# Delineating and Analyzing Locality-Level Determinants of Cholera, Haiti

Karolina Griffiths, Kenny Moise, Martine Piarroux, Jean Gaudart, Samuel Beaulieu, Greg Bulit, Jean-Petit Marseille, Paul Menahel Jasmin, Paul Christian Namphy, Jean-Hugues Henrys, Renaud Piarroux, Stanislas Rebaudet

Centre Department, Haiti, was the origin of a major cholera epidemic during 2010–2019. Although no fine-scale spatial delineation is officially available, we aimed to analyze determinants of cholera at the local level and identify priority localities in need of interventions. After estimating the likely boundaries of 1,730 localities by using Voronoi polygons, we mapped 5,322 suspected cholera cases reported during January 2015–September 2016 by locality alongside environmental and socioeconomic variables. A hierarchical clustering on principal components highlighted 2 classes with high cholera risk: localities close to rivers and unimproved water sources (standardized incidence ratio 1.71, 95% CI 1.02–2.87; p = 0.04) and urban localities with markets (standardized incidence ratio 1.69, 95% CI 1.25–2.29; p = 0.0006). Our analyses helped identify and characterize areas where efforts should be focused to reduce vulnerability to cholera and other waterborne diseases; these methods could be used in other contexts.

Haiti experienced a large and lasting cholera epidemic that began in October 2010. The epidemic originated in Centre Department in a hamlet hosting a camp of United Nations peacekeepers, then contaminated the Artibonite River and coastal plain (*1*,*2*). During 2010–2019, >820,000 suspected cholera cases and 9,792 deaths were reported (*3*). However, since February 2019 no suspected deaths have occurred, and no positive test results have been reported from >1,000 samples collected from patients with acute watery diarrhea.

Fecal contamination of water, food, or hands by toxigenic strains of the bacterium *Vibrio cholerae* O1 is the common mode of cholera transmission. The National Plan for the Elimination of Cholera, launched by the Ministry of Public Health and Population (Ministère de la Santé Publique et de la Population [MSPP]) in Haiti and co-authored by Haiti's National Drinking Water and Sanitation Directorate (La Direction Nationale de l'Eau Potable et de l'Assainissement [DINEPA]), is now in its third phase, focusing on reinforcing access to clean drinking water and sanitation (*4*,*5*). Reducing population vulnerability to cholera and other waterborne diseases at a community level is vital. However, with the constraints of limited resources, the most pertinent at-risk geographic zones must be prioritized for sustainable water sanitation and hygiene (WaSH) interventions.

Previous studies have performed spatial analyses at communal levels, but data at a finer scale are limited. High-resolution mapping is needed to understand the heterogeneous transmission patterns and to adapt specific intervention strategies at the community level (*6*–*9*). Although a previous study by Allan et al. (*6*) provided maps of the origin of cholera patients at a subcommunal level for the neighboring Artibonite Department, these maps are still at a relatively large scale, demonstrating communal sections hosting 10,000–30,000 persons. That study reported neighboring sections within the same commune with clear differences in relative risk for cholera in a mosaic pattern, highlighting the need for further locality-level investigation to guide WaSH

Author affiliations: Aix Marseille University, Marseille, France (K. Griffiths, J. Gaudart, S. Rebaudet); Université Quisqueya, Port-au-Prince, Haiti (K. Moise, J.-H. Henrys); Centre d'épidémiologie et de santé Publique des armées, Marseille (M. Piarroux); UNICEF, Kinshasa, Democratic Republic of the Congo (S. Beaulieu); UNICEF, New York, New York, USA (G. Built); Direction Nationale de l'Eau Potable et de l'Assainissement, Hinche, Haiti (J.-P. Marseille); Ministère de la Santé Publique et de la Population, Hinche (P.M. Jasmin); Direction Nationale de l'Eau Potable et de l'Assainissement, Petion-Ville, Haiti (P.C. Namphy); Sorbonne Université, Paris, France (R. Piarroux)

DOI: https://doi.org/10.3201/eid2701.191787

interventions. However, the mapping of rural localities in Haiti is sparse and no spatial delineation has been established, presenting a challenge in identifying microspots with recurrent cases and analyzing associated factors (*6*). A previous study collected fine-scale spatial data in Haiti by using spatial videos for cholera investigations; however, it was limited to intraurban areas and did not include case data (*10*).

Previous studies highlight the challenges in collecting fine-scale spatial data in Haiti, with limited official information on informal settlements (*10*–*12*). Our objective was to analyze the spatial determinants of cholera and to identify the priority localities in need of prevention interventions in the Centre Department in Haiti. We chose Centre Department because it was at the origin of the epidemic and had a high incidence of cholera that persisted for several years.

## Methods

### Study Design and Setting

We conducted an observational, ecologic study at the locality (hamlet) level in the Centre Department, Haiti (Figure 1), which covers an area of 3,487 km$^2$, is >80% rural, and in 2015 had a population of ≈746,236 (*13*). Centre is administratively subdivided into 12 civil townships, known as communes, each of which has an urban area (Figure 2, panel A). Each commune is further subdivided into communal sections, the smallest official administrative unit, which include several hundred localities. Localities are groups of residences and are the smallest spatial unit culturally used to define the place of residence for rural populations. However, the number, spelling, and delimitations of rural localities are not officially established, and only estimated geolocalization and populations are available (*14*).

The largest towns in Centre are Hinche and Mirebalais. Centre has a mean altitude of 447 m (range 84–1,820 m) and is situated northeast of the Montagnes Noires and Montagnes de Trou d'Eau mountain ranges and south of the Massif du Nord. Hundreds of rivers and streams provide natural water sources in the department.

### Materials

We obtained the deidentified case list of suspected cholera cases in Centre during January 2015–September 2016 from the MSPP. The case list provided the

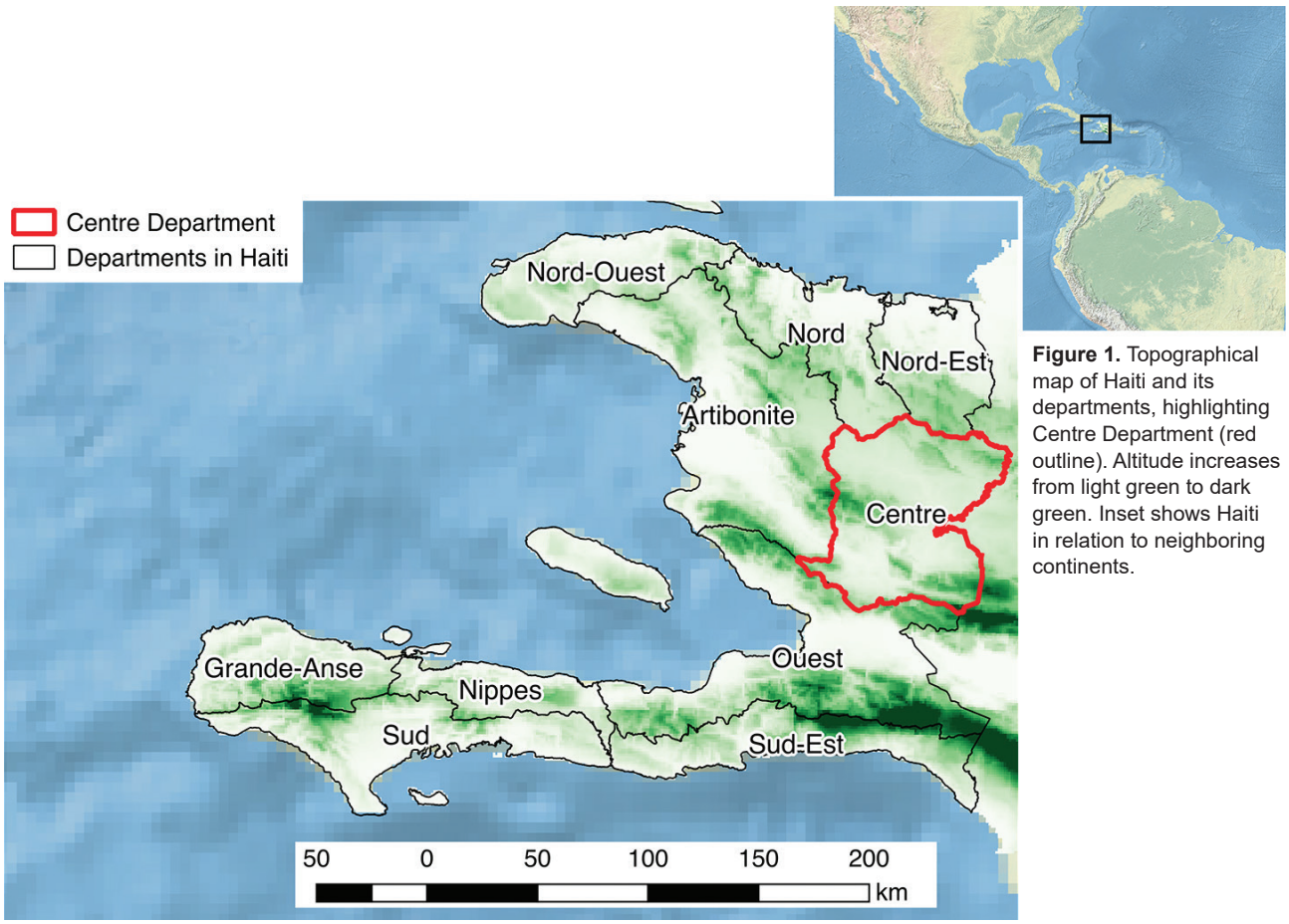



**Figure 1.** Topographical map of Haiti and its departments, highlighting Centre Department (red outline). Altitude increases from light green to dark green. Inset shows Haiti in relation to neighboring continents.

