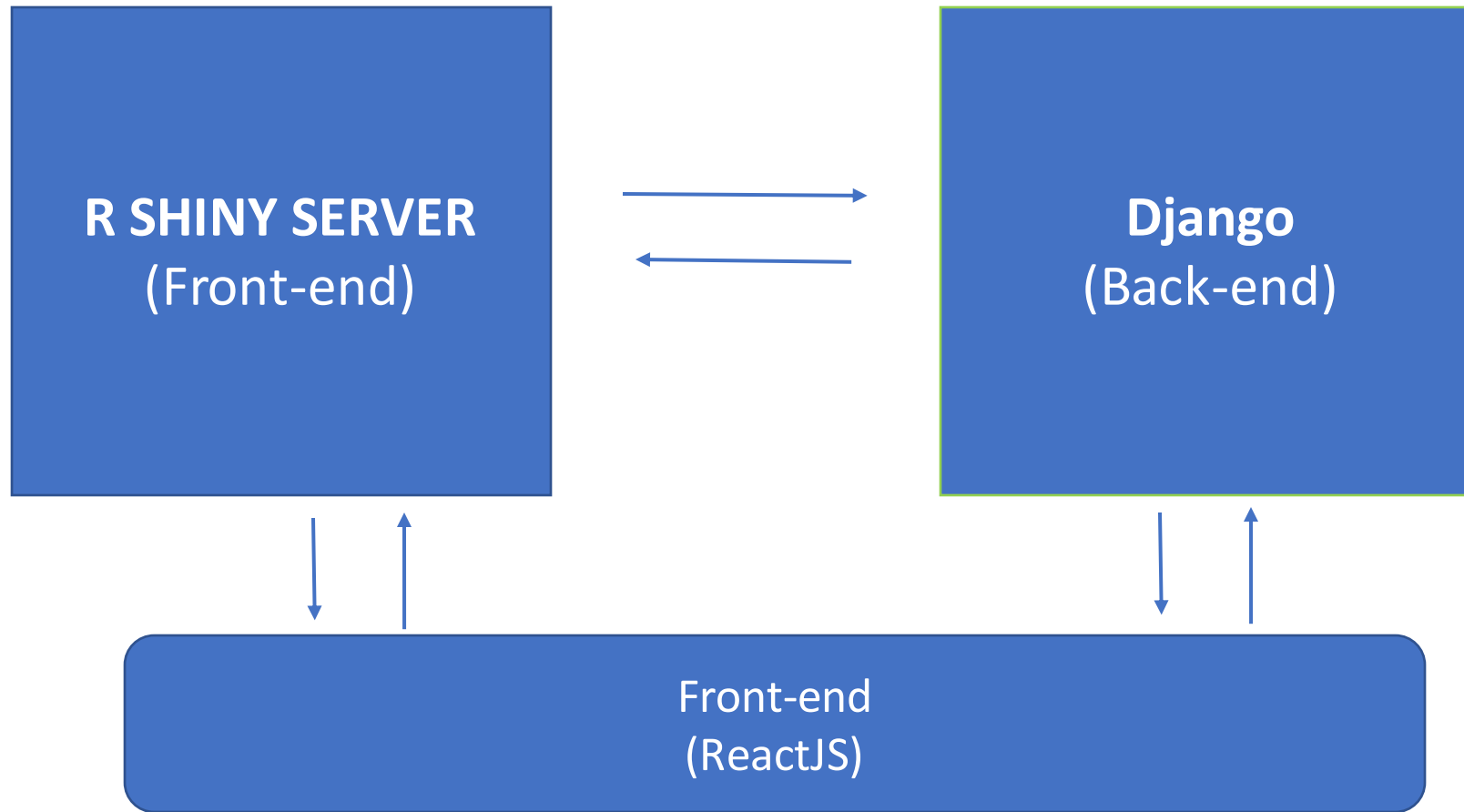
Here is a brief description of the variables needed:

- `school_id` : unique identifier of the school
- `lon` : longitude
- `lat` : latitude
- `exam` : number of individuals examined
- `pos` : number of individuals who test positive
- `specie` : STH specie (either `ascaris` , `trichuris` or `hookworm`)
- `year` : year when the survey was conducted.

Geostat NTD Hub



Geostat NTD Hub



References

Fronterre C, Amoah B, Giorgi E, Stanton MC, Diggle PJ. Design and analysis of elimination surveys for neglected tropical diseases. *J Infect Dis.* 2020;221: S554–S560.

Diggle PJ, Amoah B, Fronterre C, Giorgi E, Johnson O. Rethinking neglected tropical disease prevalence survey design and analysis: a geospatial paradigm. *Trans R Soc Trop Med Hyg.* 2021;115: 208–210. doi:10.1093/trstmh/trab020

Johnson O, Fronterre C, Amoah B, Montresor A, Giorgi E, Midzi N, et al. Model-Based Geostatistical Methods Enable Efficient Design and Analysis of Prevalence Surveys for Soil-Transmitted Helminth Infection and Other Neglected Tropical Diseases. *Clin Infect Dis.* 2021;72: S172–S179. doi:10.1093/cid/ciab192

Amoah B, Fronterre C, Johnson O, Dejene M, Seife F, Negussu N, et al. Model-based geostatistics enables more precise estimates of neglected tropical-disease prevalence in elimination settings: mapping trachoma prevalence in Ethiopia. *Int J Epidemiol.* 2022;51: 468–478. doi:10.1093/ije/dyab227

Thank you!