# Estimating spatial sub-national variations from Demographic and Health Surveys data

J. Larmarange[a], R. Vallo[b], S. Yaro[c]

P. Mselatti[d], N. Méda[c], B. Ferry[a]

[a] IRD / CEPED UMR 196 • [b] Université Montpellier EA 4205

[c] Centre Muraz • [d] UMR 145 IRD / Université de Montpellier I

**IUSSP Conference • Marrakech • October 2nd 2009**

**joseph.larmarange@ceped.org**

CePeD Centre Population & Développement

prevR

IRD Institut de recherche pour le développement

centre MURAZ

# Demographic and Health Surveys (DHS)

- DHS are well known by demographers.
- Since 1984, more than 200 DHS have been conducted in several low- and middle-income countries, mostly in Africa.

- On certain topics, HIV prevalence for example, DHS often constitute the only national source of data in general population.
- Some DHS collect longitude and latitude of surveyed clusters, but this information is rarely used for analyses.

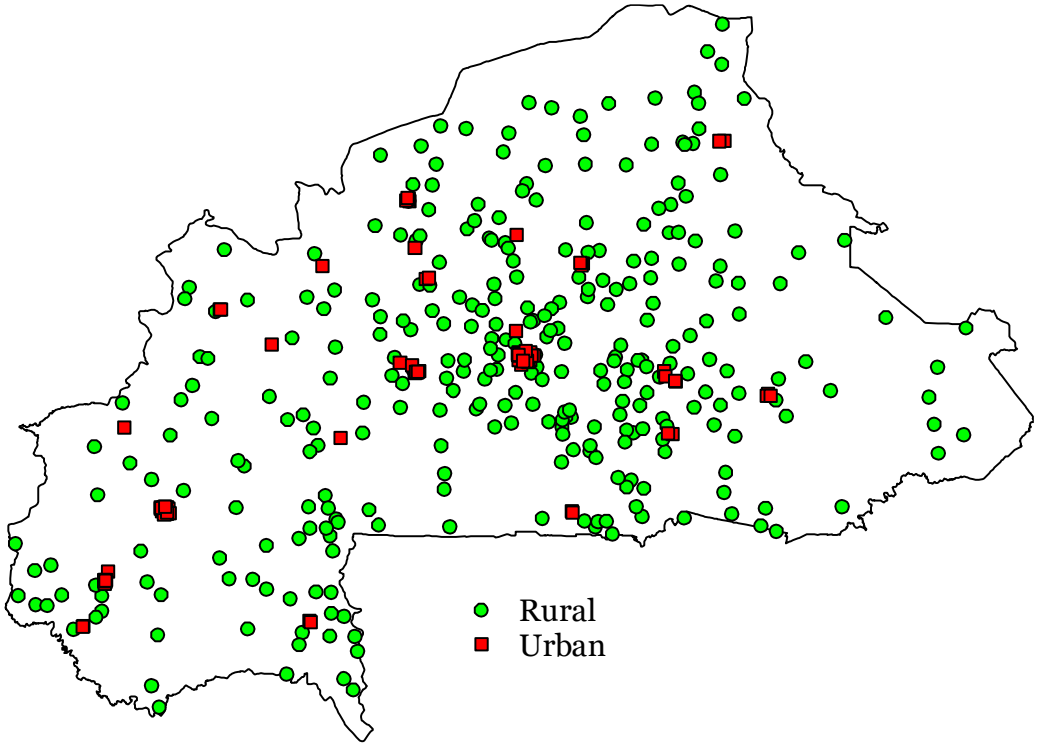
- DHS are designed to be representative at national and regional level. But maps by region are often not adapted to explore spatial variations of a phenomenon.

# DHS Sampling: example of Burkina Faso (DHS 2003)

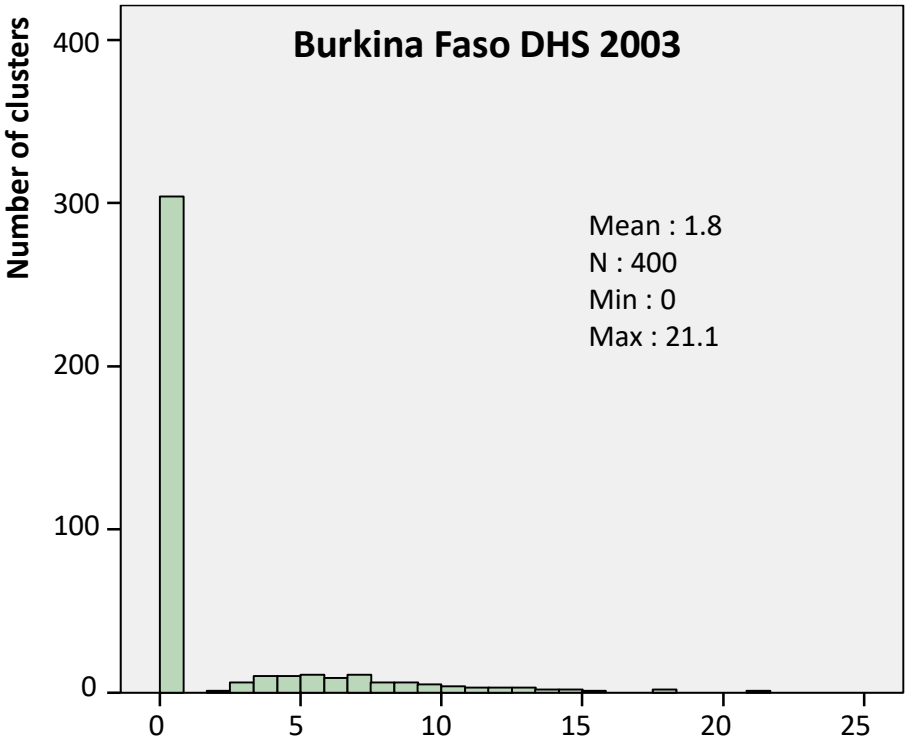
- Burkina Faso 2003
  - 400 clusters
  - 3179 households • 7151 persons tested for HIV

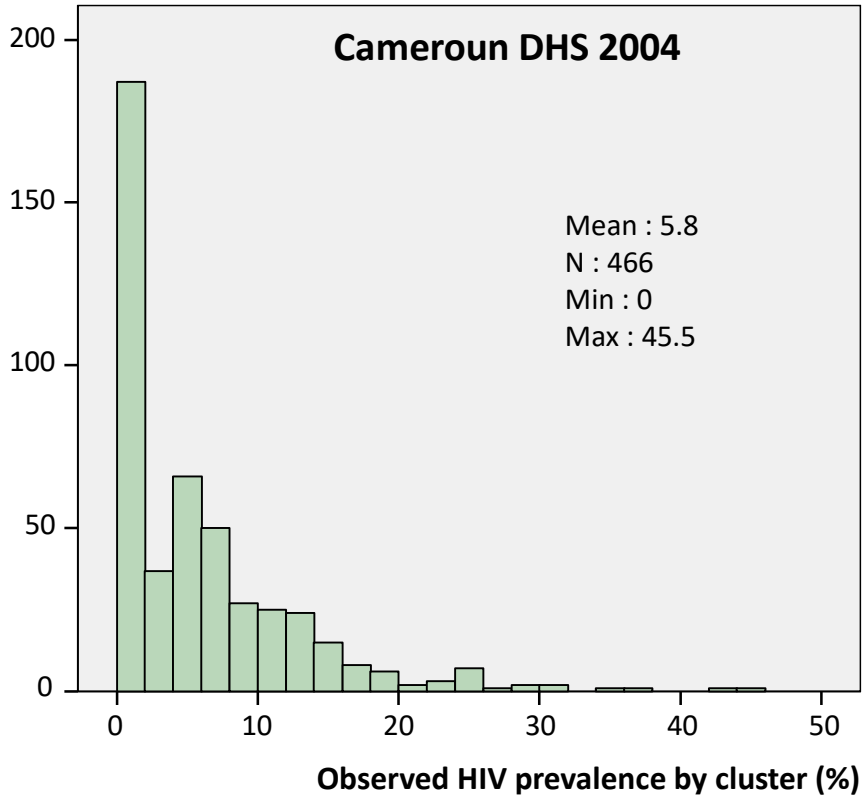
  ➔ 17.9 tested persons by cluster

- Depending of the DHS, number of tested persons by cluster varies from 10 to 35.



