LETTER

Biophysical, infrastructural and social heterogeneities explain spatial distribution of waterborne gastrointestinal disease burden in Mexico City

Andrés Baeza1, Alejandra Estrada-Barón2, Fidel Serrano-Candela3, Luis A Bojórquez2, Hallie Eakin1 and Ana E Escalante2,3

1 School of Sustainability, Arizona State University, Tempe AZ 85287, United States of America
2 Laboratorio Nacional de Ciencias de la Sostenibilidad (LANCIS), Instituto de Ecología, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico
3 Author to whom any correspondence should be addressed.

E-mail: aescalante@iecologia.unam.mx

Keywords: flooding, giardiasis, neighborhood effect, spatial statistics, water management, Mexico City

Abstract

Due to unplanned growth, large extension and limited resources, most megacities in the developing world are vulnerable to hydrological hazards and infectious diseases caused by waterborne pathogens. Here we aim to elucidate the extent of the relation between the spatial heterogeneity of physical and socio-economic factors associated with hydrological hazards (flooding and scarcity) and the spatial distribution of gastrointestinal disease in Mexico City, a megacity with more than 8 million people. We applied spatial statistics and multivariate regression analyses to high resolution records of gastrointestinal diseases during two time frames (2007–2009 and 2010–2014). Results show a pattern of significant association between water flooding events and disease incidence in the city center (lowlands). We also found that in the periphery (highlands), higher incidence is generally associated with household infrastructure deficiency. Our findings suggest the need for integrated and spatially tailored interventions by public works and public health agencies, aimed to manage socio-hydrological vulnerability in Mexico City.

1. Introduction

Today, more than 50% of the human population lives in cities. This percentage is expected to rise to 70% by the year 2050 (United Nations 2016). Urbanization has had a great impact on improving quality of life, but it poses great challenges to the planning and management of the built environment, urban services and economic policy (Dye 2008, UN-HABITAT 2017). Urbanization in the developing world is often unregulated and unplanned and occurs most rapidly in the peripheries of cities, which are often highly vulnerable to different hazards such as extreme rainfall, flooding, landslides and water scarcity (Pick and Butler 1997, IPCC 2012). Unplanned growth also results in significant spatial heterogeneity in terms of culture, economy, and infrastructure. This creates inequalities in the provision of urban services, exposure to hazards, and responses from managers and infrastructure providers (Zhou et al 2017, Ahern 2013, Pickett et al 2017, Jacobi et al 2010). In the literature regarding vulnerability, it has long been understood that risk (i.e. the probability that a potentially damaging event will occur) is a product of both environmental conditions and social processes that determine vulnerability (Turner et al 2003). In cities, these social and political processes determine the actions of governments and decision-makers, who in turn influence the creation of the built environment, socio-economic heterogeneity and the concomitant vulnerability to hydrological hazards (Eakin et al 2017).

Despite its relevance for urban planning and health policy decisions, we still lack a full understanding of the role of socio-economic attributes, hydrological processes and the built environment in the production of epidemiological vulnerability in large urban
areas (Santos-Vega et al 2016a, Vlahov et al 2007, Vlahov et al 2005). In other words, more research is needed on the association between socio-hydrological vulnerability and epidemiological outcomes. In particular, there has been a growing interest in understanding how the socio-economic heterogeneities found in large cities influence the spread and persistence of infectious diseases (Ahern et al 2005). They can cause, for example, heterogeneity in the spatial distribution of pathogens. Waterborne pathogens are of particular interest in the literature regarding epidemiology, as they can be in contact with human populations in several ways, and cause a variety of infectious diseases (Hunter 1997). From an epidemiological point of view, it is crucial to understand the role that socio-hydrological processes and water management play on epidemiological outcomes in urban areas.

