

# Optimisation de l'échantillonnage et ciblage des interventions contre la schistosomiase au Tchad : Approche géostatistique sur les données environnementales.

Serge Madjou

Doctorant en science de données en  
démographie à l'Université de Rwanda

Mots-clés : Schistosomiase, Tchad, géostatistique, échantillonnage, données environnementales, santé publique.

## Résumé :

Cette étude examine la schistosomiase, maladie tropicale négligée touchant principalement l'Afrique, dont le Tchad où plus de 2,45 millions de personnes ont besoin de traitement. L'objectif est d'analyser les facteurs environnementaux influençant la distribution de la maladie et de proposer une méthode d'échantillonnage optimale via la modélisation géospatiale, pour une meilleure évaluation des interventions. Une revue de la littérature a identifié des facteurs clés tels que la proximité aux cours d'eau et la végétation aquatique. L'analyse descriptive et le modèle géostatistique gaussien ont mis en évidence une forte corrélation entre ces facteurs environnementaux et la prévalence de la maladie, démontrant l'importance de la proximité aux points d'eau et aux zones agricoles comme foyers de transmission. Sur cette base, l'étude a permis de réduire la taille des échantillons nécessaires de 25% par rapport à un échantillonnage aléatoire simple, tout en conservant la même précision. Elle souligne l'utilité des données environnementales publiques pour optimiser le ciblage des interventions, particulièrement dans des pays à ressources limitées, et recommande l'utilisation de données d'intervention actualisées pour affiner les échantillonnages et les interventions futures.

## Introduction :

La bilharziose, appelé encore schistosomiase est une maladie tropicale négligée qui touche plus de 250 millions de personnes dans le monde, dont plus de 90% se trouvent en Afrique<sup>1</sup>. Ses impacts sur la santé et l'économie sont considérables, causant d'importants coûts publics, une dégradation de la qualité de vie et une baisse de productivité[1].

Au Tchad, elle impacte des millions d'individus, dont plus de 2.45 millions nécessitant des traitements massifs. La distribution de la maladie varie considérablement, avec des zones de forte à faible prévalence, suggérant une analyse profonde pour identifier les principaux facteurs déterminants<sup>2</sup>.

La feuille de route pour éliminer les maladies tropicales négligées [2] recommande de mieux cibler les traitements, suggérant l'utilisation de nouvelles approches pour mieux estimer le besoin. Par ailleurs, face au manque de financement, il est crucial de développer de nouvelles méthodes réduisant la taille de l'échantillon tout en améliorant la précision pour permettre des évaluations fréquentes afin de mieux mesurer l'impact des interventions.

L'objectif de ce cette étude est d'analyser les facteurs environnementaux de la schistosomiase, prédire sa distribution à une petite échelle spatiale et proposer une approche d'échantillonnage optimale pour mesurer l'impact des interventions en utilisant les techniques de modélisation géospatiale.

---

<sup>1</sup> <https://www.who.int/data/gho/data/themes/topics/schistosomiasis>

<sup>2</sup> [https://espen.afro.who.int/system/files/content/resources/CHAD\\_NTD\\_Master\\_Plan\\_2016\\_2020.pdf](https://espen.afro.who.int/system/files/content/resources/CHAD_NTD_Master_Plan_2016_2020.pdf)

## Méthodologie :

Afin d'atteindre cet objectif, nous avons dans un premier temps une revue de la littérature afin d'identifier les principaux facteurs environnementaux de la schistosomiase. Cela servira de bases d'initier la température, l'élévation, la proximité aux cours d'eau, ainsi que la végétation aquatique, la densité de la population, les caractéristiques du sol comme le Ph et la salinité, l'humidité et la précipitation comme facteurs importants.

Nous par la suite fait une analyse descriptive pour cerner la distribution des variables ainsi que l'existence éventuelle de corrélation entre les facteurs environnement entre elles et avec la prévalence de la schistosomiase. Enfin la modélisation géospatiale multiniveaux a été effectué pour identifier les facteurs environnementaux qui expliquent la distribution de la prévalence et faire la prédiction de la prévalence sur une échelle spatiale le plus petit possible.

Sur la base de la prédiction, nous avons constitué des strates qui permettent un échantillonnage optimal.

Plusieurs sources de données ouvertes suivantes seront utilisées :

- [Enquête de cartographie](#) sur la schistosomiase réalisée en 2015
- [WorldPop](#) pour des estimations de populations dégagés aux unités spatiales les plus petites ainsi que des informations environnementales importants
- [WorldClim](#) qui fournit des données la température
- [IHME](#) qui fournit des données sur les l'environnement bâti, notamment l'accès à l'eau potable ainsi qu'à des assainissements appropriés
- [Google Earth Engine](#) qui fournit des données allant de du climat aux données géophysique
- [ISRIC](#) qui fournit des données sur les caractéristiques du sol
- [Enquête démographique de santé](#) utilisé pour extraire certaines variables sur les caractéristiques des ménages à l'échelle de province

## Résultats :

Les résultats de l'analyse descriptive ont mis en lumière l'existence d'une forte corrélation entre certaines variables (voir Annexe 1). Cela a permis de sélectionner, parmi les variables fortement corrélées, quelques-unes pour la modélisation. L'analyse du variogramme empirique a révélé quelques points en dehors de la bande de confiance, suggérant l'existence d'une autocorrélation spatiale.