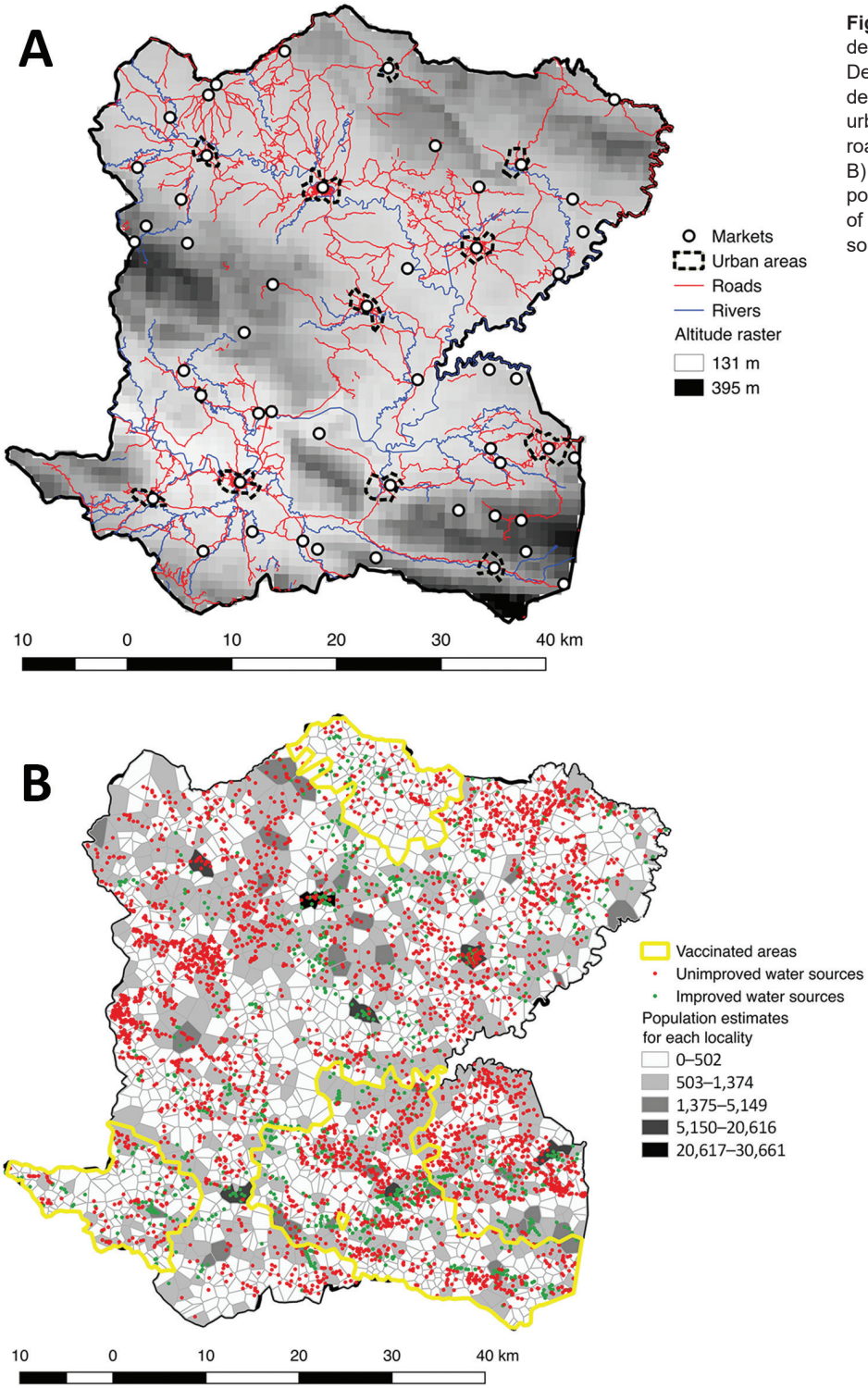



**Figure 2.** Geographic and demographic details of Centre Department, Haiti. A) Outline of the department's 12 communes, principal urban areas, altitude raster, rivers, roads, and position of markets. B) Polygon outline and estimated population for each locality, positions of unimproved and improved water sources, and vaccinated areas.

locality of residence for each patient and was collected by the Health Departmental Directorate of the MSPP to guide case-area targeted interventions conducted by rapid response teams (*15*). All patient identifiers were removed previously.

The mapping of rural localities in Haiti is incomplete (*6*). We collected point global positioning system (GPS) coordinates for each locality from field visits by mobile response teams, authors' field visits, or satellite photos and from geographic repositories,

including Haitian Institute of Statistics and Computer Science (http://ihsi.ht/publication_cd_atlas.htm), Index Mundi (https://www.indexmundi.com), OpenStreetMap (https://www.openstreetmap.org), Google Earth (https://earth.google.com), and Google Maps (https://www.google.fr/maps). We verified GPS coordinates against official data from Centre National d'Information Geospatiale (CNIGS), when available. Further data validation was performed by using computerized participatory mapping techniques to integrate spatial information from local community stakeholders in each commune, either individually or in small groups. These stakeholders included local healthcare professionals from mobile response teams, WaSH technicians, and local government personnel. We checked and unified locality names and used satellite imagery–assisted visual mapping techniques to validate the estimated GPS coordinates. This process ensured coherence with local knowledge of the terrain and identified inconsistencies, which was vital for rural areas without road access.

As part of a collaborative project financed by World Vision (https://www.worldvision.org), DINEPA and the nongovernmental partner organization Haiti Outreach (https://www.haitioutreach.org) collected an inventory of water sources and provided GPS coordinates by visiting each water point and validating data with local partners (*16*) (Figure 2, panel B). Water sources were classified as improved or unimproved according to the Joint Monitoring Program for Water Supply Sanitation and Hygiene (*17*,*18*).

We obtained the location of rivers, roads, and altitude for Centre from CNIGS (Figure 2, panel A). We obtained history of previous oral cholera vaccination campaigns by communal section from MSPP (Figure 2, panel B). We geolocated markets during field investigations in Centre.

We calculated Voronoi polygons from estimated point coordinates for each locality. These polygons, or proximity diagrams, consist of all locations closer to the locality point coordinate than any other point coordinate, providing estimated boundaries. We used Voronoi polygons as the basis for all variables to maintain spatial unit consistency and to reduce aggregation bias. We estimated the number of houses per polygon by using a satellite-based house detection shapefile from CNIGS and completed maps by using Google Earth and OpenStreetMap. We multiplied the number of houses by the mean number of household members in the area to estimate the population and incidence rates in each locality (*13*). The distances from each house to an unimproved water source, improved water source, river, and road were calculated and averaged for each polygon. The mean altitude was calculated for each polygon by using raster terrain analysis and rounded to the nearest meter.

### Statistical Analysis

We assessed variables for normal distribution by using histograms, quantile-quantile plots, and Shapiro-Wilk test. For continuous nonnormally distributed variables, we calculated the median and interquartile range (IQR); for categorical variables, we calculated counts and percentages. We categorized continuous variables, including altitude and distances to a road, to an improved water source, to an unimproved water source, and to a river, by using information from quartiles and histograms to form 4 classes. We performed univariate nonparametric statistical tests by using Kruskal-Wallis rank test and Spearman correlation coefficients on cholera incidence and environmental variables. We performed spatial analysis for hotspot detection by using SatScan (https://www.satscan.org).

We performed multistep nonsupervised analysis to classify the localities according to their environmental and spatial characteristics. Hierarchical clustering on the principal components of a multiple correspondence analysis (MCA) was detailed previously (*19*,*20*); we used these data to classify neighborhoods in towns in Haiti (*12*). The first step is an MCA, which is an exploratory method that considers the relationship between variables and reduces complex datasets into fewer dimensions (*21*). We performed MCA by using the original categorical variables and the categorized continuous variables. Active variables included the presence of a market, urban or rural location, vaccination status, and area averaged: altitude, distance to a road, distance to an improved water source, distance to an unimproved water source, and distance to a river. We retained quantitative information only as supplementary variables and did not use these in the determination of the principal components. To reduce basal noise and ensure a more stable classification, we retained the principal components that summarized 95% of the data. We performed hierarchical ascendant classification on the first 16 principal components' coordinates, which provided classes independent of the number of cholera cases. Then, we compared these classes to cholera cases in a general additive model (GAM) with quasi-Poisson distribution. For spatial autocorrelation, we performed Moran I tests on the number of cases and the GAM residuals. To model spatial dependence, we tested a trend-surface GAM, fitting the geographic location by

using 2 dimensional splines on latitude and longitude coordinates, as previously demonstrated (*22*–*24*). We accounted for the increasing population by using an offset of the log population and estimating standardized incidence ratios (SIRs) for each class. We considered p<0.05 statistically significant.

We used QGIS version 2.14.3 (QGIS Development Team, http://qgis.osgeo.org) as a geographic information system (GIS) for mapping. We performed all statistical analyses by using R version 3.3.0 (R Foundation for Statistical Computing, https://www.r-project.org). We used the FactoMineR package in R for classification analysis (*19*) and mgcv for GAMs, with generalized cross validation criteria for smoothing parameter estimations and the gam.check function to verify residual plots (*22*,*25*).

All data remained anonymous with no patient identifiers, in accordance with national and international ethics guidance (*26*). Ethics approval was obtained from the National Bioethics Committee in Haiti, MSPP (reference no. 1516-73).