# Observed prevalence by cluster




*Observed prevalence = Cluster prevalence + Random error (RE)*

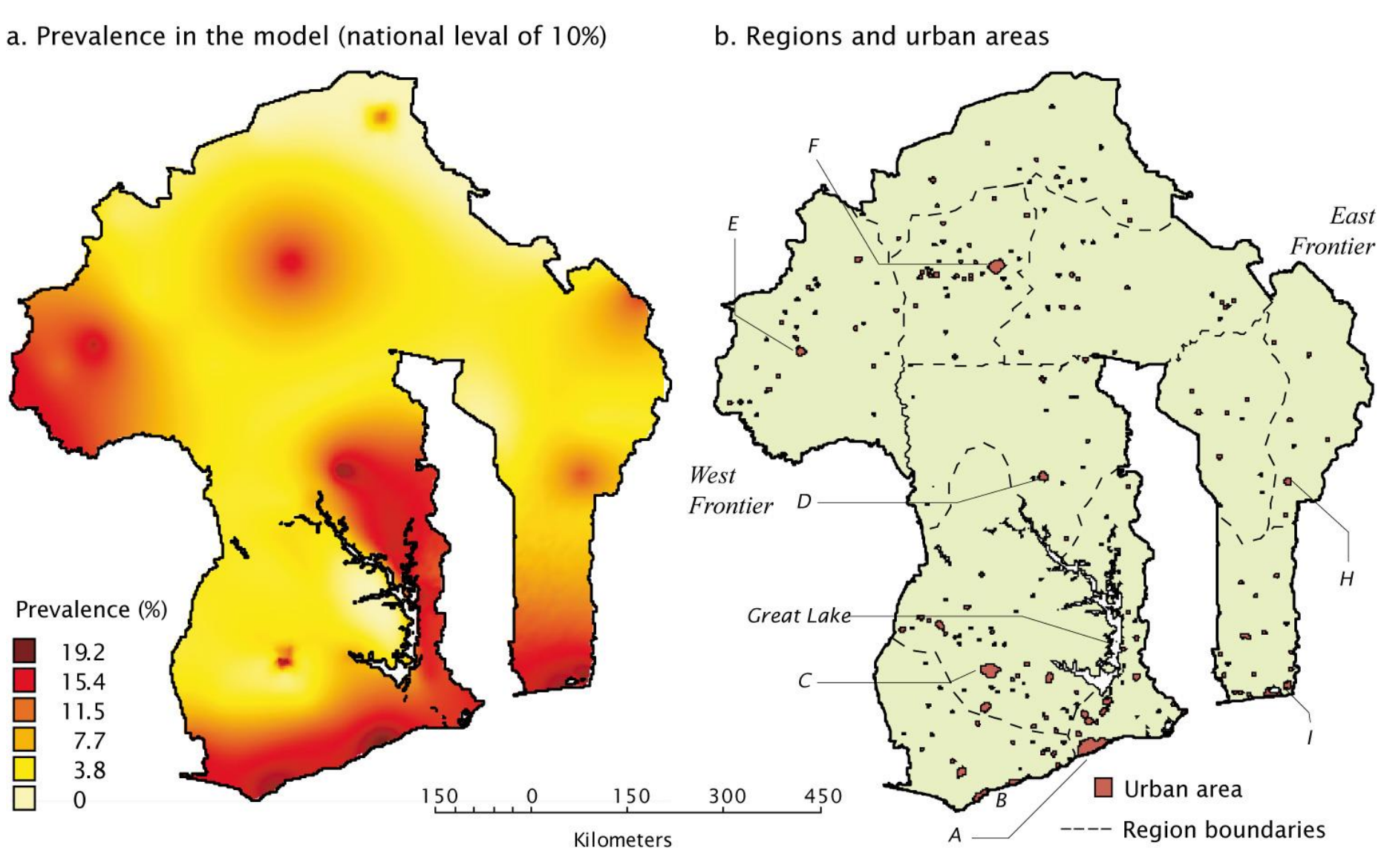
# Approach developed

- Inspired by trend-surface mapping.
- Observed prevalence = regional trend + local component + RE
- Is it possible to estimate for each cluster a regional trend ?

- To test the methodology, elaboration of a fictive country with a fictive epidemic.

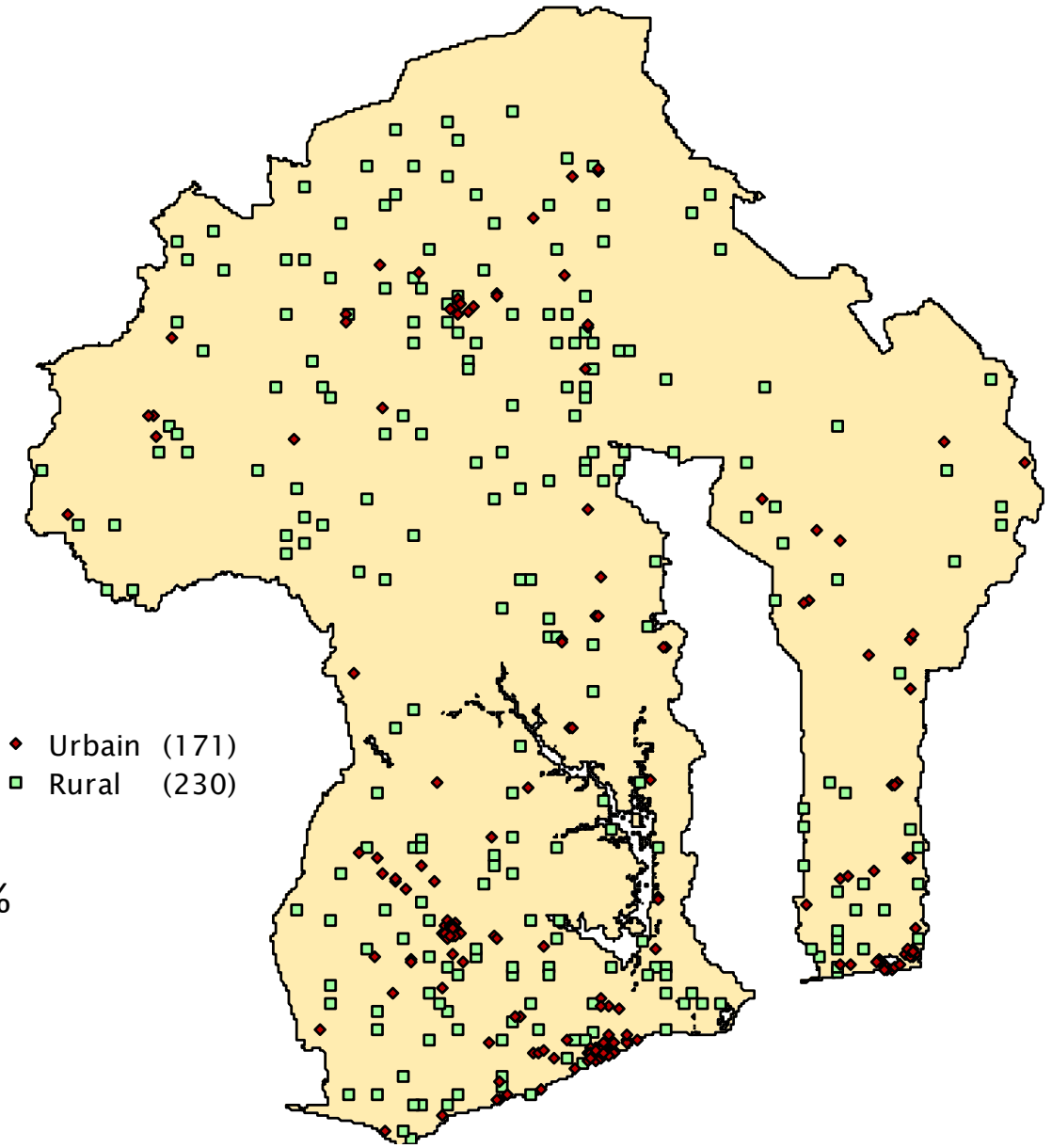
- Objective:
  - not a precise estimation for each point of the final map,
  - but to reproduce the main spatial variations of the phenomenon.

