

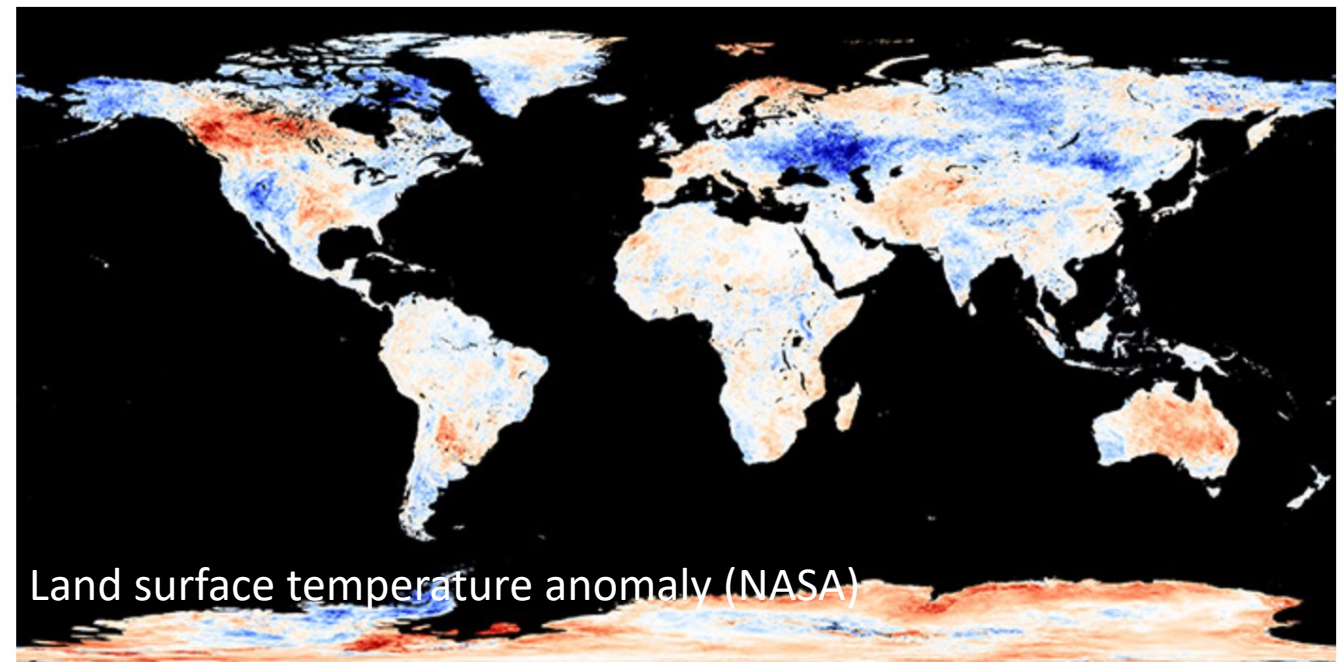
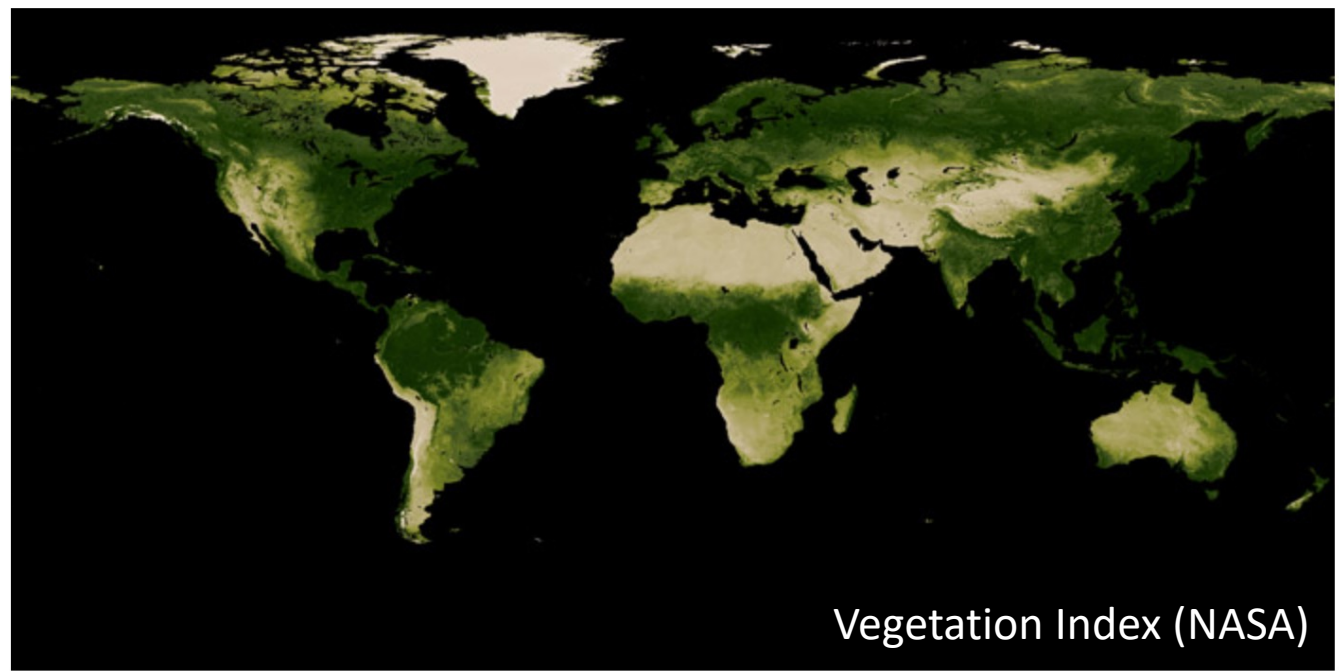
# Comparison of geostatistical methods for spatial data