## Results

A total of 5,322 suspected cholera cases were recorded in Centre during January 2015–September 2016, and 1,730 localities were identified and mapped. The median locality size was 1.77 km$^2$ (IQR 1.15–2.58 km$^2$) and median population 300 (IQR 153–530) persons. Among 1,730 localities, 689 (40%) had ≥1 suspected cholera case (Table 1). The median incidence ratio for all localities was 0 (IQR 0–61.9) per 10,000 persons. Incidence ranged from 0–6,050.9/10,000 persons; 25 localities had an incidence >1,000/10,000 persons (Figure 3). In univariate analysis, the categorical variables statistically significant for incidence were altitude, distance to an unimproved water source, distance to an improved water source, distance to road, distance to a river, presence of market, rural or urban location (p<0.0001), and cholera vaccination (p<0.0001 for all variables) (Table 1; Appendix, https://wwwnc.cdc.gov/EID/article/27/1/19-1787-App1.pdf).

### Hierarchical Clustering on Principal Components

Hierarchical clustering on MCA provided 4 different classes of localities (Table 2; Appendix Figure 1), demonstrating different environmental and spatial characteristics, which we mapped (Figures 4, 5). The categorical variables that best characterize the portioning into 4 classes were distance to a river, presence of a market, altitude, distance to an unimproved water source, urban or rural location, and distance to a road (p<0.0001 for all).

Class 1 (n = 621) was most strongly associated with being far from a river, at high altitude, and far from a road (Table 2). Class 1 localities also were remote and farther than average from both unimproved water sources (median 576 m) and improved water sources (median 1,213 m), had no markets, were rural, and had more vaccinated persons.

Class 2 (n = 941) was associated with being a medium distance to the river and at the lowest altitudes (Table 2). Class 2 localities also were associated with being a medium distance from a road, having no markets, being rural, and having more unvaccinated persons. These rural intermediary localities were slightly closer than the average distance to both improved water sources (median 712 m) and unimproved water sources (median 439 m) and lower than average distances to a road (median 390 m).

Class 3 (n = 61) was most notably associated with being closer to rivers (<200 m) and unimproved water sources (<150 m). Class 3 localities also had lower than average distances to a road (median 210 m).

Class 4 (n = 77) was most strongly associated with markets and urban localities. These localities were closer than average to unimproved water sources (median 347 m) and improved (median 338 m) sources and a lower than average distance to a road (median 135 m).

**Table 1.** Summary descriptive statistics for each locality and variables used for classification analysis used in delineating and analyzing locality-level determinants of cholera, Haiti

| Locality information | All localities, n = 1,730 | Association with incidence, p value |
|---|---|---|
| Median no. suspected cholera cases (IQR) | 0 (0–2) | – |
| Median no. houses (IQR) | 65 (33–115) | – |
| Median estimated population (IQR) | 300 (152–530) | – |
| Median estimated incidence/10,000 (IQR) | 0 (0–61.85) | – |
| Median altitude, m (IQR)*† | 360 (264–620) | <0.0001 |
| Median distance to nearest improved water source, m (IQR)*† | 842 (464–1,476) | <0.0001 |
| Median distance to nearest unimproved water source, m (IQR) *† | 451 (307–702) | <0.0001 |
| Median distance to nearest river, m (IQR)*† | 1,263 (554–2,706) | <0.0001 |
| Median distance to nearest road, m (IQR) | 638 (185–1,493) | <0.0001 |
| No. markets (%)* | 50 (2.9) | <0.0001 |
| No. rural locations (%)* | 1,686 (97.5) | <0.0001 |
| No. cholera vaccinations administered (%)* | 556 (32.1) | <0.0001 |

*Active variables used in classification analysis, statistically significant in nonparametric univariate analysis (p<0.05).
†Values transformed to categorical variables by using quartiles to form 4 classifications for analysis.

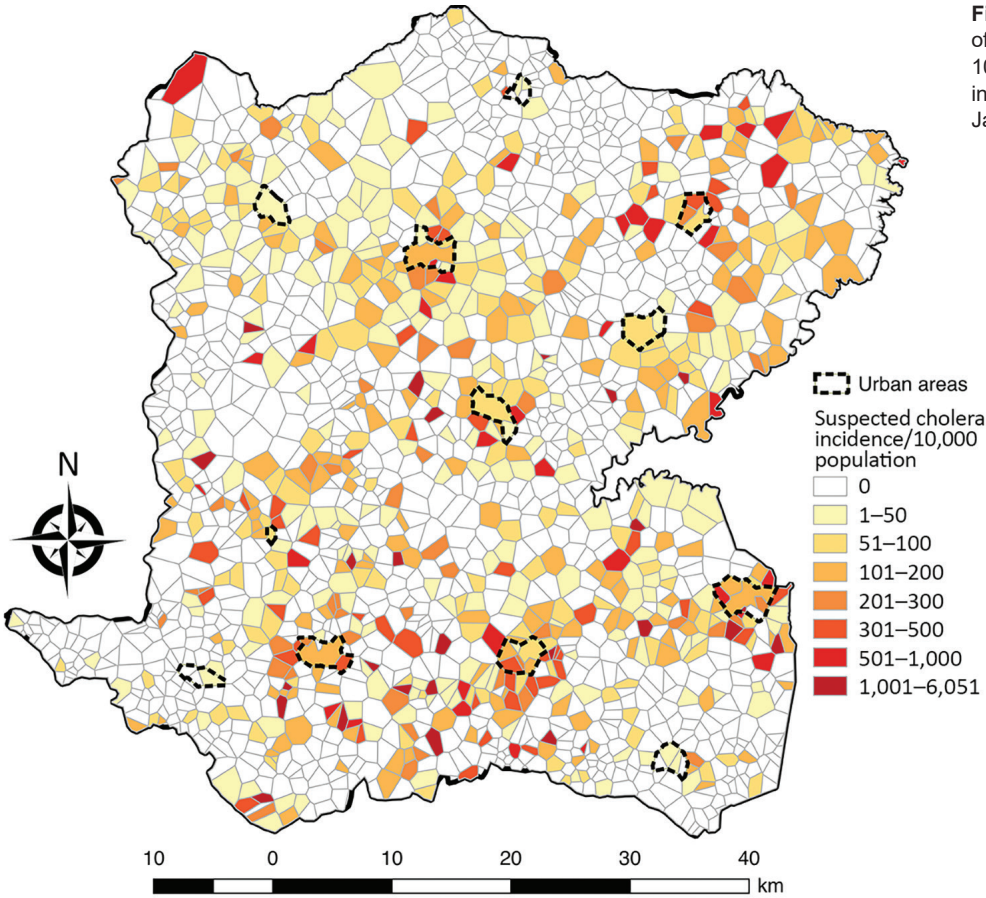



**Figure 3.** Estimated incidence of suspected cholera cases per 10,000 persons for each locality in Centre Department, Haiti, January 2015–September 2016.

Taking class 1 as the reference and taking geographic coordinates and population into consideration, we used a quasi-Poisson GAM to compare the classes and estimate SIRs (Table 2; Appendix). The model confirmed the statistically significantly higher cholera incidence in class 3 (SIR 1.71, 95% CI 1.02–2.87; p = 0.0425) and class 4 (SIR 1.69, 95% CI 1.25–2.29; p = 0.0006). We found class 2 had a slightly increased risk for cholera compared with class 1, with an SIR of 1.28 (95% CI 0.96–1.71), although this difference was not statistically significant at the 5% level (p = 0.0896) (Figure 6).

**Table 2.** Characteristics for each classification identified for analysis of locality-level determinants of cholera, Haiti*

| Information for each locality | Class 1, n = 651 | Class 2, n = 941 | Class 3, n = 61 | Class 4, n = 77 |
|---|---|---|---|---|
| Median no. suspected cholera cases (IQR) | 0 (0–1) | 0 (0–3) | 0 (0–3) | 6 (1–17) |
| Median no. houses (IQR) | 46 (23–86) | 79 (42–127) | 52 (26–79) | 146 (75–315) |
| Median estimated population (IQR) | 212 (106–396) | 364 (194–585) | 240 (120–364) | 673 (346–1,452) |
| Median estimated incidence, cases/10,000 population (IQR) | 0 (0–15.80) | 0 (0–72.31) | 0 (0–111) | 83.0 (8.6–197.2) |
| Median altitude, m (IQR)*† | 665 (520–816) | 287 (232–352) | 284 (227–414) | 339 (265–461) |
| Median distance to nearest improved water source, m (IQR)*† | 1,213 (665–1,819) | 712 (429–1,203) | 721 (263–1,357) | 338 (233–585) |
| Median distance to nearest unimproved water source, m (IQR)*† | 576 (331–1,080) | 439 (317–593) | 163 (136–183) | 347 (274–533) |
| Median distance to nearest river, m (IQR)†‡ | 2,932 (2,047–4,434) | 805 (447–1,346) | 177 (156–193) | 578 (353–1,407) |
| Median distance to nearest road, m (IQR)†‡ | 1,544 (882–2,462) | 390 (133–816) | 210 (70–594) | 135 (29–279) |
| No. (%) markets† | 0 | 0 | 2 (3) | 48 (62) |
| No. (%) rural locations† | 651 (100) | 941 (100) | 57 (93) | 40 (52) |
| Standardized incidence ratio (95% CI), p value | Referent | 1.28 (0.96–1.71), 0.0896 | 1.71 (1.02–2.87), <0.05 | 1.69 (1.25–2.29), <0.01 |

*Class 1 was the reference class, situated at high altitude, and furthest from river and sources. Class 2 was not much more at risk, characterized by a medium distance to roads. Classes 3 and 4 were much more at-risk. Class 3 localities were closer to a river or an unimproved water source and class 4 localities were more often urban with a market.
†Active variables used in classification analysis.
‡Transformed to categorical variables for analysis.