# The fictive country

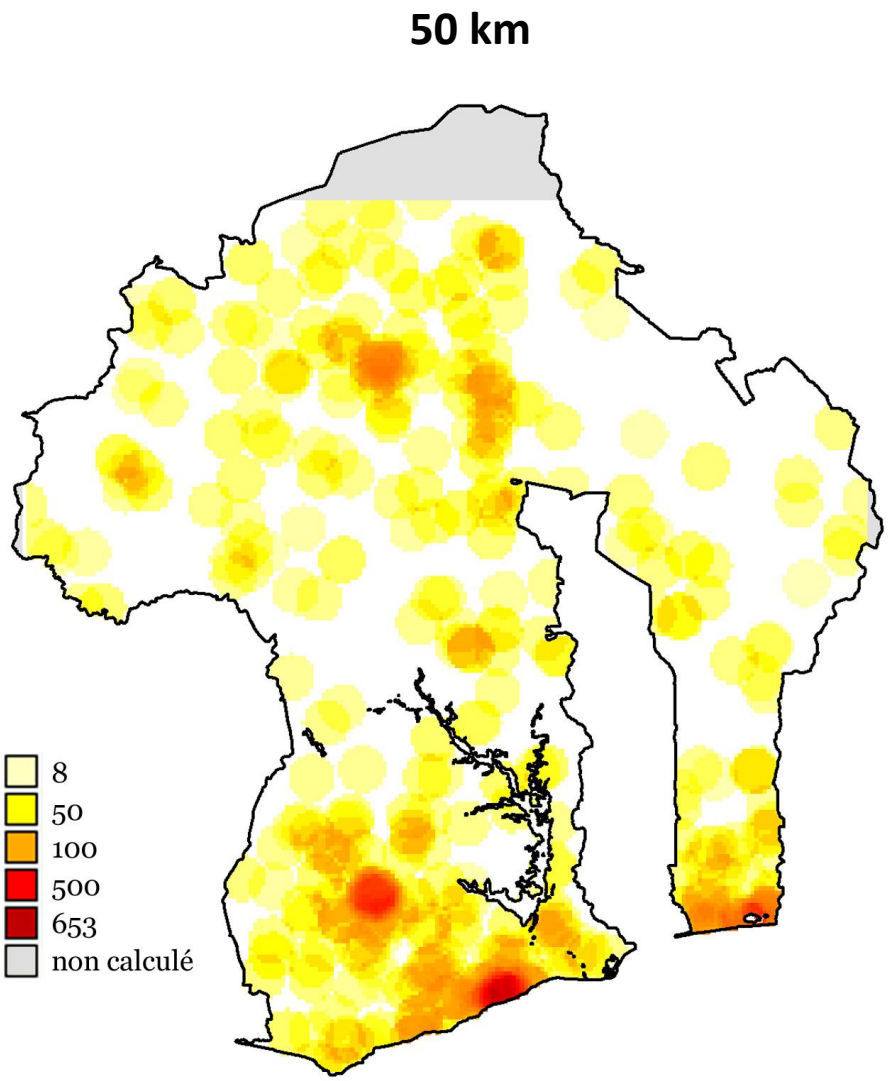


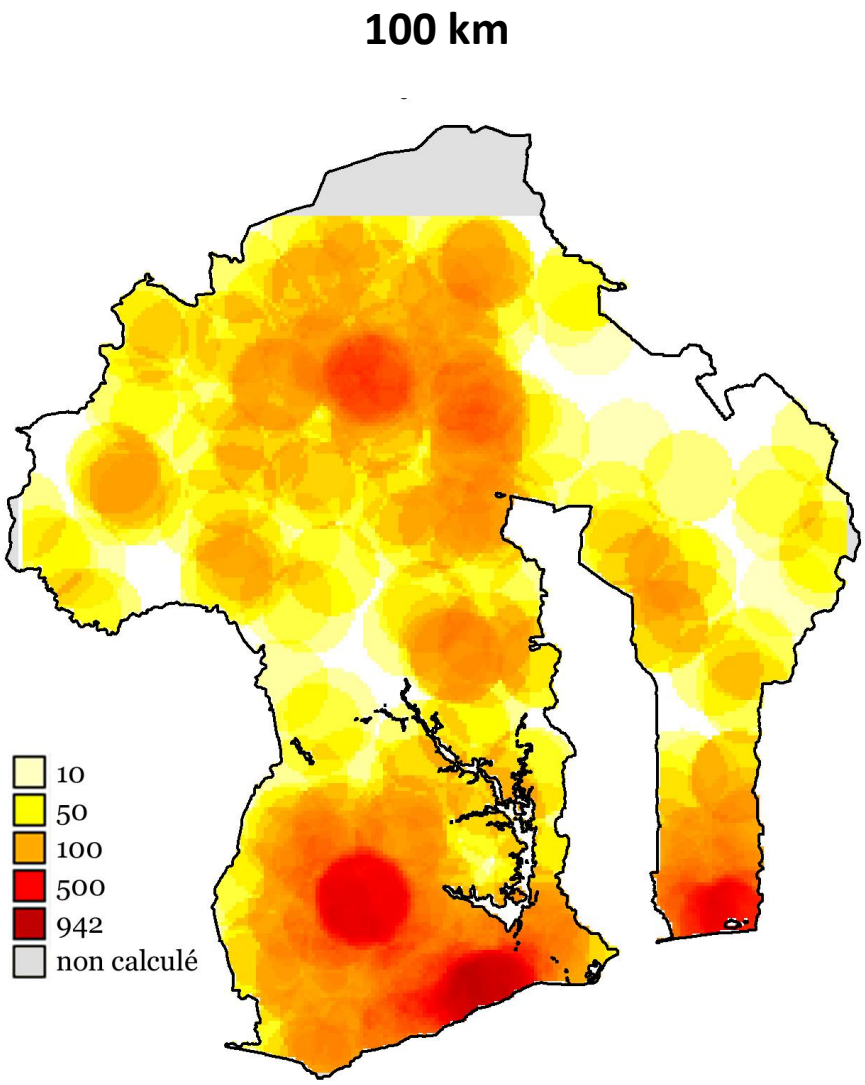
# DHS simulation

- One example with:
  - National prevalence of 10%
  - 8000 tested individuals
  - 400 clusters