Sevvandi Kandanaarachchi, Spencer Wong, Jennifer Flegg, Nick Golding

AASC 2022

# Geospatial data is everywhere

- Climate
- Disease
- Energy
- Satellite data



# The original problem

- Data at different locations  $(\mathbf{x}_i, y_i)$ 
  - $\mathbf{x}_i$  - location data  $(x_{i1}, x_{i2})$ , 2D or 3D
  - $y_i$  - quantity of interest (dependent variable)
    - Temperature, disease prevalence
- Want to predict  $y_k$  at a different  $\mathbf{x}_k$ , where data is not present
- Or map it to a whole region/country
- Covariates of interest
  - Altitude, population density

# The original problem

- $y_i = f(x_i, u_i, \epsilon_i)$
- $u_i$  - covariates
- $\epsilon_i$  - error term
- Methods adopt certain assumptions and approximations
  
- General principle – Tobler – “everything is related to everything else, but near things are more related than distant things”

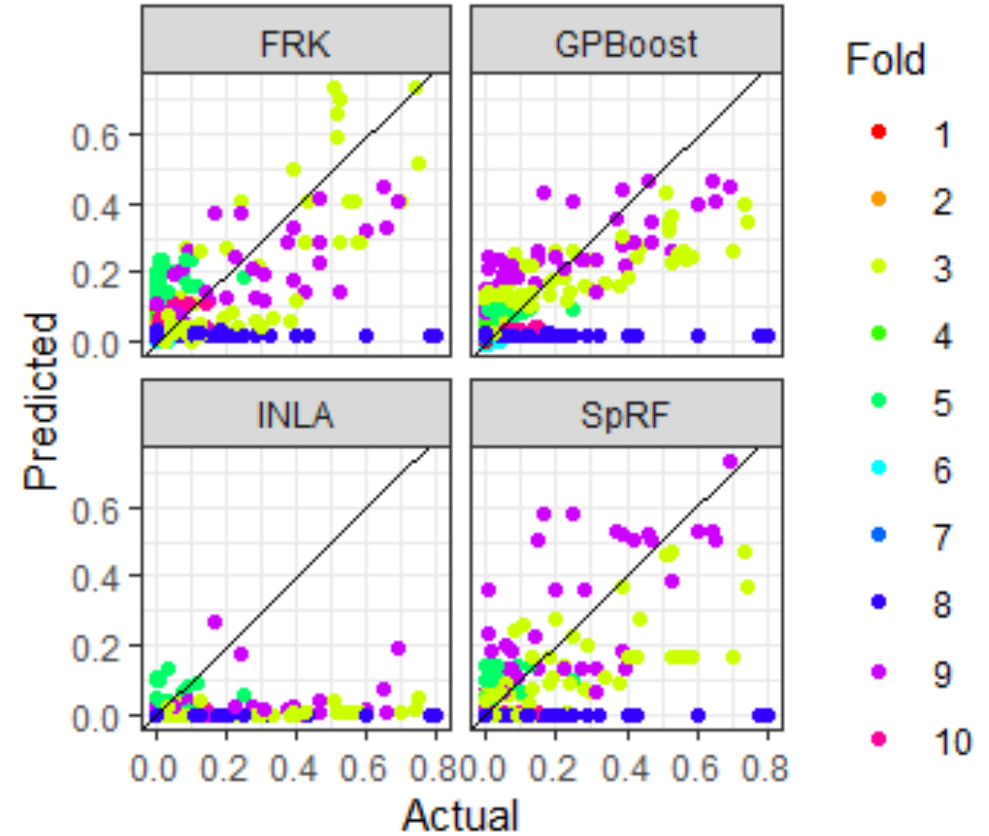
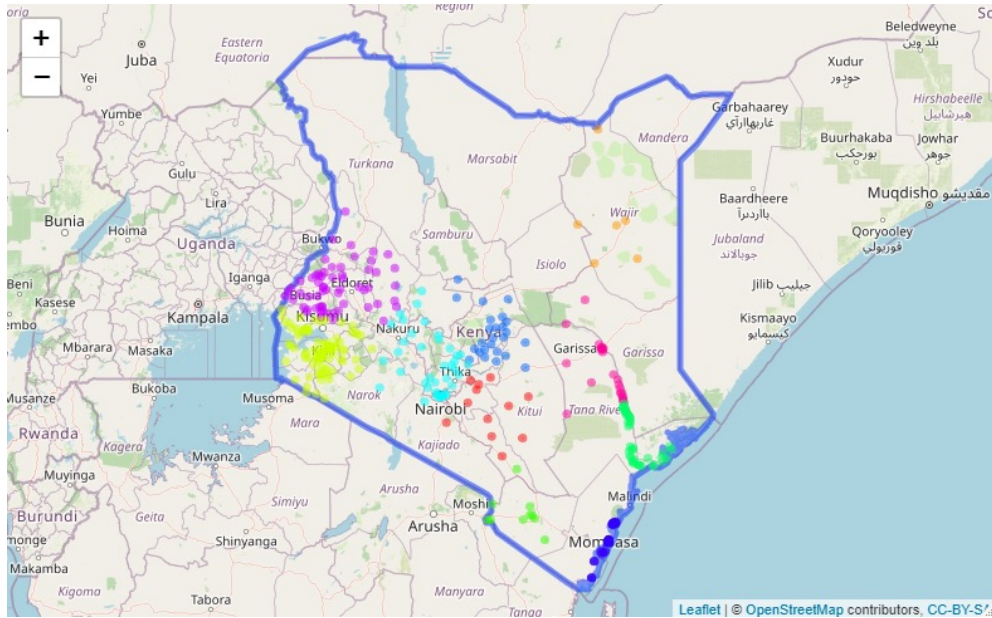
# Methods we compare

- INLA – Integrated Nested Laplace Approximations
- SpRF – Spatial Random Forests
- FRK – Fixed Rank Kriging
- GPBoost – Tree boosting with Gaussian Processes
  