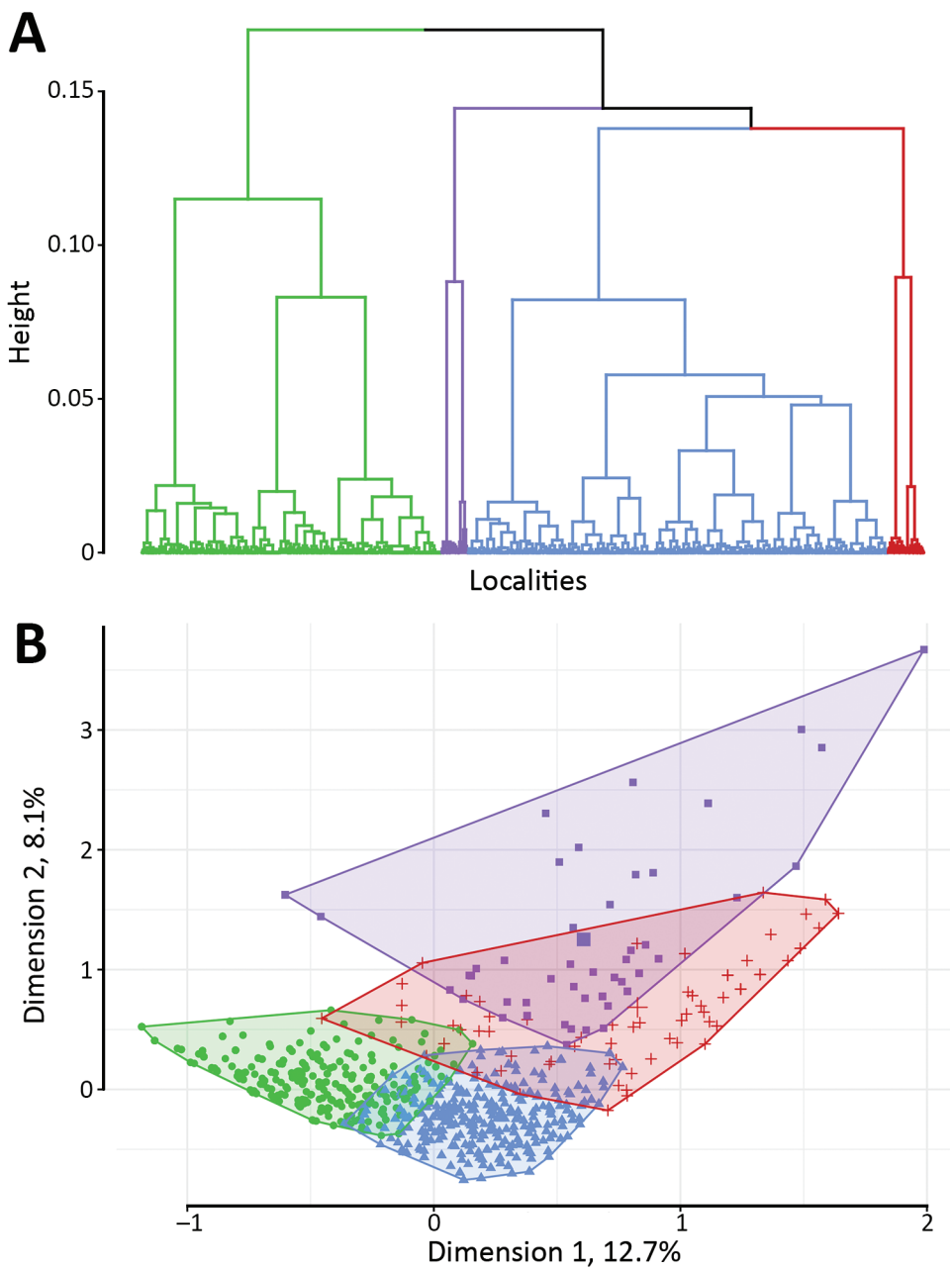



**Figure 4.** Classification analysis of localities regarding environmental variables based on hierarchical clustering on principal components of multiple correspondence analysis, Centre Department, Haiti. A) Cluster dendrogram demonstrating the division of localities into 4 classes: green, class 1; blue, class 2; purple, class 3; red, class 4. Height indicates the order at which the clusters were joined. B) Factor map demonstrating the 4 classes on the first 2 dimensions of the multiple correspondence analysis with the following variables: altitude, distance to an unimproved water source, distance to an improved water source, distance to road, distance to a river, presence of market, rural or urban, and cholera vaccination. The x and y axes represent the first 2 dimensions of the multiple correspondence analysis; the percentage of the total dataset inertia is represented by each dimension. Each point is a locality, with the shaded areas representing the 4 classes, as in panel A.

## Discussion

Our analysis identified 2 main classes of localities at higher risk for cholera: localities close to both rivers (<200 m) and unimproved water sources (<150 m) and urban localities with markets. These criteria are simple ways of identifying high-risk localities for source-based interventions within a large department like Centre. We identified total of 138 higher risk localities, accounting for 8% of all localities.

Reducing population vulnerability to diarrheal illnesses in a low-resource setting has complex challenges, which was confounded in Centre by the large geographic area and difficult terrain. Prioritizing strategies on an easily identifiable number of sources and localities might have implications for preventing cholera outbreaks in Haiti and elsewhere. Whereas previous research emphasized the importance of commune-level data focused on WaSH interventions (27), we highlight the importance of fine-scale mapping and using the spatial unit most logical at an operational level. Mapping of localities in Centre Department previously was unavailable, rendering the analysis of case lists impossible. Mapping performed in this study opens doors for further analysis related to

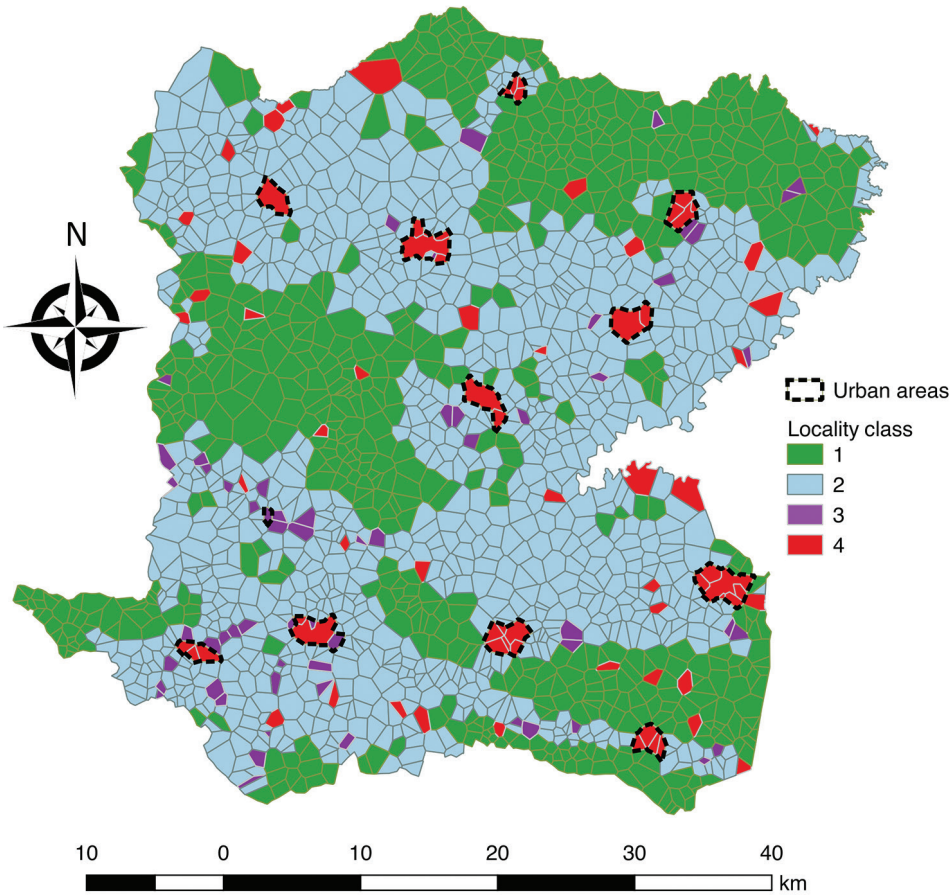



**Figure 5.** Localities of Centre Department, Haiti, mapped according to hierarchical clustering on principal components classification analysis. Classification analysis determined 2 large lower-risk classes: class 1 localities were remote and higher altitude, and class 2 localities were rural intermediary areas. Two high-risk classes were identified: class 3 localities are closer to rivers and unimproved water sources, and class 4 localities have markets and are in urban areas.

cholera and to other health conditions. Furthermore, cholera can be considered as a proxy for vulnerability to waterborne diseases, and targeting these high-risk localities can prevent other disease outbreaks.

Cholera case distribution seems to be linked to geographic patterns, specifically related to environmental hygiene and water source contamination (*28*). Previous studies have demonstrated a substantial association between cholera and populations living in proximity to water bodies (*29*–*32*), lakes (*28*,*33*,*34*), and rivers and to river density (*32*). Increasing elevation has been inversely linked to cholera distribution (*35*). However, due to geographic heterogeneity, the relationship between environmental factors does not fully explain cholera distribution in Haiti. Understanding environmental causation between water sources and cholera is confounded by the difficulties of sampling rivers and water sources and challenges in determining whether these sources are perennial reservoirs or temporary sources of pathogens (*36*,*37*). Other studies have highlighted the change in local populations' preferences and behaviors to type of water access (*38*,*39*). Furthermore, alongside the proximity to environmental factors, the role of population movement and social factors in the spatial determination of cholera must be considered (*28*,*40*).