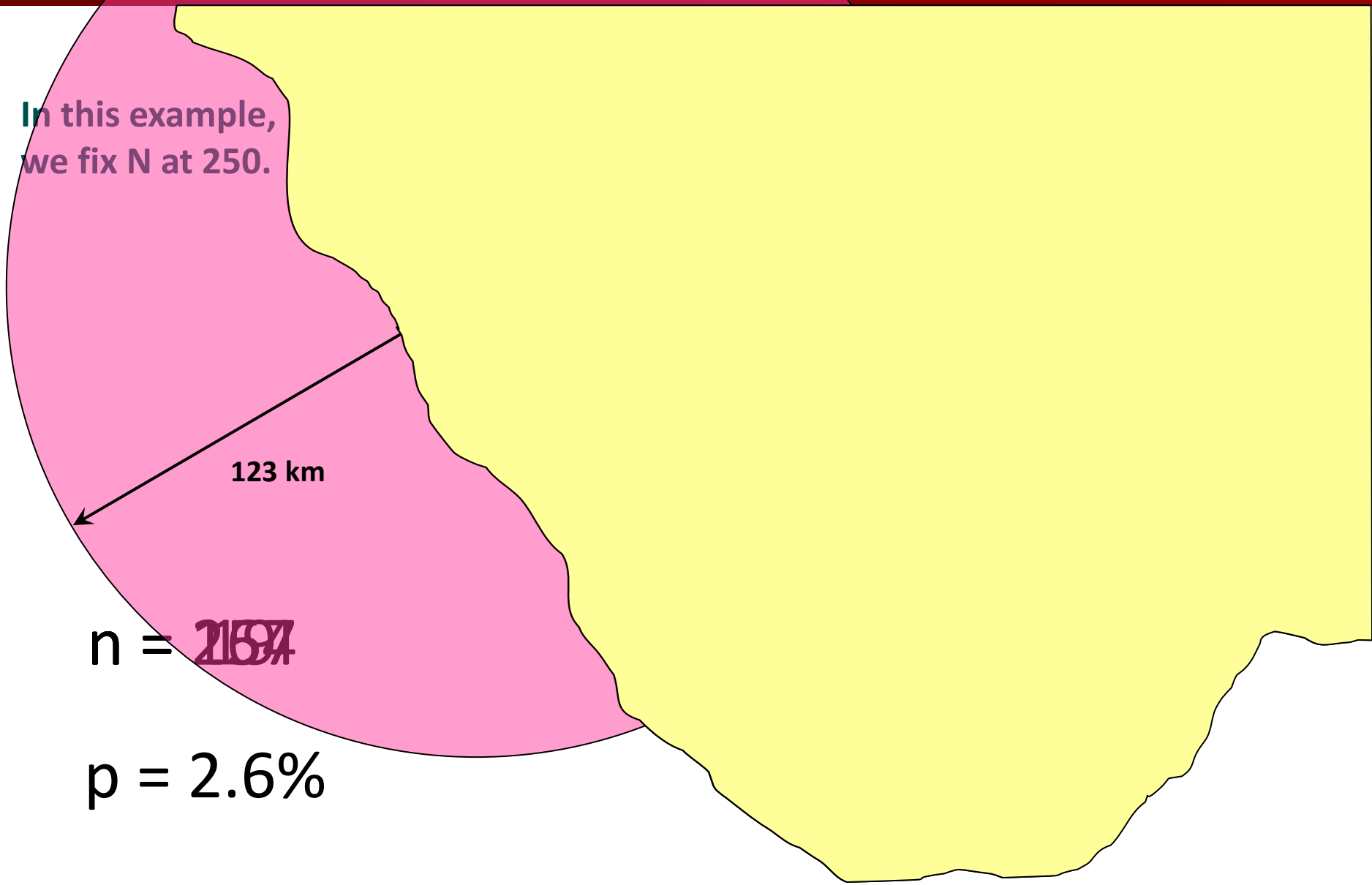


# Observations in a radius of 50 or 100 kilometers

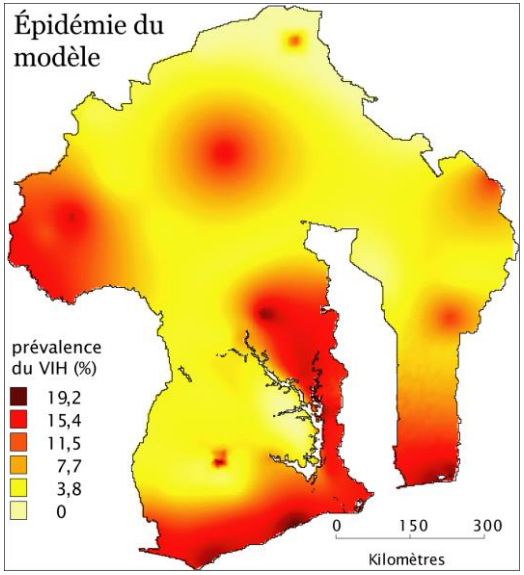


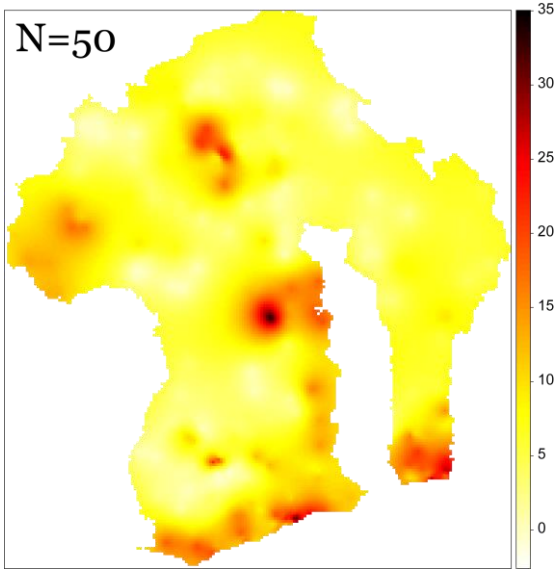



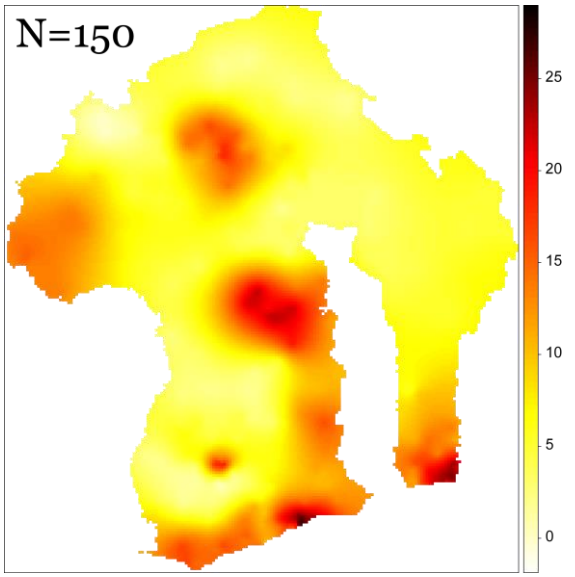
# N method: circle of same number of observations