- Standard parameters for all methods

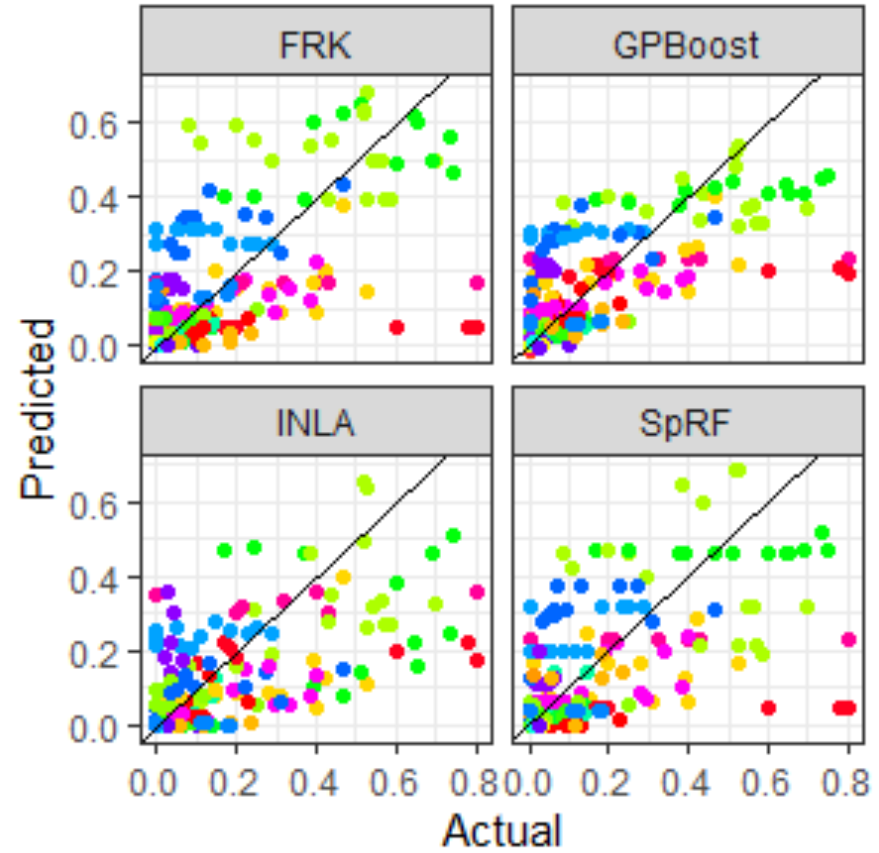
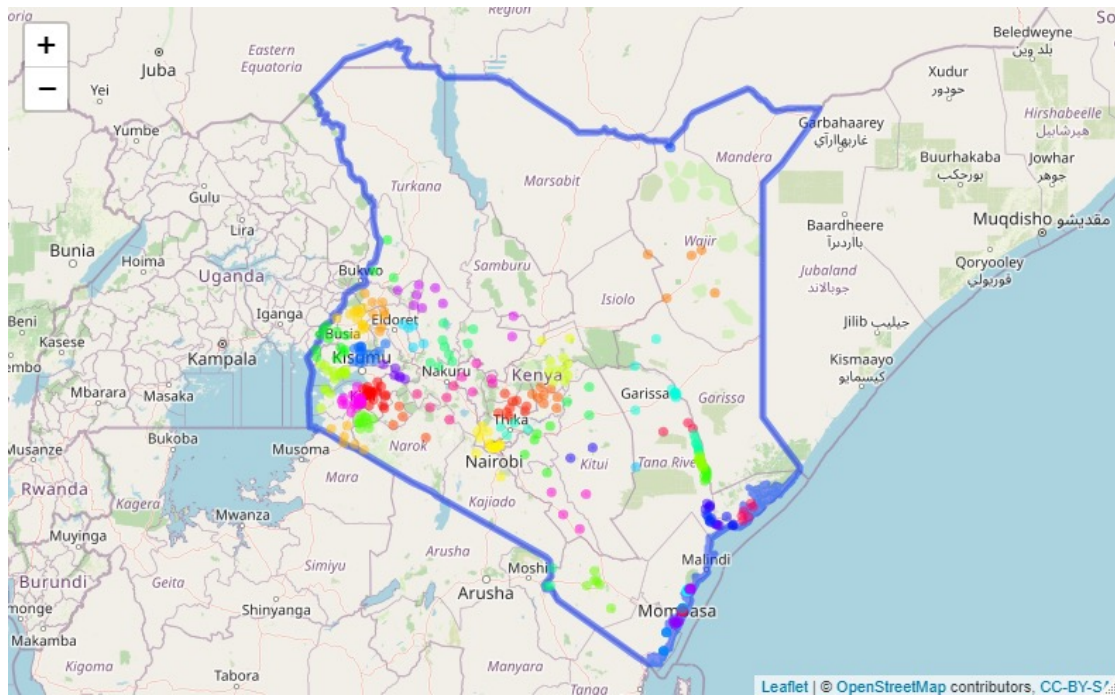
# Computation