Consistent with our results, previous cholera outbreak investigations have reported hotspots in proximity to busy markets (*9*,*12*) and main roads (*32*,*41*), presuming the role of population mobility in creating cholera transmission hubs through the fecal–oral route (*9*,*29*). These social and environmental characteristics must be considered when examining the spatial determinants of cholera in a region. This reinforces the notion that context matters. Our results demonstrate 2 different high-risk classes and highlight the heterogeneous factors across localities, even within the same communal section. We cannot point to a single clear high-risk variable but instead noted patterns of risk depending on the socioenvironmental context of each locality. Our study adds to the literature by using this adaptive contextualized approach at a fine scale across a department, not just limited to smaller urban areas, as seen previously (*9*,*12*).

Numerous reasons can explain an increased cholera risk in localities with markets (*9*). The regular large

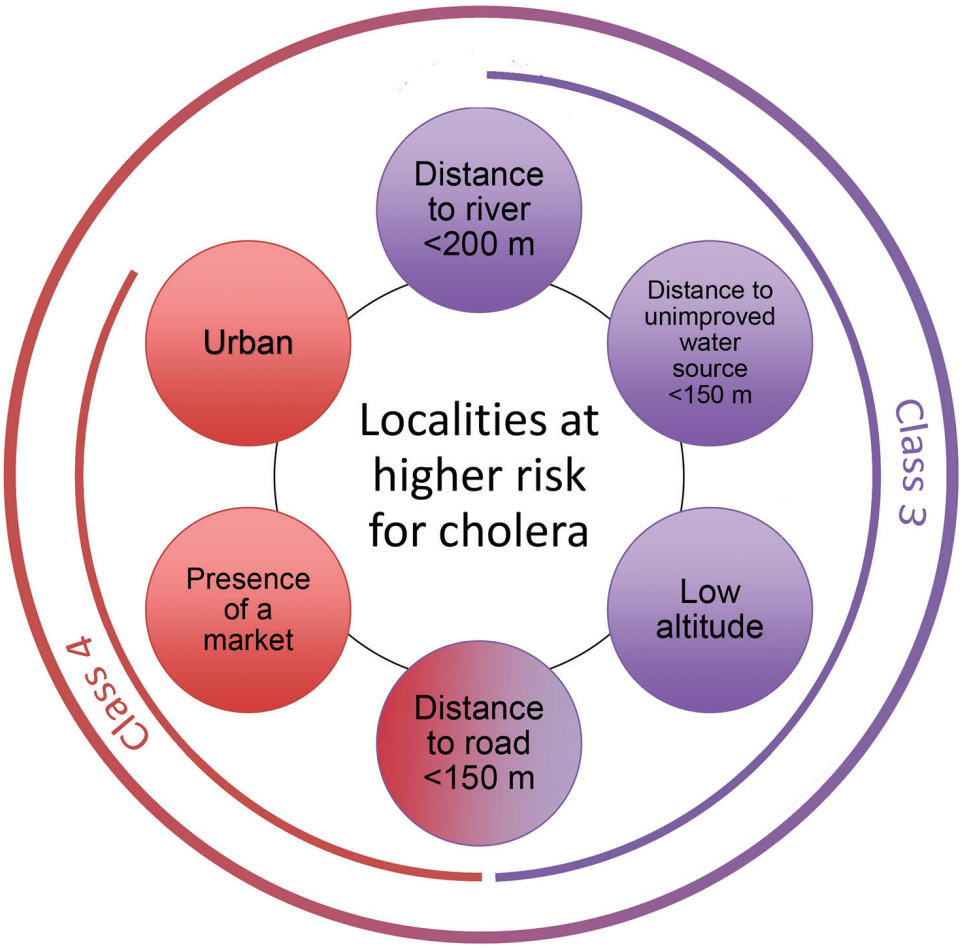



**Figure 6.** Comparison of the most strongly associated risk factors for cholera between 2 high-risk classes, Centre Department, Haiti. The most strongly associated risk factors for suspected cholera cases were distance <200 m to a river, distance <150 m to an unimproved water source, low altitude, distance to a road <150 m, presence of a market, and urban areas. Class 3 and class 4, both at higher risk for cholera, were close to a road but differed for the other identified risk factors.

flux of persons increases the risk for contact with cholera cases. In addition, the poor sanitary conditions of marketplaces increase risk because of inadequate waste disposal, lack of handwashing points, no or poor drainage systems, and the poor hygienic standards of food stalls and latrines, if available. Open defecation areas have been highlighted near Mirebalais market as a source of cholera (*42*). However, the association to markets might be confounded by the geographic proximity of treatment centers, which both often are located in urban areas where a concentration of ill patients aids cholera transmission via contamination of food and water.

Our results highlight the importance of distance to a road in both high-risk classes. Road distance <150 m was statistically significantly associated with class 3 ($p<0.0006$) and class 4 ($p<0.0001$). This finding can be interpreted in 2 ways. First, it confirms that cholera is propagated by population movement along main roads. However, main transport hubs also link localities and treatment centers; therefore, cases from localities with easier road access also are more likely to be notified. We used average estimated distances from a locality to a road as an accessible proxy for population mobility, but other information could be used, such as travel time to the nearest town (*43*). Information on actual population mobility by using mobile phone data could be a promising tool to examine spatial spread and improve outbreak preparedness strategies (*40*).

Of note, the association with distance to an improved water source was not a decisive factor in the classification. Increasing distance from an improved water source was associated more with the low-risk class 1, but unimproved water sources were also at a greater median distance. Furthermore, proximity to an improved water source was associated with the high-risk class 4. This finding suggests that the presence of improved water sources in a locality might not fully prevent diarrheal illnesses unless all other potential sources are considered for WaSH improvements. An improved water supply should be available within a 30-minute round trip, according to post-2015 Sustainable Development Goals

(*18*), but this availability cannot improve cholera incidence if an unimproved water source is still used for reasons of accessibility and affordability. Our study at the locality level did not consider individualized or household methods of drinking water collection or treatment because no piped water networks are available outside towns and those in towns do not function all the time. Instead, we studied the distance from a house to a water source, including rivers, averaged for the whole locality. However, this variable on geographic accessibility could be confounded by subsequent contamination during travel or storage.

Our study demonstrates the feasibility of fine-scale geographic analysis over a large, mainly rural area to study incidence and spatial epidemiology of cholera in relation to socioenvironmental characteristics and water sources. We used a grassroots approach incorporating expertise from local health and sanitation experts to define a list of localities and comprehensive documentation of water sources. Our approach has potential benefits for future studies by researchers in public health and other disciplines.

As can be common in resource-limited settings with multiple users, the case list required a lengthy data cleaning process. Limitations of the case list include possible missing data from patients who did not go to healthcare centers. However, we believe the case list provided the most comprehensive information available and impressive detail considering the setting constraints.

As demonstrated in previous studies, our field study corroborates that risk for cholera is associated with intertwined socioeconomic and environmental factors and highlights marketplaces located near water bodies and roads in high-density neighborhoods, such as Mirebalais (*29*), as risk factors. By analyzing a wide range of variables, without preceding presumptions on their relationship or correlation, we were able to conduct an exploratory analysis by using MCA. This analysis enabled inclusion of numerous variables despite collinearity and classification of the reduced dimensions in a multivariate model to outline the spatial determinants of cholera.

One limitation of our study was the lack of information regarding excrement management, a recognized predisposing risk factor (*29*,*44*). We attempted to retrieve this information by identifying the location of municipal latrines, but heterogeneous excrement disposal activities make this process futile because open defecation is a common practice (*45*). A recent survey of 13,405 households in Haiti reports only 31% of households have improved, nonshared toilets; in urban areas, 43% have such amenities, but half as many (23%) are found in rural areas (*45*). Another limitation is that trend-surface GAM does not fully address spatial autocorrelation by smoothing the spatial coordinates; however, it does account for trends in geographic data. Nonetheless, trend-surface GAM is a recognized method to model the spatial dependence in the systematic part of the model (*22*–*24*). Furthermore, we were unable to incorporate temporal analyses due to the relatively short timeframe (only 1 dry season). Therefore, we cannot account for all factors associated with cholera incidence, particularly the movement of cholera between localities. In addition, the geographic analysis could not incorporate meteorological information, which is unavailable at the locality level. However, our aim was not to model cholera incidence and dynamics of transmission, which previously has been studied (*46*–*50*). Instead of identifying individual patterns of risk, we used classification analysis to reduce the dimensions of numerous correlated variables to define local-scale risk profiles within a large area affected by cholera. Our study provides information to guide strategies to reduce vulnerability to diarrheal illnesses within a realistic setting, taking the social and environmental context into account.

With no confirmed cholera cases in Haiti since February 2019, the focus in this postepidemic period must be on reducing the vulnerability of the population of Haiti to cholera and other diarrheal illnesses. Our results highlight different typologies of risk at the locality level across a department, defining high risk by access to unimproved water sources and presence of markets in urban localities. Focusing hygiene awareness and prevention strategies in localities with known high-risk factors can help concentrate limited resources and improve efficiency in the fight against future cholera and other waterborne disease epidemics in Haiti and elsewhere.

**Acknowledgments**

We thank the staff from the department heath directorate, the technicians of La Direction Nationale de l'Eau Potable et de l'Assainissement, and the nongovernmental organizations Haiti Outreach, OXFAM-GB, and UNICEF for providing data and their invaluable support with field study organization.

## About the Author

Dr. Griffiths is an academic general practitioner at the Aix-Marseille University, France. Her research interests include medical statistics and epidemiology.