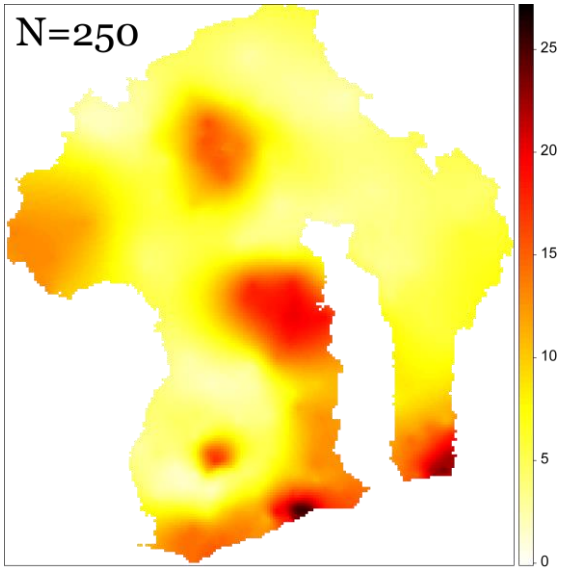


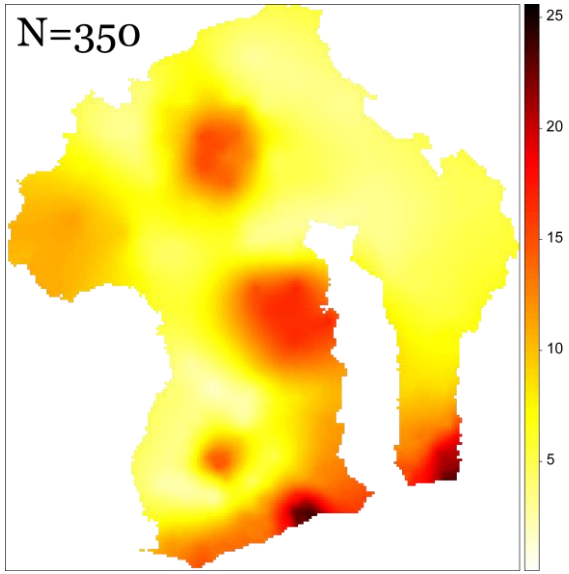
# N method and spatial interpolation by kriging