- Data – latitude, longitude, number of people tested, number of people positive
- Prevalence = number positive/number tested
- Kenya data in 2009
  - 10 and 50-fold CV
  - Map of Kenya
- Africa
  - Africa maps

# 10-fold CV

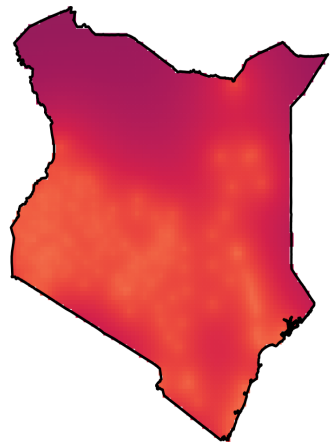
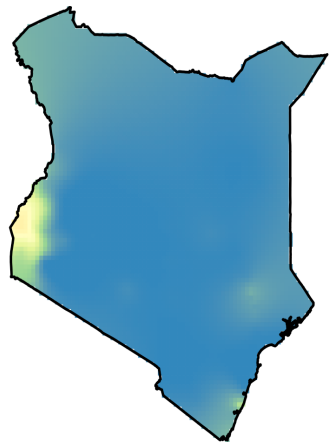
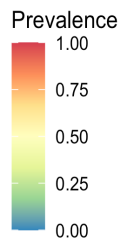


# 50-fold CV

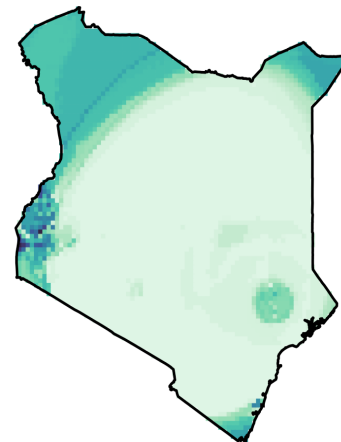
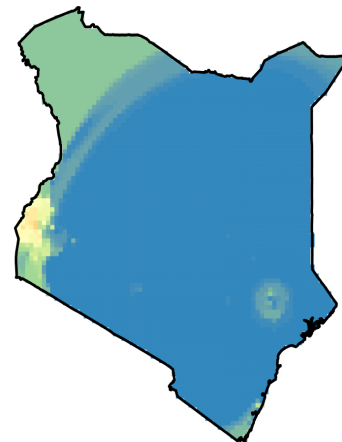
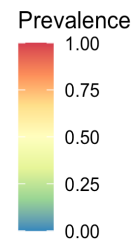