## References


1. Piarroux R, Barrais R, Faucher B, Haus R, Piarroux M, Gaudart J, et al. Understanding the cholera epidemic, Haiti. Emerg Infect Dis. 2011;17:1161–8. https://doi.org/10.3201/eid1707.110059
2. Gaudart J, Rebaudet S, Barrais R, Boncy J, Faucher B, Piarroux M, et al. Spatio-temporal dynamics of cholera during the first year of the epidemic in Haiti. PLoS Negl Trop Dis. 2013;7:e2145. https://doi.org/10.1371/journal.pntd.0002145
3. Ministère Santé Publique et de la Population (MSPP). Department of Epidemiology, Laboratories and Research (DELR). Report of the National Cholera Surveillance Network, 40th epidemiological week 2019 [in French] [cited 2019 Dec 1]. http://mspp.gouv.ht/site/downloads/Profil%20statistique%20Cholera%2040SE%202019.pdf
4. Republic of Haiti, Ministry of Public Health and Population, National Directorate for Water Supply and Sanitation. National plan for the elimination of cholera in Haiti 2013–2022 [in French] [cited 2019 Dec 1]. http://www.paho.org/hq/index.php?option=com_docman&task=doc_view&gid=20326&Itemid=270&lang=en
5. Republic of Haiti, Ministry of Public Health and Population, National Directorate for Water Supply and Sanitation. National plan for the elimination of cholera: development of the long term plan 2018–2022 [in French] [cited 2019 Dec 1]. http://mspp.gouv.ht/site/downloads/Plan%20a%20long%20terme%202018%202022%20pour%20l%20elimination%20du%20Cholera%20en%20Haiti.pdf
6. Allan M, Grandesso F, Pierre R, Magloire R, Coldiron M, Martinez-Pino I, et al. High-resolution spatial analysis of cholera patients reported in Artibonite Department, Haiti in 2010–2011. Epidemics. 2016;14:1–10. https://doi.org/10.1016/j.epidem.2015.08.001
7. Blackburn JK, Diamond U, Kracalik IT, Widmer J, Brown W, Morrissey BD, et al. Household-level spatiotemporal patterns of incidence of cholera, Haiti, 2011. Emerg Infect Dis. 2014;20:1516–9. https://doi.org/10.3201/eid2009.131882
8. Mukandavire Z, Smith DL, Morris JG Jr. Cholera in Haiti: reproductive numbers and vaccination coverage estimates. Sci Rep. 2013;3:997. https://doi.org/10.1038/srep00997
9. Luquero FJ, Banga CN, Remartínez D, Palma PP, Baron E, Grais RF. Cholera epidemic in Guinea-Bissau (2008): the importance of "place". PLoS One. 2011;6:e19005. https://doi.org/10.1371/journal.pone.0019005
10. Curtis A, Blackburn JK, Smiley SL, Yen M, Camilli A, Alam MT, et al. Mapping to support fine scale epidemiological cholera investigations: a case study of spatial video in Haiti. Int J Environ Res Public Health. 2016;13:187. https://doi.org/10.3390/ijerph13020187
11. Curtis A, Blackburn JK, Widmer JM, Morris JG Jr. A ubiquitous method for street scale spatial data collection and analysis in challenging urban environments: mapping health risks using spatial video in Haiti. Int J Health Geogr. 2013;12:21. https://doi.org/10.1186/1476-072X-12-21
12. Rebaudet S, Griffiths K, Trazillio M, Lebeau A-G, Abedi AA, Bulit G, et al. Cholera spatial-temporal patterns in Gonaives, Haiti: From contributing factors to targeted recommendations Adv Water Resour. 2017;108:377–85. https://doi.org/10.1016/j.advwatres.2016.12.012
13. Haitian Ministry of Economy and Finance and Haitian Institute of Statistics and Information. Total population, aged 18 and over, households and densities estimated in 2015 [in French] [cited 2019 Dec 1]. http://www.ihsi.ht/pdf/projection/Estimat_PopTotal_18ans_Menag2015.pdf
14. Milbin D. Geographical names of zones in Haiti. Ninth United Nations Conference on the standardization of geographical names [in French] [cited 2019 Dec 1]. https://unstats.un.org/unsd/geoinfo/UNGEGN/docs/9th-uncsgn-docs/crp/9th_UNCSGN_e-conf-98-crp-36.pdf
15. Rebaudet S, Bulit G, Gaudart J, Michel E, Gazin P, Evers C, et al. The case-area targeted rapid response strategy to control cholera in Haiti: a four-year implementation study. PLoS Negl Trop Dis. 2019;13:e0007263. https://doi.org/10.1371/journal.pntd.0007263
16. Haiti Outreach. FRAPE water infrastructure monitoring system [cited 2019 Dec 1]. https://frape.mwater.co/#/?locale=en
17. World Health, Organization (WHO) and the United Nations Children's Fund (UNICEF). Progress on Drinking Water, Sanitation and Hygiene: 2017 Update and SDG Baselines [cited 2019 Dec 1]. https://www.unicef.org/publications/files/Progress_on_Drinking_Water_Sanitation_and_Hygiene_2017.pdf
18. Water Supply and Sanitation Collaborative Council. WASH post-2015: proposed targets and indicators for drinking-water, sanitation and hygiene [cited 2019 Dec 1]. https://www.unicef.org/wash/files/2_WSSCC_JMP_Fact_Sheets_2_UK_LoRes.pdf
19. Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. J Stat Softw. 2008;25:1–18. https://doi.org/10.18637/jss.v025.i01
20. Husson F, Josse J, Pagès J. Principal component methods—hierarchical clustering—partitional clustering: why would we need to choose for visualizing data? Technical Report—Agrocampus: Applied Mathematics Department, September 2010 [cited 2019 Dec 1]. http://factominer.free.fr/more/HCPC_husson_josse.pdf
21. Greenacre M. Correspondence analysis in medical research. Stat Methods Med Res. 1992;1:97–117. https://doi.org/10.1177/096228029200100106
22. Wood SN. Generalized additive models: an introduction with R. Boca Raton (FL): Chapman & Hall/CRC; 2006.
23. Hastie T, Tibshirani R. Generalized additive models. Boca Raton (FL): Chapman & Hall/CRC; 1999.
24. Cressie NAC. Statistics for spatial data. New York: John Wiley & Sons; 1993.
25. Wood S. mgcv: mixed GAM computation vehicle with automatic smoothness estimation 2019 [cited 2019 Dec 1]. https://CRAN.R-project.org/package=mgcv
26. World Health Organization. Ethical standards and procedures for research with human beings [cited 2016 Feb 15]. http://www.who.int/ethics/research
27. Barzilay EJ, Schaad N, Magloire R, Mung KS, Boncy J, Dahourou GA, et al. Cholera surveillance during the Haiti epidemic—the first 2 years. N Engl J Med. 2013;368:599–609. https://doi.org/10.1056/NEJMoa1204927
28. Rebaudet S, Sudre B, Faucher B, Piarroux R. Environmental determinants of cholera outbreaks in inland Africa: a systematic review of main transmission foci and propagation routes. J Infect Dis. 2013;208(Suppl 1):S46–54. https://doi.org/10.1093/infdis/jit195
29. Olanrewaju OE, Adepoju KA. Geospatial assessment of cholera in a rapidly urbanizing environment. J Environ Public Health. 2017;2017:6847376. https://doi.org/10.1155/2017/6847376
30. Mengel MA, Delrieu I, Heyerdahl L, Gessner BD. Cholera outbreaks in Africa. Curr Top Microbiol Immunol. 2014;379:117–44. https://doi.org/10.1007/82_2014_369
31. Birmingham ME, Lee LA, Ndayimirije N, Nkurikiye S, Hersh BS, Wells JG, et al. Epidemic cholera in Burundi:

patterns of transmission in the Great Rift Valley Lake region. Lancet. 1997;349:981–5. https://doi.org/10.1016/S0140-6736(96)08478-4
32. Bompangue Nkoko D, Giraudoux P, Plisnier P-D, Tinda AM, Piarroux M, Sudre B, et al. Dynamics of cholera outbreaks in Great Lakes region of Africa, 1978–2008. Emerg Infect Dis. 2011;17:2026–34.
33. Bompangue D, Giraudoux P, Piarroux M, Mutombo G, Shamavu R, Sudre B, et al. Cholera epidemics, war and disasters around Goma and Lake Kivu: an eight-year survey. PLoS Negl Trop Dis. 2009;3:e436. https://doi.org/10.1371/journal.pntd.0000436
34. Bompangue D, Giraudoux P, Handschumacher P, Piarroux M, Sudre B, Ekwanzala M, et al. Lakes as source of cholera outbreaks, Democratic Republic of Congo. Emerg Infect Dis. 2008;14:798–800. https://doi.org/10.3201/eid1405.071260
35. Luque Fernandez MA, Schomaker M, Mason PR, Fesselet JF, Baudot Y, Boulle A, et al. Elevation and cholera: an epidemiological spatial analysis of the cholera epidemic in Harare, Zimbabwe, 2008–2009. BMC Public Health. 2012;12:442. https://doi.org/10.1186/1471-2458-12-442
36. Alam MT, Weppelmann TA, Weber CD, Johnson JA, Rashid MH, Birch CS, et al. Monitoring water sources for environmental reservoirs of toxigenic *Vibrio cholerae* O1, Haiti. Emerg Infect Dis. 2014;20:356–63. https://doi.org/10.3201/eid2003.131293
37. Rebaudet S, Piarroux R. Monitoring water sources for environmental reservoirs of toxigenic *Vibrio cholerae* O1, Haiti. Emerg Infect Dis. 2015;21:169–70. https://doi.org/10.3201/eid2101.140627
38. MacRae D, Whittington D. Assessing preferences in cost-benefit analysis: reflections on rural water supply evaluation in Haiti. J Policy Anal Manage. 1988;7:246. https://doi.org/10.2307/3323827
39. Patrick M, Steenland M, Dismer A, Pierre-Louis J, Murphy JL, Kahler A, et al. Assessment of drinking water sold from private sector kiosks in post-earthquake Port-au-Prince, Haiti. Am J Trop Med Hyg. 2017;97(Suppl):84–91. https://doi.org/10.4269/ajtmh.16-0692
40. Bengtsson L, Gaudart J, Lu X, Moore S, Wetter E, Sallah K, et al. Using mobile phone data to predict the spatial spread of cholera. Sci Rep. 2015;5:8923. https://doi.org/10.1038/srep08923
41. Troeger C, Gaudart J, Truillet R, Sallah K, Chao DL, Piarroux R. Cholera outbreak in Grande Comore: 1998–1999. Am J Trop Med Hyg. 2016;94:76–81. https://doi.org/10.4269/ajtmh.15-0397
42. Maire de Mirebalais, Regional Office for Potable Water and Sanitation, ESA Consultance. Analysis of factors of persistence of cholera and the elaboration of an intervention plan for the town of Mirebalais [in French] [cited 2019 Dec 1]. https://en.calameo.com/read/005285856479e515bd3ba
43. Weiss DJ, Nelson A, Gibson HS, Temperley W, Peedell S, Lieber A, et al. A global map of travel time to cities to assess inequalities in accessibility in 2015. Nature. 2018;553:333–6.
44. Esrey SA, Potash JB, Roberts L, Shiff C. Effects of improved water supply and sanitation on ascariasis, diarrhoea, dracunculiasis, hookworm infection, schistosomiasis, and trachoma. Bull World Health Organ. 1991;69:609–21.
45. Haitian Institute for Children. ICF International. Mortality, morbidity and use of services survey 2016–2017 [in French] [cited 2019 Dec 1]. https://www.dhsprogram.com/pubs/pdf/FR326/FR326.pdf
46. DuBois AE, Sinkala M, Kalluri P, Makasa-Chikoya M, Quick RE. Epidemic cholera in urban Zambia: hand soap and dried fish as protective factors. Epidemiol Infect. 2006;134:1226–30. https://doi.org/10.1017/S0950268806006273
47. Dunkle SE, Mba-Jonas A, Loharikar A, Fouché B, Peck M, Ayers T, et al. Epidemic cholera in a crowded urban environment, Port-au-Prince, Haiti. Emerg Infect Dis. 2011;17:2143–6. https://doi.org/10.3201/eid1711.110772
48. Kanungo S, Sur D, Ali M, You YA, Pal D, Manna B, et al. Clinical, epidemiological, and spatial characteristics of *Vibrio parahaemolyticus* diarrhea and cholera in the urban slums of Kolkata, India. BMC Public Health. 2012;12:830. https://doi.org/10.1186/1471-2458-12-830
49. Morof D, Cookson ST, Laver S, Chirundu D, Desai S, Mathenge P, et al. Community mortality from cholera: urban and rural districts in Zimbabwe. Am J Trop Med Hyg. 2013;88:645–50. https://doi.org/10.4269/ajtmh.11-0696
50. Grandesso F, Allan M, Jean-Simon PSJ, Boncy J, Blake A, Pierre R, et al. Risk factors for cholera transmission in Haiti during inter-peak periods: insights to improve current control strategies from two case-control studies. Epidemiol Infect. 2014;142:1625–35. https://doi.org/10.1017/S0950268813002562

Address for correspondence: Karolina Griffiths, Montpellier University, 2 rue école de Médecine, 34000 Montpellier, France; email: karolina.griffiths@umontpellier.fr

Article DOI: https://doi.org/10.3201/eid2701.191787

# Delineating and Analyzing Locality-Level Determinants of Cholera, Haiti

## Appendix

### SatScan Analysis

Spatial analysis for cholera hotspot detection was computed by using SaTScan version 9.4.2 (*1*). A hotspot was defined as a statistically high-risk cluster in which incidence of suspected cholera cases is higher than in the neighboring areas and was a purely elliptic spatial analysis performed by using a discrete Poisson model. We estimated p values by using Monte Carlo inference (Appendix Table 2).

The results demonstrate a spatial heterogeneity and identified 9 major hotspot ellipses that are comprised of a total of 262 localities. We mapped the location of these 9 hotspot ellipses (Appendix Figure 2), then compared them with the classification of high-risk clusters identified by the hierarchical clustering on principal components (HCPC) methodology. However, due to the diverse differing characteristics of each hotspot, the variables that most contributed to the clustering are not apparent; therefore, interpretation of the numerous clusters was more difficult. We categorized a binary outcome variable by using the SatScan results, so that locality within a hotspot is 1 and locality outside of a hotspot is 0.

Univariate Kruskal-Wallis tests showed a statistically significant difference in cholera hotspots between the 4 classes of localities, $x^2(3) = 18.8$, $p = 0.0003$. Lower risk classes 1 and 2 and higher risk classes 3 and 4 among localities identified by the HCPC methodology were compared with the binary outcome “hotspot” variable in a binomial general additive model. To model spatial dependence, the geographic location was fitted by using splines as a trend surface and we considered $p<0.05$ statistically significant.

Using lower risk classes as the referent, the general additive model (GAM) showed a statistically significant difference for higher-risk classes being associated with a cholera hotspot.

The GAM confirms the statistically significant higher odds of the higher risk classes being in a cholera hotspot, odds ratio (OR) 3.57 (95% CI 1.63–7.83, $p = 0.001$).

## Moran I Inferential Test

Moran I inferential test for spatial autocorrelation was calculated on the raw data, including total number of cases (taking into account the population) and GAM residuals with the spatial coordinates as a smooth term. We used the R package spdep (R Foundation for Statistical Computing, https://www.r-project.org) to compute the Moran I autocorrelation statistic. For the raw total number of cases data, the $p<0.05$ was considered statistically significant ($p = 0.043$) and rejected the null hypothesis of complete spatial randomness, which demonstrates spatial dependence (or autocorrelation) in the raw data.

Thus, we fitted the geographic location by using splines as a trend-surface (2-dimensional spline on spatial projection coordinates). Then we repeated the Moran I test on the GAM residuals, which had a nonstatistically significant p value ($p = 0.055$), seeming to suggest that the model helps to account for the trends of spatial autocorrelation in the data.

Although trend-surface GAM does not fully address the issue of spatial autocorrelation by including a smooth of the spatial locations (projected coordinates), it does account for trends in geographic data (*2*). Trend-surface GAM is a recognized method to model the spatial dependence in the systematic part of the model (*3*,*4*).

## Visual Examples of Locality in Each Class

We provide an example of a locality most representative of each of the 4 main classes of the HCPC (Appendix Figure 3). This paragon helps visualize the differences between each class and provides an example with relevant features. The features portrayed in the figure include the incidence rate (circle size and attributable to incidence rate); area-averaged distance from houses to rivers, roads, improved water sources, unimproved water sources; marketplaces; and vaccinated polygons. Further information, including altitude, is provided in Appendix Table 1.

## References


1. Kulldorff M; Information Management Services Inc. SaTScan: software for the spatial, temporal, and space-time scan statistics [cited 2019 March 1]. https://www.satscan.org

2. Cressie NAC. Statistics for spatial data. New York, New York: John Wiley & Sons; 1993.

3. Hastie TJ, Tibshirani RJ. Generalized additive models. New York, New York: Chapman & Hall/CDC; 1990.

4. Wood SN. Generalized additive models, an introduction with R, second edition. Boca Raton, FL: Chapman & Hall/CDC; 2006.


**Appendix Table 1.** Univariate analysis, nonparametric Spearman correlation coefficients, and p values used to delineate and analyze locality-level determinants of cholera, Haiti

| Variable | Variable | Correlation coefficient | p value* |
|---|---|---|---|
| Incidence | Mean altitude | −0.219877577 | 0 |
| Incidence | Distance to a road | −0.231237548 | 0 |
| Incidence | Distance to unimproved water source | −0.118602227 | <0.0001 |
| Incidence | Distance to improved water source | −0.185149843 | <0.0001 |
| Incidence | Distance to river | −0.147558669 | <0.0001 |

*Quantitative variables.

**Appendix Table 2.** List of cholera hotspots detected by SaTScan, Haiti*

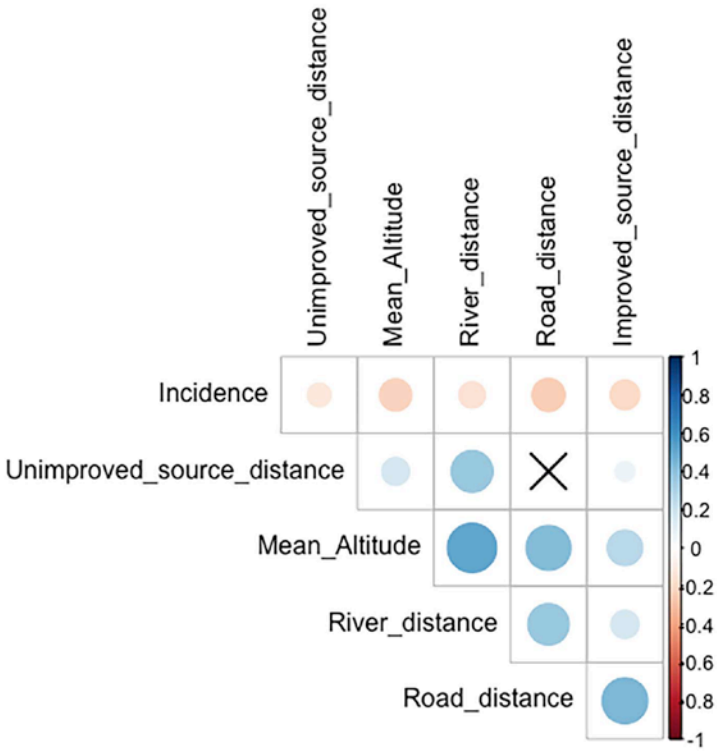
| High-risk hotspot | Relative risk | p value |
|---|---|---|
| 1 | 2.15 | <0.0001 |
| 2 | 2.81 | <0.0001 |
| 3 | 6.53 | <0.0001 |
| 4 | 2.10 | <0.0001 |
| 5 | 14.60 | <0.0001 |
| 6 | 21.42 | <0.0001 |
| 7 | 16.20 | <0.0001 |
| 8 | 9.73 | <0.0001 |
| 9 | 7.85 | 0.0002 |
| 10 | 3.67 | 0.999 |

*M. Kulldorff, https://www.satscan.org (*1*).