Le modèle géostatistique gaussien a réussi à démontrer l'effet significatif de facteurs environnementaux tels que la proximité aux points d'eau et aux zones agricoles, qui constituent un foyer de transmission de la maladie, ainsi que la présence de végétation herbacée et d'aires protégées, lesquelles offrent un habitat favorable aux hôtes vecteurs.

La prédiction de la prévalence à une échelle géographique plus restreinte a permis de définir des strates qui réduisent la taille des échantillons de 25% par rapport à un échantillonnage aléatoire simple, tout en conservant la même précision.

## Discussion :

Cette étude a mis en évidence l'importance d'utiliser des données environnementales publiquement disponibles et des techniques environnementales pour mieux cibler les interventions et optimiser l'échantillonnage pour les évaluations, notamment dans des pays à ressources limitées comme le Tchad. Elle renforce les résultats d'études menées dans d'autres contextes qui aboutissent aux mêmes conclusions [3].

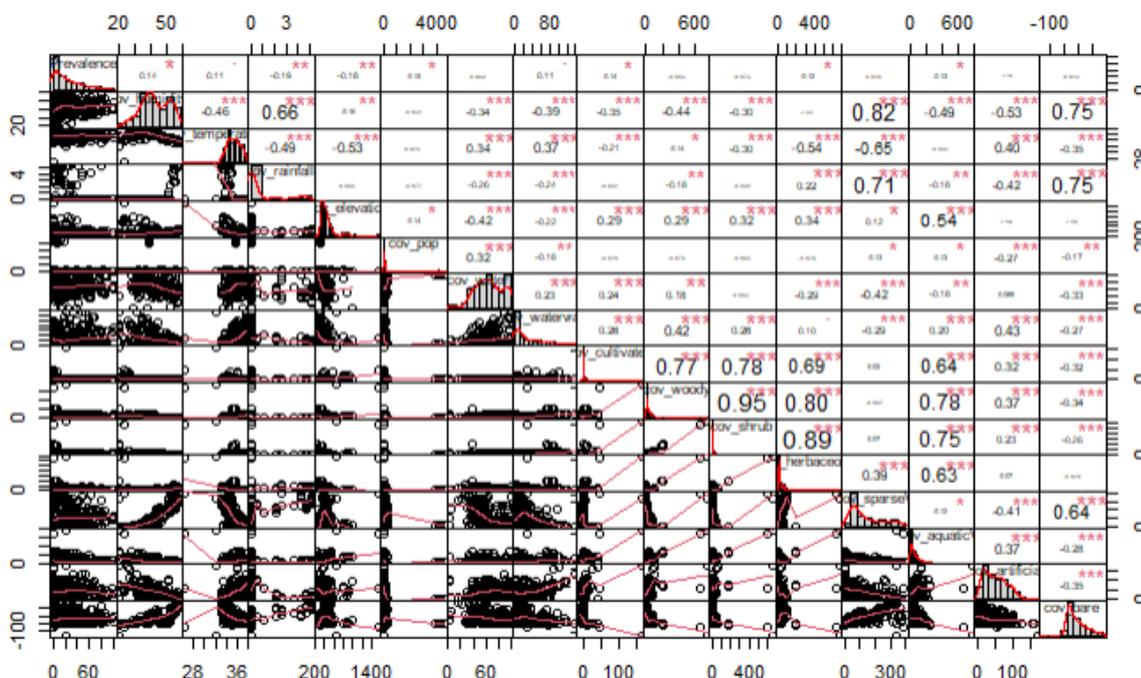
Toutefois, le fait de définir un échantillonnage basé uniquement sur les prédictions des enquêtes de référence peut ne pas être suffisant pour obtenir un échantillon représentatif de la situation actuelle. Il est

donc important de procéder à une projection des prévalences en utilisant les données des interventions afin d'obtenir des bases de données reflétant la situation au moment de l'enquête avant la sélection de l'échantillon. Cela peut permettre de réduire considérablement la taille de l'échantillon nécessaire et de mieux définir les futures interventions sur la base d'informations plus proches de la situation actuelle [4].