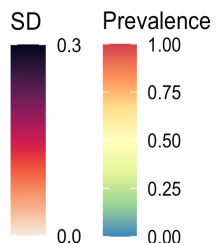
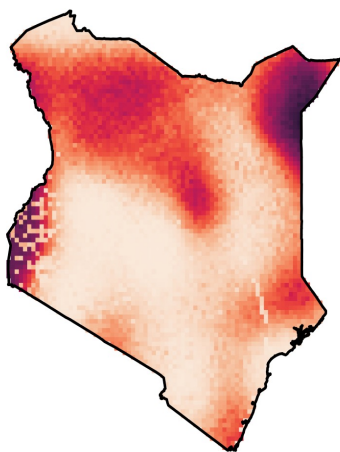
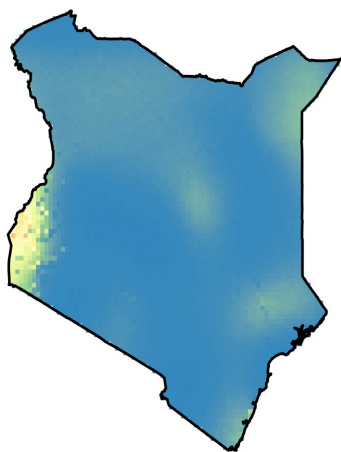
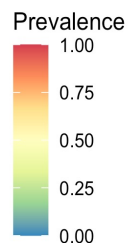
GPBoost



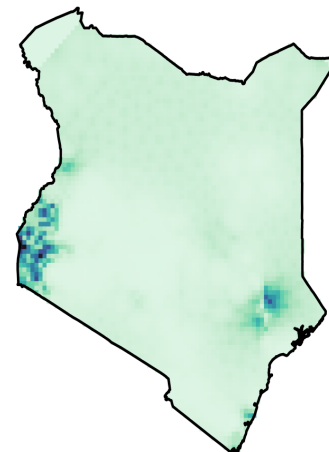
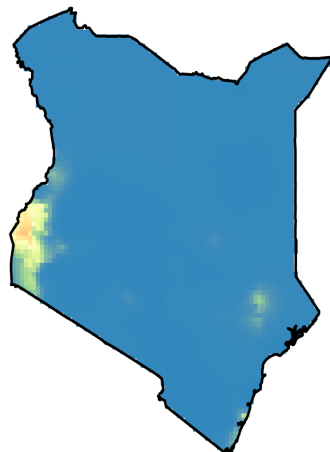
SpRF