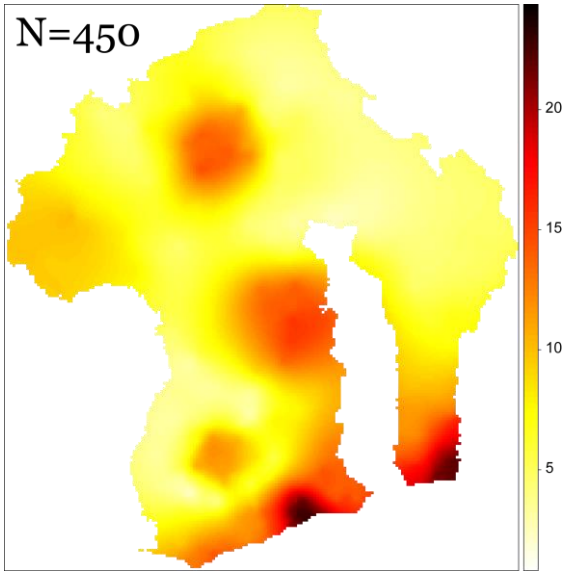


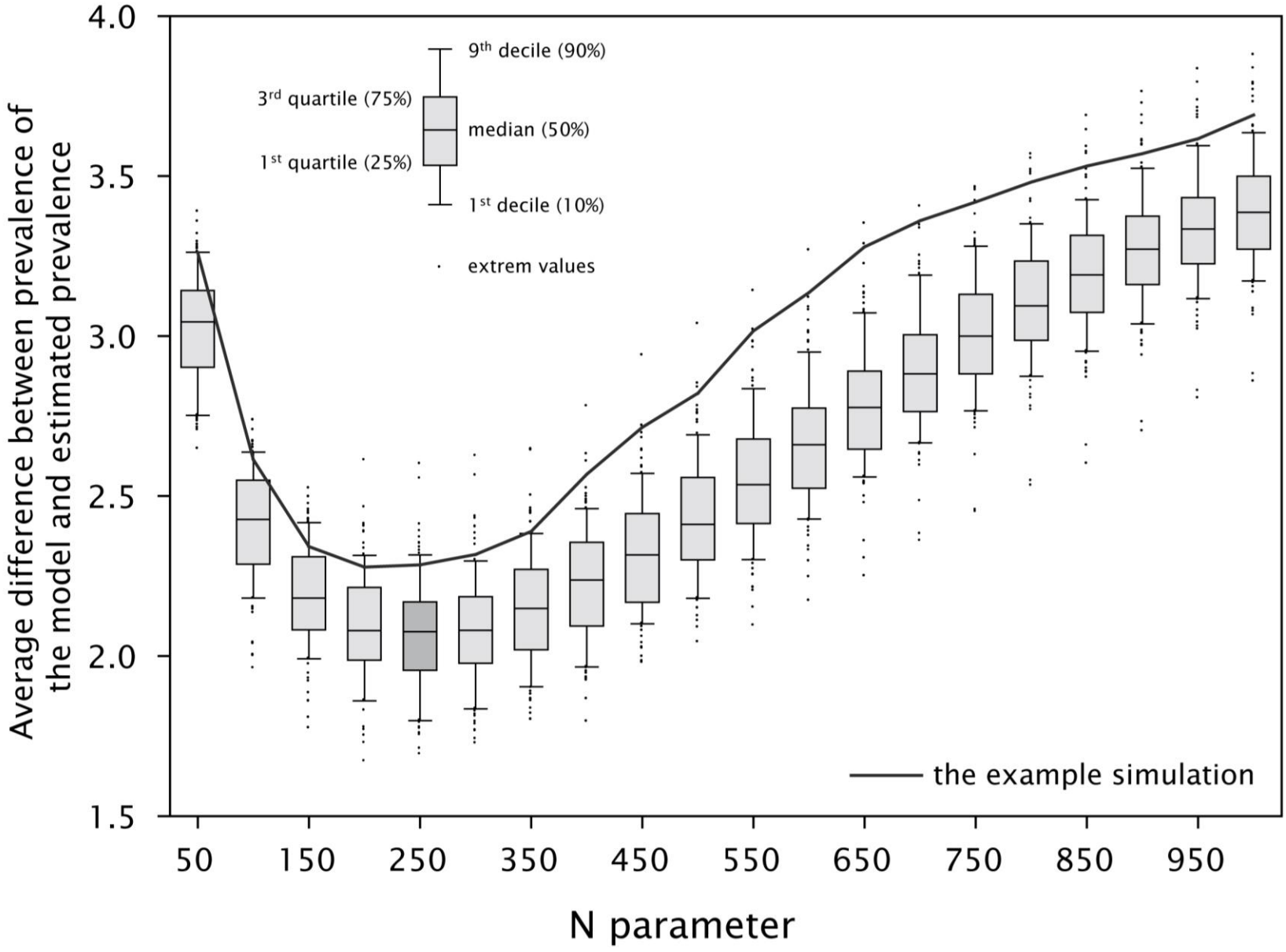



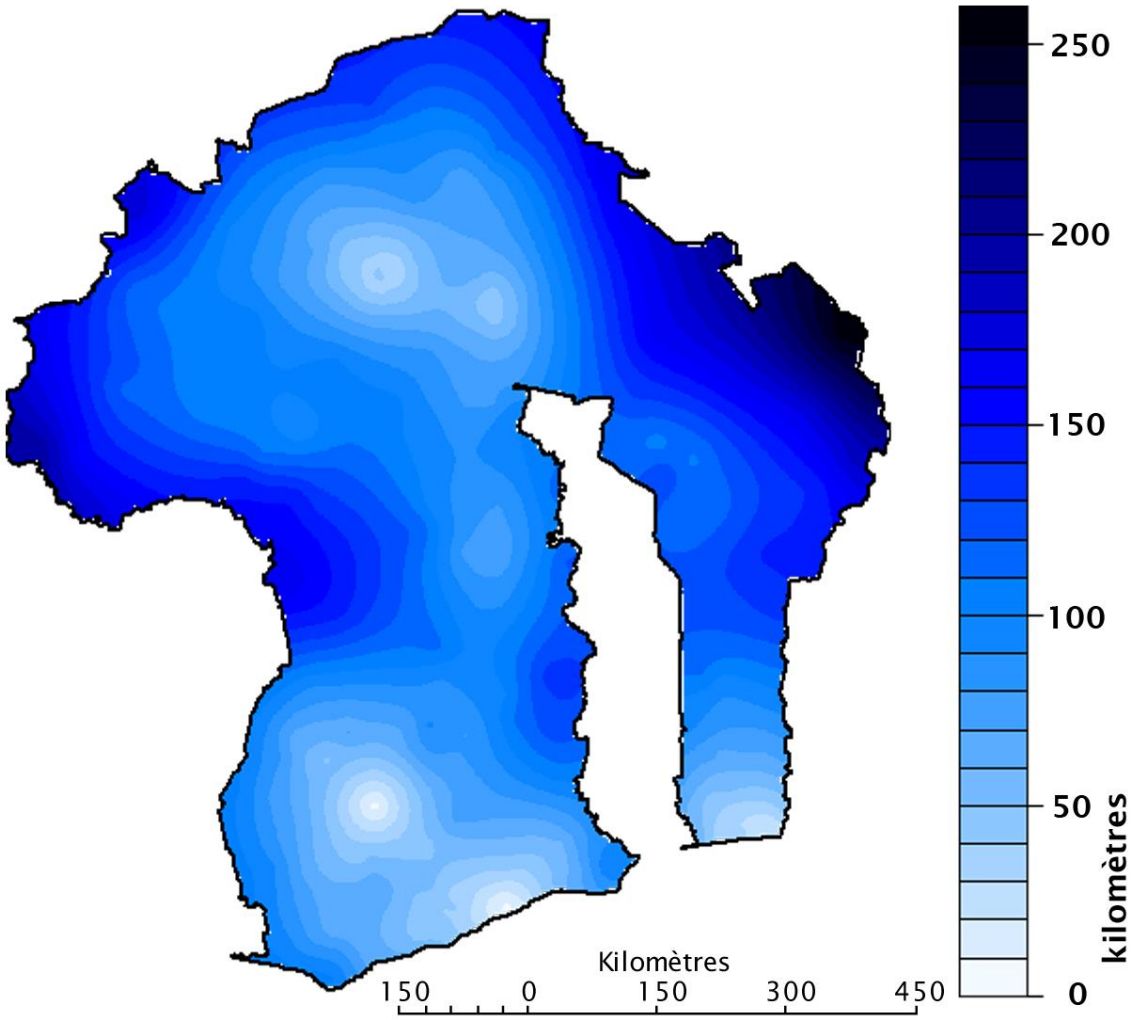



# Average difference by N parameter for 100 DHS



# Ring radius for our example with N=250

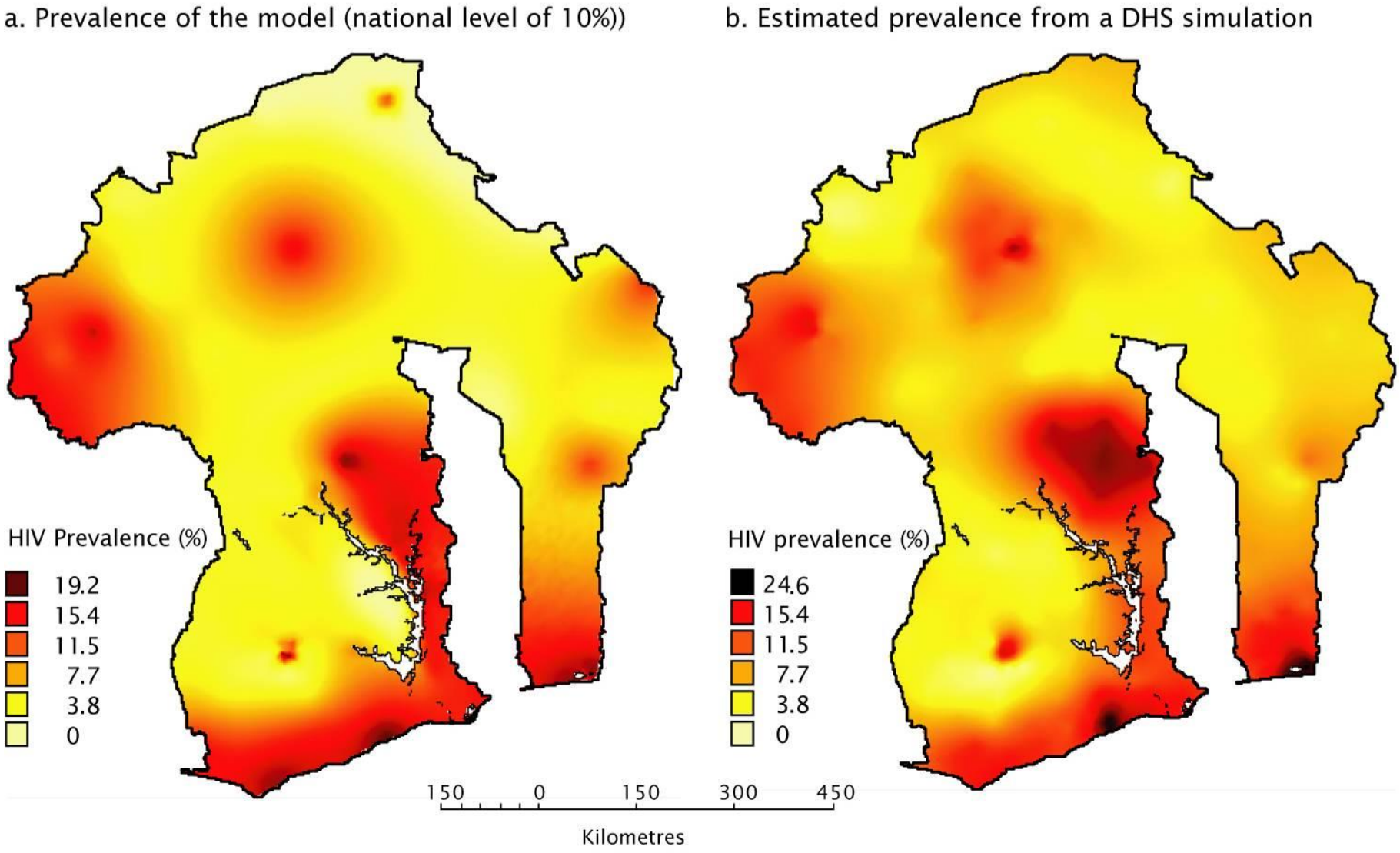



- Area with few observations :
  - High ring radius
  - We add a second parameter, R, corresponding to a maximum ring radius.
  - We decide to use the 90th percentile.
  - Here, R=128 km