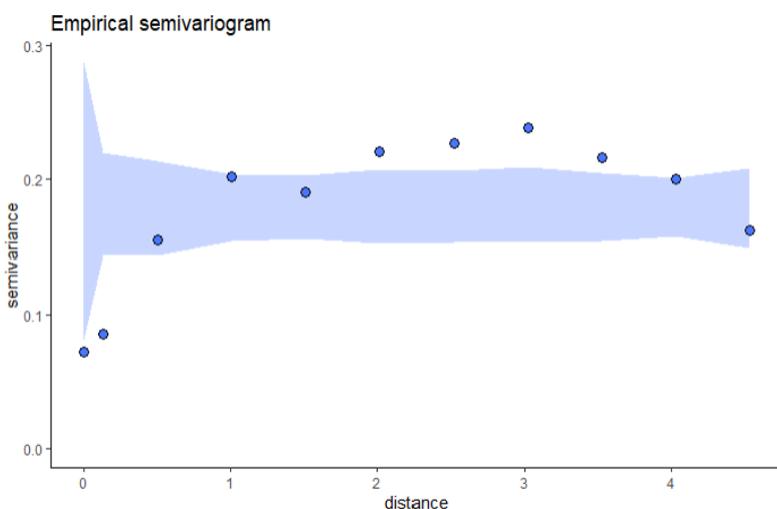
## Conclusion

L'utilisation de l'approche géostatistique pour améliorer le ciblage des individus nécessitant une intervention s'avère prometteuse. Elle permet une meilleure compréhension de la répartition spatiale et, par conséquent, optimise l'allocation des ressources pour lutter contre cette maladie. Cette méthode est également porteuse d'espoir pour la définition d'un échantillon optimal, exigeant moins de ressources.

## Autocorrélation des variables



## Variogramme empirique



## Modèle géostatistique initial

```
Geostatistical linear model
Call:
linear.model.MLE(formula = Prevalence ~ 1, coords = ~longitude +
  latitude, data = data_Chad_any, kappa = 0.5, start.cov.pars = c(1.3,
  0.2))

            Estimate StdErr z.value  p.value
(Intercept) 22.7687  5.2441  4.3418 1.413e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-likelihood: -768.1208

Covariance parameters Matern function (kappa=0.5)
            Estimate StdErr
log(sigma^2)  6.2491 0.2514
log(phi)      -0.3823 0.3826
log(tau^2)    5.4462 0.6332
```

## Modèle GLM simple

```
Call:
glm(formula = cbind(Examined, Examined - Positive) ~ 1 + cov_temperature +
  cov_elevation + cov_water + cov_waterway + cov_cultivated +
  cov_woody + cov_shrub + cov_herbaceous + cov_sparsev + cov_aquaticv +
  cov_artificial + cov_bare + cov_herbaceous, family = binomial(),
  data = data_Chad_any)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.6015459  0.9971242   1.606 0.108238
cov_temperature -0.0396236  0.0266421  -1.487 0.136946
cov_elevation  -0.0003746  0.0002315  -1.618 0.105681
cov_water       0.0024390  0.0010104   2.414 0.015784 *
cov_waterway   0.0023924  0.0007353   3.254 0.001140 **
cov_cultivated -0.0473173  0.0086515  -5.469 4.52e-08 ***
cov_woody      -0.0001443  0.0012910  -0.112 0.910999
cov_shrub      0.0006228  0.0016399   0.380 0.704125
cov_herbaceous -0.0034028  0.0008774  -3.878 0.000105 ***
cov_sparsev    0.0007618  0.0002914   2.614 0.008946 **
cov_aquaticv   0.0004748  0.0004149   1.144 0.252511
cov_artificial 0.0007654  0.0006723   1.138 0.254939
cov_bare      -0.0004131  0.0003169  -1.304 0.192372
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 758.03 on 189 degrees of freedom
Residual deviance: 682.82 on 177 degrees of freedom
(24 observations effacées parce que manquantes)
AIC: 1609.1

Number of Fisher Scoring iterations: 4
```

## Reference bibliographique

- [1] G. L. Nascimento et M. R. F. de Oliveira, « Severe forms of schistosomiasis mansoni: epidemiologic and economic impact in Brazil, 2010 », *Trans. R. Soc. Trop. Med. Hyg.*, vol. 108, n° 1, p. 29-36, 2014.
- [2] World Health Organization, « Ending the neglect to attain the Sustainable Development Goals: a road map for neglected tropical diseases 2021–2030 », 2020.
- [3] O. Johnson *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.*, vol. 72, n° Supplement\_3, p. S172-S179, juin 2021, doi: 10.1093/cid/ciab192.
- [4] C. Fronterrière, O. Johnson, E. Giorgi, A. Montresor, et P. J. Diggle, « Survey Design and Geospatial Analysis using Data on Baseline Prevalence, Environmental Risk-Factors and Treatment History Drastically Reduces the Cost of STH Impact Surveys », *Public and Global Health*, preprint, janv. 2023. doi: 10.1101/2023.01.03.23284146.