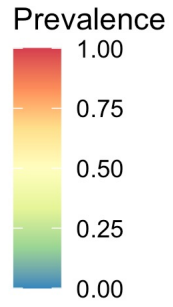
FRK



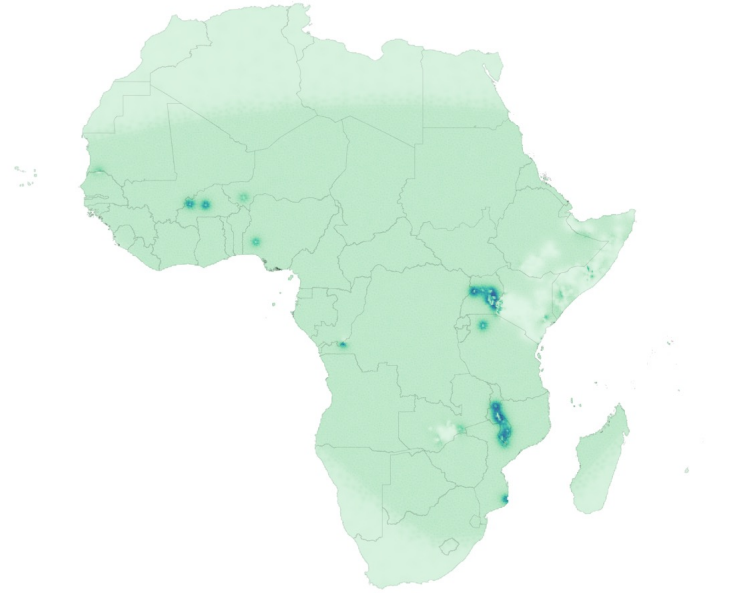
INLA



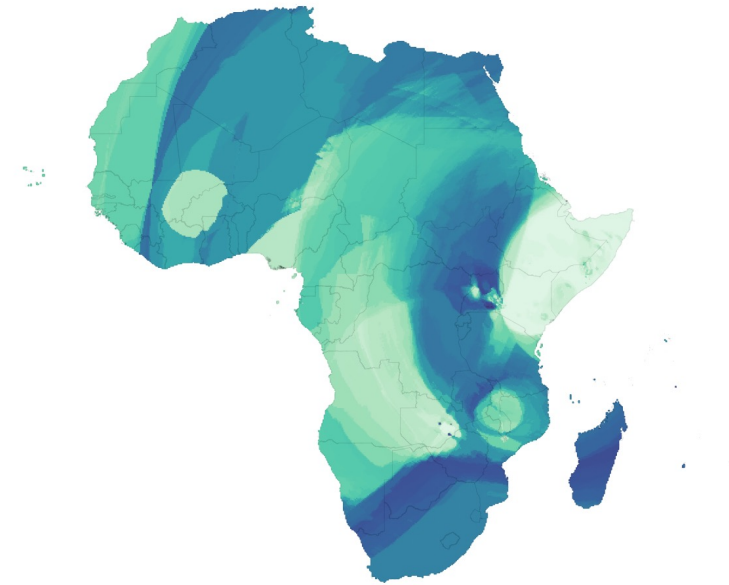
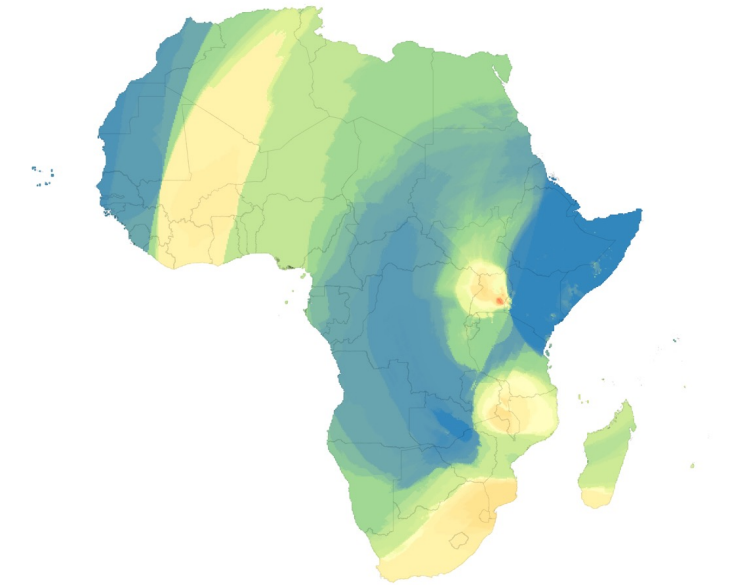
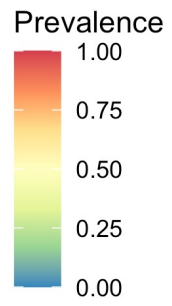
INLA