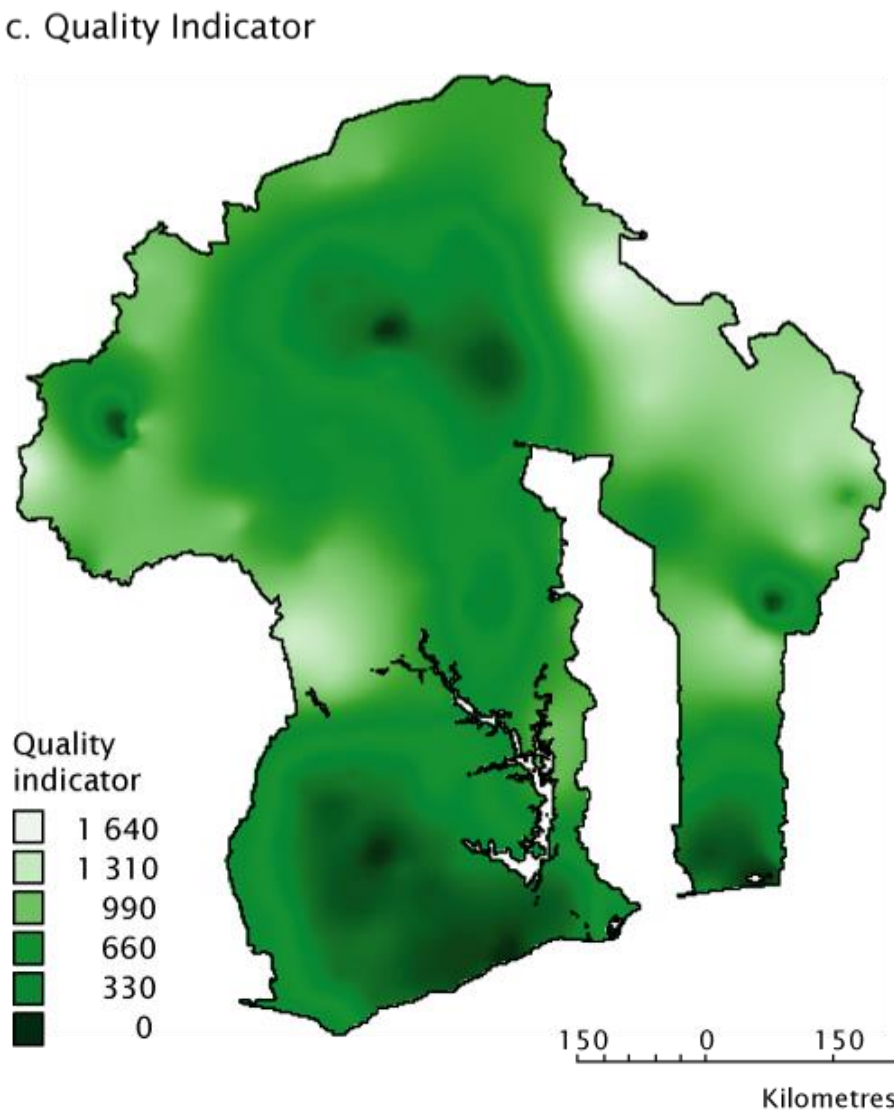
# Taking into account main urban agglomerations

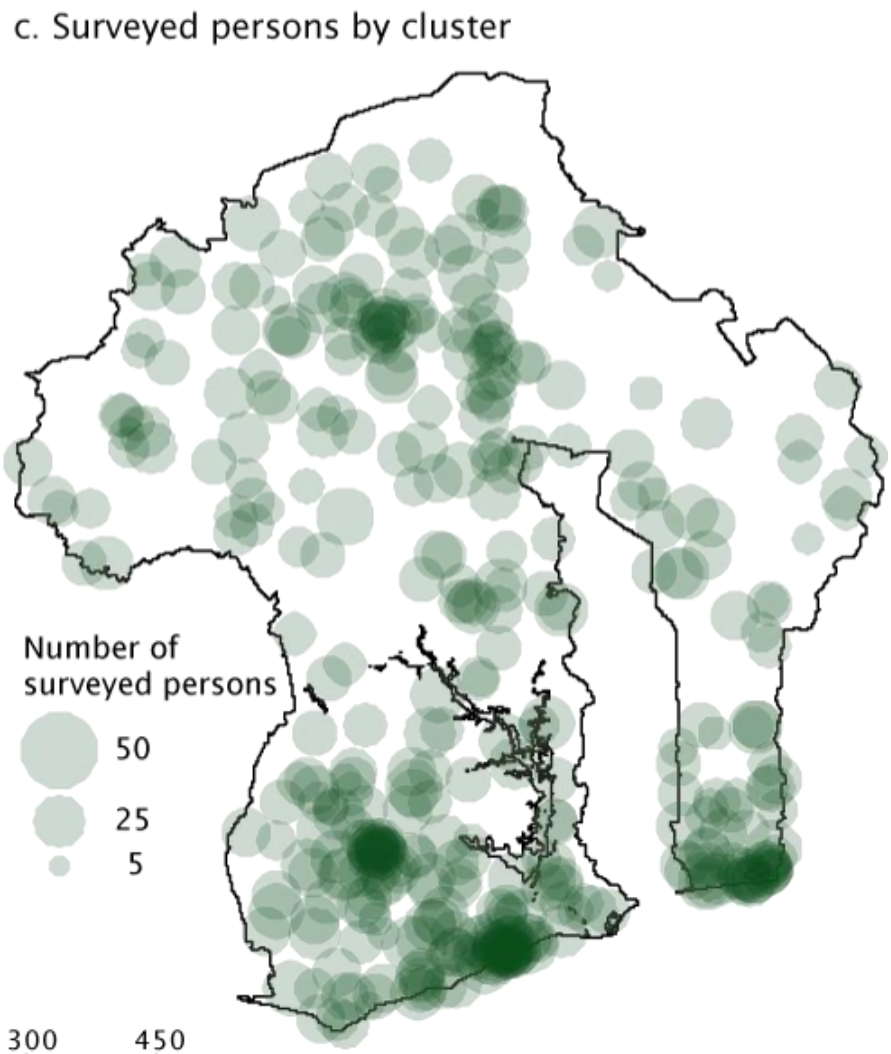
| Agglomeration | Clusters | Observations | Observed prevalence |
|---|---|---|---|
| A | 24 | 489 | 21.2 % |
| I | 18 | 352 | 23.3 % |
| C | 17 | 346 | 15.2 % |
| F | 7 | 151 | 18.0 % |
| E | 3 | 48 | 16.3 % |
| H | 2 | 32 | 10.3 % |