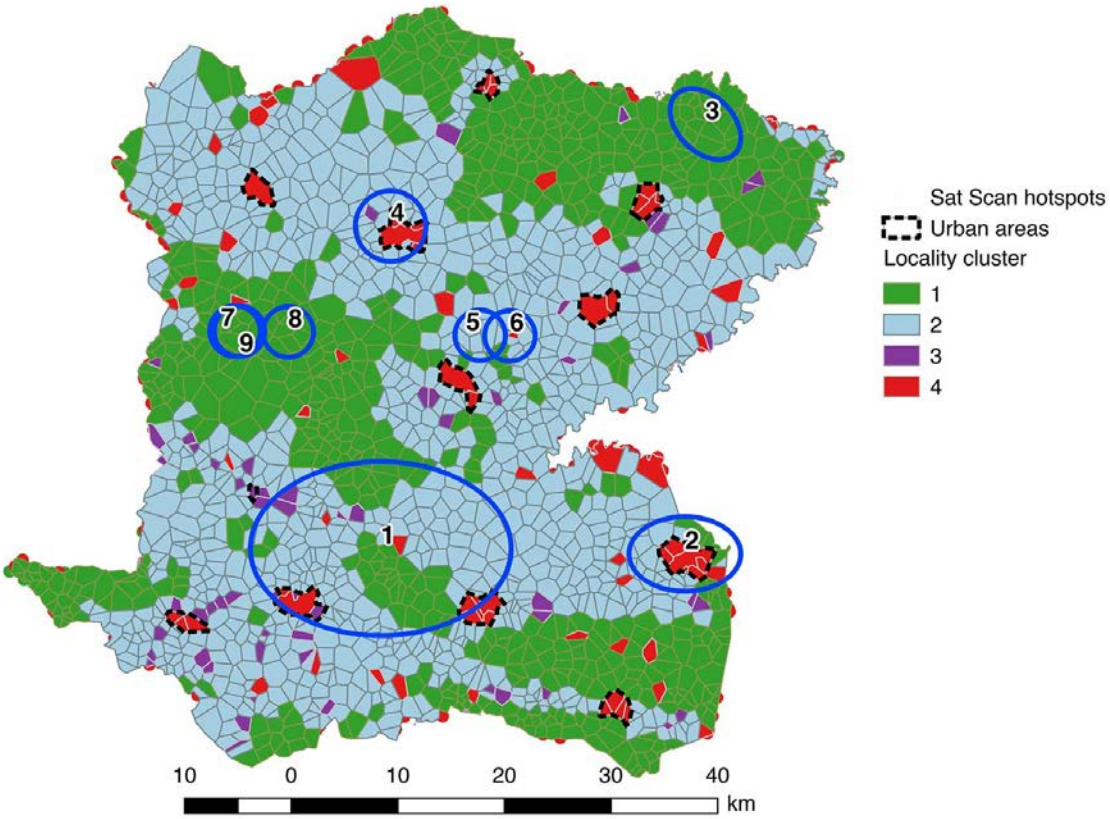
**Appendix Table 3.** Variables used to create visualization of representative localities for each class in a spatial analysis of cholera, Haiti

| | Class 1 locality | Class 2 locality | Class 3 locality | Class 4 locality |
|---|---|---|---|---|
| Variables for each locality | Nan Chambo | Palmari | Moro | Cerca la Source Ville |
| No. suspected cholera cases | 1 | 1 | 7 | 34 |
| No. houses | 65 | 113 | 41 | 1,117 |
| Estimated population | 300 | 521 | 189 | 5,149 |
| Estimated incidence/10,000 persons | 33.37 | 19.20 | 370 | 66 |
| Altitude, m in meters*† | 764 | 414 | 224 | 376 |
| Distance to nearest improved water source, m*† | 2,194 | 385 | 675 | 155 |
| Distance to nearest unimproved water source, m*† | 154 | 455 | 189 | 292 |
| Distance to nearest river, m*† | 2,073 | 575 | 189 | 508 |
| Distance to nearest, m*† | 4,785 | 1,778 | 1,106 | 9 |
| Presence of market * | No | No | No | Yes |
| Rural locality* | Yes | Yes | Yes | No |

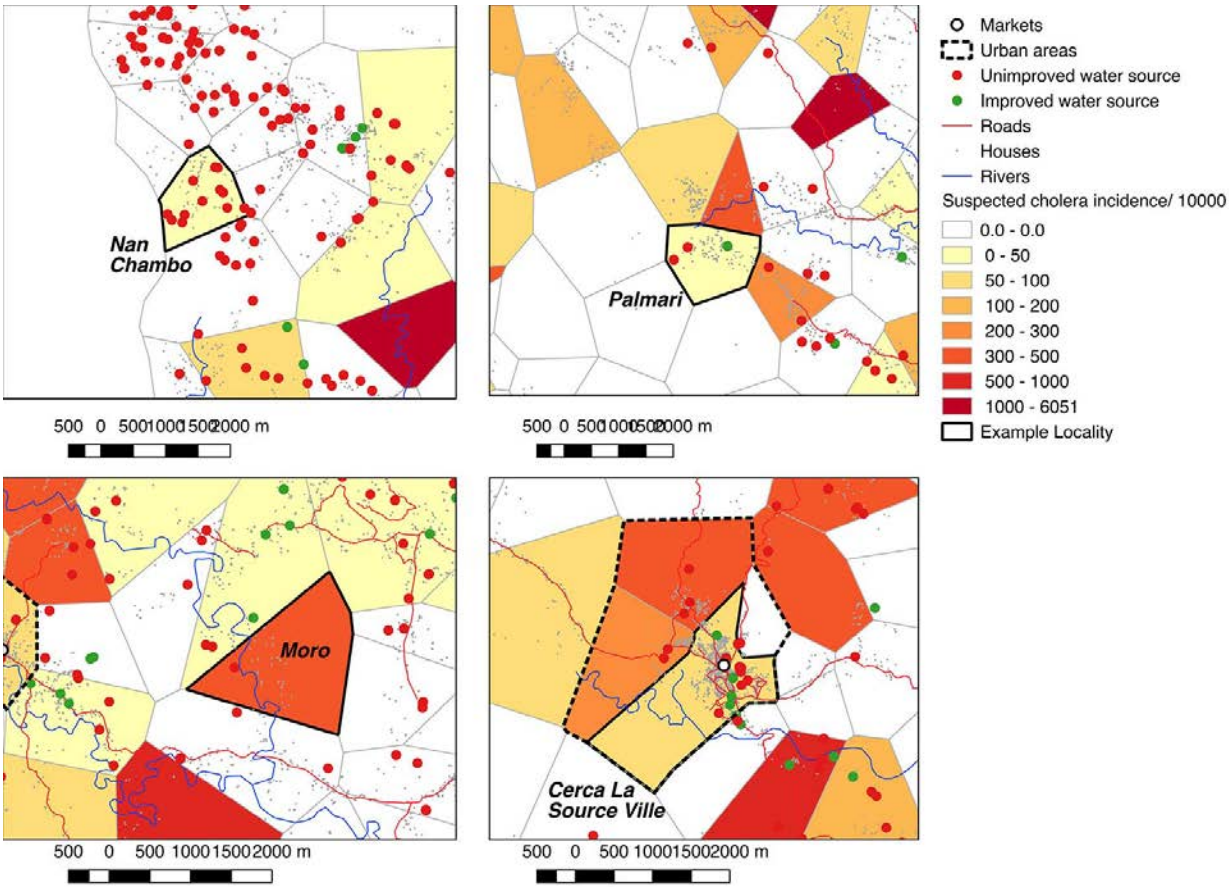
*Active variables used in classification analysis.
†Transformed to categorical variables for analysis.




**Appendix Figure 1**. Correlation matrix of factors associated with high risk for cholera, Centre Department, Haiti. We used Spearman correlation coefficients to perform univariant analysis on cholera incidence and environmental variables. Blue indicates positive correlations (i.e., lower risk for cholera) and red indicates negative correlations (i.e., higher risk for cholera). Color intensity and the size of the circle are proportional to the correlation. Legend color shows the correlation coefficients and the corresponding colors. X indicates non-significative result.




**Appendix Figure 2.** Locality map demonstrating the 4 classes identified by hierarchical clustering on principal components, Centre Department, Haiti. Numerals in blue circles indicate the 9 cholera hotspots identified by SatScan.




**Appendix Figure 3.** Maps demonstrating fine scale data on factors associated with risk for cholera, Centre Department, Haiti. Nan Chambo (top left) is a remote class 1 locality at high altitude, far from rivers and roads; Palmari (top right) is a rural intermediary class 2 at lower altitude and a medium distance from a river; Moro (lower left) a class 3 locality close to rivers and unimproved water sources; and Cerca La Source Ville (lower right) is a class 4 urban locality that has markets.