In this paper, we present statistical analyses that show the relation between socio-hydrological risk factors in Mexico City and the spatial distribution of risk of gastrointestinal diseases associated with waterborne pathogens. Thus, this work contributes to socio-hydrology, a growing body of research aimed at understanding the dynamics and co-evolution of coupled human-water systems (Sivapalan et al 2012). The ultimate goal of socio-hydrological research is to elucidate the two-way coupling, feedbacks and cross-scalar temporal and spatial dynamics between social processes and hydrological systems. The first step is to analyze multiscale, time-space patterns and dynamics of social-hydrological processes and to explain and interpret socio-hydrological responses in terms of outcomes relevant to human well-being (Sivapalan et al 2014: 226). Applied to flood research, the concept of socio-hydrological vulnerability indicates that the undesirable outcomes of flooding—including human exposure to pathogens—are the product not only of unexpected or anomalous hydrological events, but also of social processes such as urbanization, infrastructure design and maintenance, policy priorities and human behavior. Ideally, the direction and nature of feedbacks can be made explicit through socio-hydrological analyses. First, patterns of social and hydrological association have to be made visible. Under this framework, socio-hydrological risk is defined as the potential for harm which results from the combination of both hydrological processes (e.g. precipitation, runoff, subsidence) and social processes (e.g. the built environment, infrastructural conditions, human behavior and decision-making).

We propose that the spatial heterogeneity of the factors that shape the vulnerability of Mexico City to hydrological hazards is directly related to the spatial heterogeneity of gastrointestinal disease incidence (GDI). To test this hypothesis, we conducted a set of spatial statistical analyses using data from public records of GDI for the 2007–2014 period. The results show strong spatial differentiation in the type of socio-hydrological risk factors that correlate with gastrointestinal disease incidence between the center and the periphery of the city. To these we assigned two geomorphological categories: lowlands for the center and highlands for the periphery. We concluded that health and water authorities must foster dialogue and initiatives, in order to evaluate how they can tackle together the challenges posed by water infrastructure management and health problems associated with waterborne pathogens.

2. Data and methods

2.1. Data collection and handling

We compiled a database of GDI, socio-hydrological risk factors, socio-economic variables and location for the 2432 census blocks. The location of census blocks was related to two general geomorphological categories: ‘highlands’ and ‘lowlands.’ Highlands correspond to the hillslopes, mostly peripheral to the city, and lowlands correspond to the former lakebed of the Mexico City Basin, mostly in the center. We included all available data of the spatial resolution of census blocks for the entirety of the city.

2.1.1. Gastrointestinal disease incidence (GDI)

Data of gastrointestinal disease cases for 279 primary care public clinics were obtained from the annual reports of the Secretaria de Salud (Secretariat of Health) for the 2007–2014 period. The data belonged to patients from neighborhoods in close vicinity to each primary care clinic, and included clinic identifier, healthcare agency (IMSS, ISSSTE and SS), and causative pathogenic agent (parasitic, bacterial or undetermined) (table S1 available at stacks.iop.org/ERL/13/064016/mmedia). Close vicinity was defined using the criteria established by health authorities in Mexico, which delimited ‘areas of influence’ as the census blocks located at the nearest distance from a primary care clinic (spatial distribution of public clinics is presented in figure S3). To calculate how the total number of cases of an area of influence were distributed, we assumed that GDI per primary care clinic (i.e. the total number of cases in a census block was divided by its population, assuming the entire population at risk) was distributed in each census block proportional to its population.

2.1.2. Flooding records

Official records of flooding events were obtained from the Sistema de Aguas de la Ciudad de Mexico (SACMEX) for the 2007–2014 period. Since all data was at neighborhood scale, we needed to calculate the probability of occurrence for each census block. Considering that neighborhoods encompass all or part of several census blocks, we calculated (1) the proportion of area in each neighborhood that intersected with each census block (http://github.com/sostenibilidad-unam/colonias_to_agebs); and (2) the probability of...
flooding events occurring in each neighborhood as the ratio of flooding events in a neighborhood and in the city. Hence, the probability of occurrence of flooding events in a census block was the sum of probabilities of occurrence in the neighborhoods encompassed by that census block, weighted by the area of the intersection between the neighborhood and the census block.