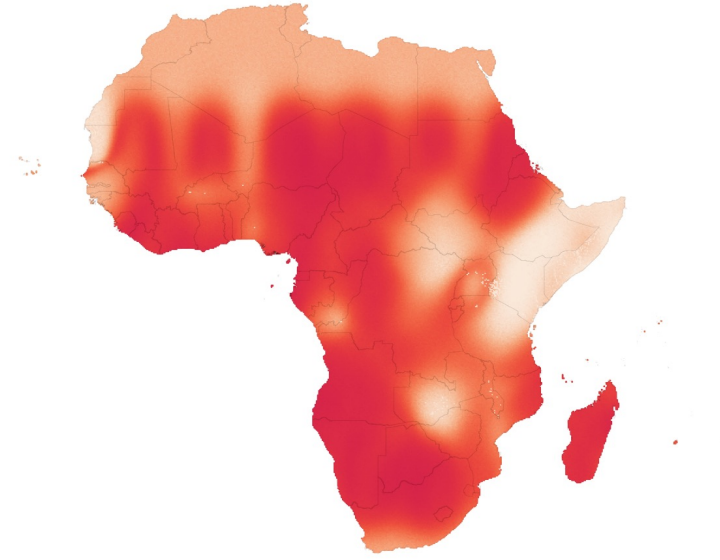
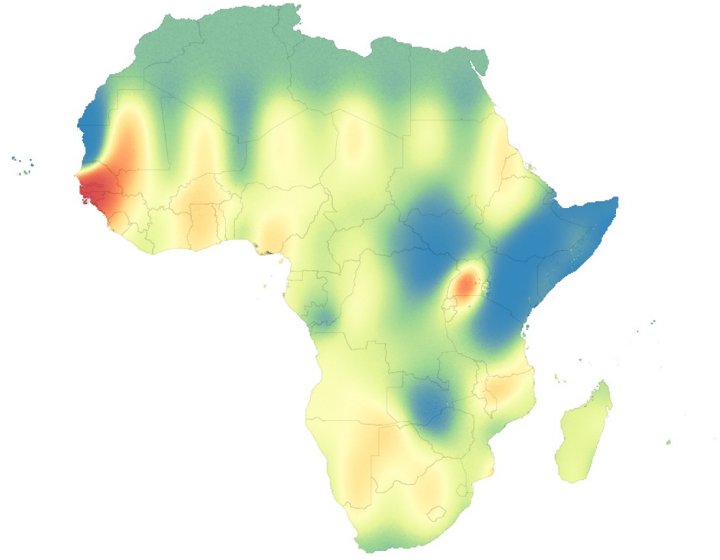
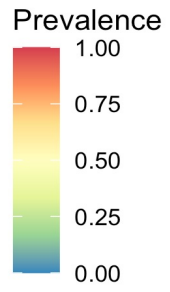
Africa



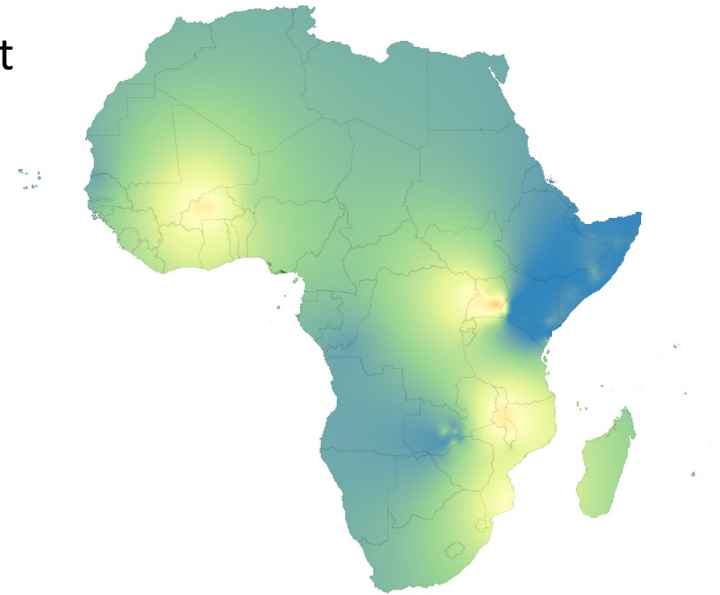
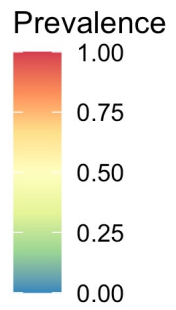
SpRF



FRK



GPBoost



# Reflections

- Maps get a bit messier when we model Africa
- Need to tune the parameters
- Sparsity of points
- See the artefacts of the methods more when you go for bigger scales
- For bigger scale maps – users need to understand the methods more
- Not off the shelf– at this point for bigger regions!

Thank you!

Questions?