# Initial epidemic of the model and estimated regional trends

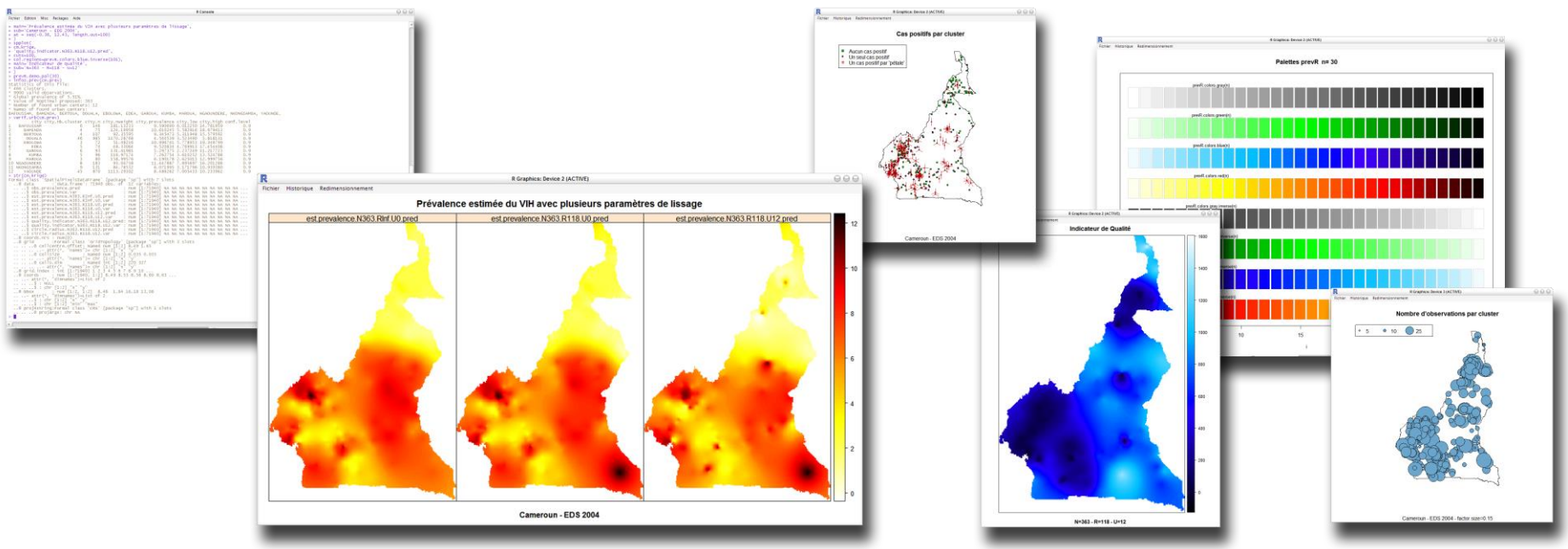


# Complementary maps



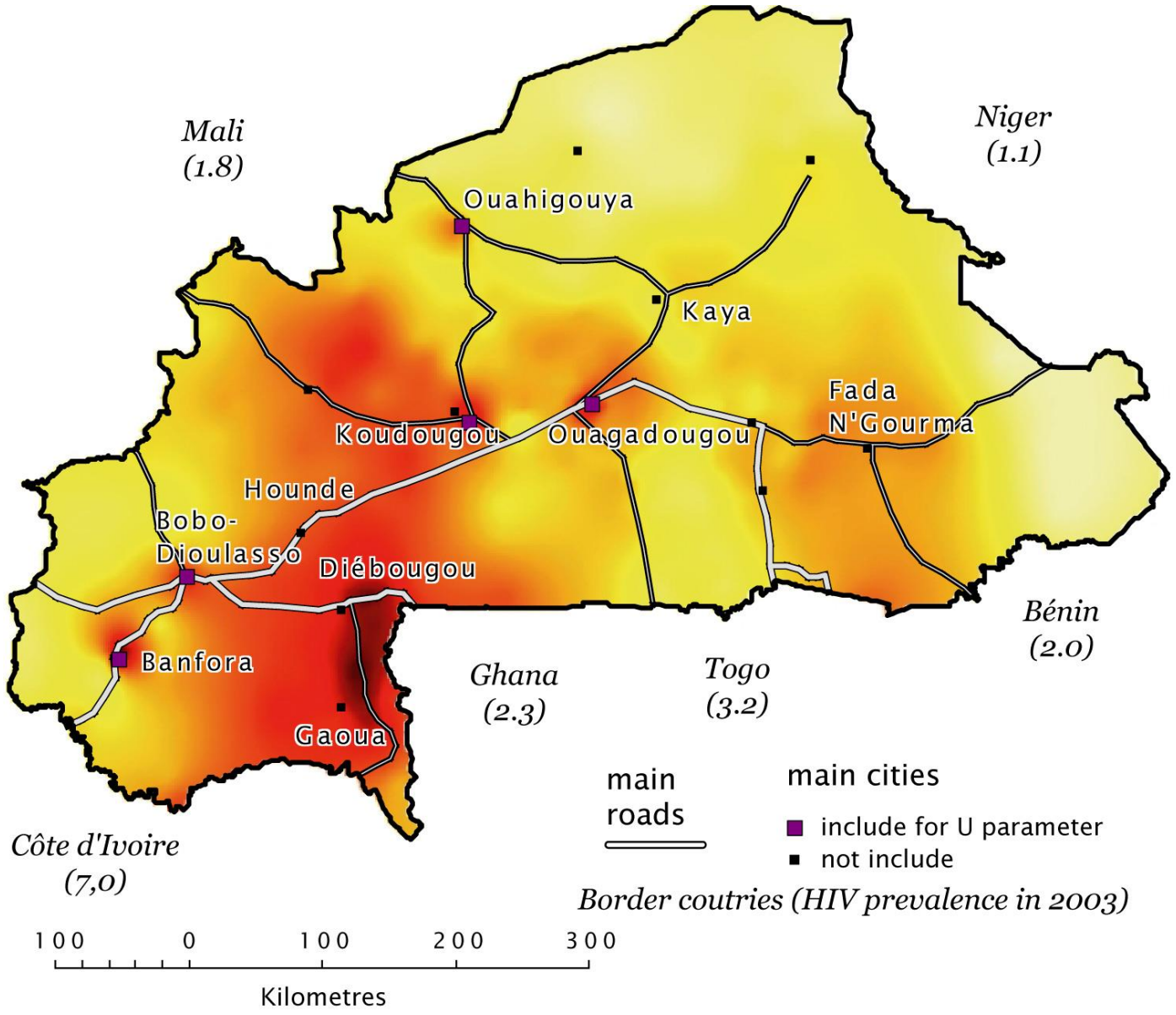



# prevR: doing theses analysis with R




- Package for the free and open-source statistical software R
- Available for free on http://www.ceped.org/prevR
- Bilingual: French and English
- Data import, analysis, export in standard formats
- Can be used with step by step functions

# Application in Burkina Faso: HIV prevalence in 2003



# Discussion

- Radius of same number of observations allows to take into account the heterogeneity of the distribution of the observations.
- Method easy to use with prevR and not specific to HIV prevalence.

- Results are limited:
  - indirect estimation from neighbored clusters
  - descriptive approach and not explicative
  - optimal value of N estimated from a fictive country but applied to real one
- But coherent:
  - with current knowledge of HIV epidemics

- This approach constitutes a first step in a research:
  - Having from DHS a first picture of spatial variations of a phenomenon before doing a specific survey in order to understand and to explain it.

# Thank you

http://www.ceped.org/prevR