2.1.3. Residential services and socioeconomic variables

We compiled data on the number of homes per census block with access to piped water and sewage system, as well as on construction materials, toilet facilities, household overcrowding, level of education of the inhabitants and percentage of people under 18. Data was taken from the Population Censuses conducted in 2005 and 2010 (INEGI 2011). To complement this information, we created a purchasing power index as a proxy for income, using the methodology established by CONEVAL (CONEVAL 2013). Finally, we obtained data on the location of sewers and informal food stands in the city from the Mexico City Statistical Department. All the socio-economic variables that reported number of people per census block were normalized by population size to use proportion instead of absolute values.

2.2. Data analyses

Since geo-referenced data face a type of effect known as spatial dependence (Anselin 1988, Anselin 2001), conventional statistical tools are not appropriate to analyze the spatial association of GDI with socio-hydrological risk factors at the city scale. The concept of spatial dependence or autocorrelation derives from the premise that geographic units which are closer in space have a tendency to influence each other and consequently have similar attributes. In other words, the magnitude of the variable of interest may be determined by the values of the same or other variables in locations nearby. Therefore, a proper analysis of spatial data that takes into account these spatial dependencies is needed for this case.

We relied on the use of a statistical modeling approach that considers spatial dependence when determining univariate and multivariate correlations and when using regression analysis (Anselin 1988). We first conducted univariate statistical tests to measure the spatial dependence of the GDI records and of the socio-hydrological risk factors. We then conducted bivariate statistical tests to evaluate the global dependence of GDI on each factor. We then evaluated the local influence of each factor on the GDI in each census block using spatial correlation methods. Finally, we conducted regression analyses to formally evaluate the relative contribution of each factor, including the spatial association of the data, in explaining GDI. All analyses were run for two different periods of time: 2007–2009 and 2010–2014.

2.2.1. Spatial autocorrelation of GDI

To evaluate the spatial dependence of GDI and of the socio-hydrological risk factors independently, a univariate Moran Index (Moran’s I) was calculated (Anselin 1988, Anselin 1996). Moran’s I is used to identify the global degree of spatial association of data. That is, how much the magnitude of an indicator in one location is influenced by the magnitude of the indicator in an area close to it (Anselin 1996, Anselin 2001). To calculate the Moran’s I, a contiguity weight matrix had to be specified (appendix A). This matrix was also used for the subsequent statistical analyses.

2.2.2. Global spatial association of GDI and socio-hydrological risk factors

In order to determine the existence of a spatial relationship between GDI and the analyzed factors, we conducted a bivariate statistical test using the bivariate Moran Index (Moran 1985). This index captures the influence of an independent variable (e.g. flooding events) on the magnitude of a dependent variable (i.e. GDI) including not only the effect of the dependent variable on each analyzed census block, but also the effect of the neighbors, which were defined by the contiguity weight matrix. This indicator reports in a single value the global spatial association between GDI and each socio-hydrological factor.

2.2.3. Local indicators of spatial association (LISA) of GDI and socio-hydrological risk factors

The next step was to identify areas of the city with different levels of disease burden (GDI) that responded similarly to a particular socio-hydrological factor. For this we calculated the Local Indicators of Spatial Association (LISA) (Anselin 1995). These indicators were calculated in order to evaluate the spatial dependence of the relation between the level of GDI in each census block and the magnitude of the socio-hydrological factor or variable in the nearest neighborhoods, which are defined, again, by the contiguity matrix. These indicators signify location differences in the association between the dependent (GDI) and independent variables (analyzed factors). Depending on the sign of the indicator (positive or negative), these local associations can express positive-positive, positive-negative, negative-positive or negative-negative associations. Only positive-positive associations with GDI are reported in the main text; all other associations are reported in the supplementary material (figures S1(a), (b), S2(a) and (b)).

2.2.4. Spatial regression model

To evaluate the relative contribution of each socio-hydrological factor in explaining the variance of GDI in space, we constructed a multivariate regression model that considered the socio-hydrological risk factors altogether (i.e. socio-economic, demographic and hydrological) in a single equation. We first specified a
model that included all the variables that were first analyzed independently. We then conducted an Akaike information criteria test in order to obtain a model that considered only the variables which explained most of the variance in the data, without over-fitting. Next, we tested for multicollinearity in the model specification by using the variance inflation factor (VIF) and the tolerance (TOL) (Marquardt 1970) diagnostic measures of multicollinearity (table S2). To evaluate if a spatial parameter had to be included in the final model, a Moran’s test was conducted on the residuals of the regression. If this test reports a significative value, it means that the regression residuals are spatially autocorrelated. Thus, a Lagrange multiplier test was applied to determine the type of spatial dependence to include in the model; it could be included as a residual parameter or as a spatial lag parameter (Anselin 2001, Anselin 2003, Elhorst 2014) (appendix A). In addition to the regression model, we tested for variance of the different factors in highlands and lowlands, for both studied periods, with an F-test (Gujarati and Porter 2008).

3. Results

3.1. Spatial distribution of GDI
The spatial distribution of GDI for each group of pathogens over the entire city reveals that the incidence records are concentrated in the city center—in the lowlands—and at the periphery of the city—in the highlands (figure 1). This pattern is consistent for the different categories of pathogens (parasitic, bacterial, and non-well defined or undetermined) and for the two analyzed periods (2007–2009 and 2010–2014). We also observed an overall reduction in cases from period 1 to period 2 (table S3). Infections caused by protozoans (parasitic) were more frequently observed in the city center. Nonetheless, a significant number of cases was also observed at the periphery (figure 1(b)). The incidence of bacteria-related infections (bacterial) was less commonly observed and mostly concentrated in the center city, in the lowest part of the watershed.

3.2. Spatial autocorrelation of variables and GDI
We found significant spatial autocorrelation between the studied variables and GDI for all groups of causative pathogens at the global or entire city scale (table 1). The concentration of incidences reported in the city center as well as in the periphery contributes to this high spatial autocorrelation. Similarly, socio-hydrological factors are significantly autocorrelated. In particular, the spatial autocorrelation of the flood events is high and significant in the Moran’s I. This is due to the fact that most of the flooding events are concentrated in the city center (table S3).

Results from the bivariate analyses indicate a spatial association of most of the analyzed socio-hydrological risk factors with GDI. Of the eleven factors analyzed using the bivariate Moran Index, eight presented a significant spatial autocorrelation with GDI (table 1). Of these, five were positively correlated with GDI: number of flooding events, number of sewer drains, homes without toilet facilities, street food stands and level of education. The variables that were negatively correlated with GDI are: percentage of people < 18 years old and homes with dirt floor and overcrowding.

3.3. The socio-hydrological risk factors associated with GDI differ at the local scale
The results from the local indicators of spatial association (LISA) test showed that the factors spatially associated with high GDI differed significantly between the city center (lowlands) and periphery (highlands) (figure 2). Specifically, incidence in the center had a positive correlation (positive/positive) with flooding reports, percentage of homes without toilet facilities and food stands. In the periphery, there was a negative association between GDI and income in the defined neighborhoods. Homes without toilet facilities also presented a significant positive association with GDI in the periphery of the city. A full display of the analysis, including all factors and correlations, is presented in the supporting material (figures S1(a), (b), S2(a) and (b)).

3.4. The influence of the spatial component in the city center explains a 50% increase in the risk of GDI
The spatial pattern observed by mapping the distribution of cases in the city, as well as the differences evidenced by the LISA test, suggest strong differences between the center and the periphery in terms of both the GDI and the factors that contribute to it.

Results from the regression model developed for lowlands reported that flooding is the most consistent factor associated with GDI (table 2). The model also evidences the existence of spatial dependence of GDI. This spatial dependence is captured in the regression as the spatial lag parameter (rho) for the dependent variable (GDI), which is a weighted average of GDI in neighboring locations. Similar results were obtained when the regression was applied to each type of pathogen independently. Moreover, we observed the significant association of other factors (homes with dirt floors, homes without water supply, level of education, sewer drains, street food stands, overcrowding) with GDI in the 2007–2009 period, none of which had the level of consistency which flooding had, either in time or in relation to the explanatory agents.

In the case of the regression model developed for highlands, different factors were significantly associated with GDI, and these factors changed depending on the pathogen group which was analyzed. Nonetheless, one factor was significant in all regression cases: the percentage of homes with dirt floors. Additionally,
Figure 1. Distribution of GDI at census block scale in Mexico City for the 2010–2014 period. (a) Total GDI; (b) GDI attributable to parasitic (protozoans) microorganisms; (c) GDI attributable to bacterial microorganisms; (d) GDI attributable to undetermined or not well-defined causal agent. For (a) and (d), red census blocks correspond to those with incidence rate > 1 (case/1000 inhabitants per year). For (b) and (c), tones of red indicate low (light) or high (dark) incidence rates. Municipality limits are represented by solid black lines. Grey background depicts elevation differences.

the percentage of homes without toilet facilities explains GDI for the parasitic cases in 2007–2009 and for all cases in 2010–2014. Finally, for the highlands, the regression did not include a residual or lag spatial parameter, which indicates that the spatial association is not as strong in the highlands.

4. Discussion

Spatial statistical analyses conducted on GDI records of the highly populated Mexico City revealed significant association between incidence and socio-hydrological risk factors, such as flooding and lack of...
Table 1. Bivariate spatial autocorrelation index between GDI and analyzed factors. The table shows the autocorrelation coefficient obtained from the bivariate Moran’s I for each analyzed period (2007–2009; 2010–2014). The asterisks show the level of significance: ***, 99%; **, 95%; *, 90%.

| Factor                        | 2007–09 Causative pathogenic agent | 2007–09 Causative pathogenic agent |
|-------------------------------|-----------------------------------|-----------------------------------|
|                               | All | Parasitic | Bacterial | Undetermined | All | Parasitic | Bacterial | Undetermined |
| Flooding                      | 0.068*** | 0.019** | 0.068*** | 0.076*** | 0.156*** | 0.060*** | 0.057*** | 0.170*** |
| Homes with dirt floor         | −0.031*** | −0.009 | −0.052*** | −0.035*** | −0.046*** | −0.013** | −0.020*** | −0.057*** |
| Homes without water supply    | −0.007 | 0.008 | −0.026*** | −0.010 | 0.000 | 0.010 | −0.005 | −0.004 |
| Homes without toilet facilities | 0.023* | 0.023* | 0.017** | 0.022* | 0.058*** | 0.025* | 0.009 | 0.067** |
| Homes without sewage          | 0.007 | 0.012 | −0.007 | 0.006 | 0.001 | 0.008 | −0.001 | −0.001 |
| % people < 18 years old       | −0.041*** | −0.011 | −0.053*** | −0.046*** | −0.013** | 0.008 | −0.019 | −0.019** |
| Level of education            | 0.055*** | 0.007 | 0.087*** | 0.063*** | 0.029** | 0.003 | 0.012 | 0.036** |
| Income Index                  | −0.006 | −0.019 | 0.039*** | −0.003 | −0.011 | −0.017 | 0.001 | −0.010 |
| Sewer drains                  | 0.115*** | 0.029** | 0.090*** | 0.123*** | 0.107*** | 0.036** | 0.237** | 0.133*** |
| Street food stands            | 0.11*** | 0.052** | 0.028** | 0.125*** | 0.109*** | 0.278** | 0.034** | 0.128** |
| Overcrowding                  | −0.04*** | 0.002 | −0.074*** | −0.050*** | −0.043*** | −0.002 | −0.019** | −0.056*** |

Figure 2. Clusters of bivariate LISA between GDI and the analyzed factors for the 2010–2014 period. (a) census blocks with significant positive/positive association of GDI with flooding (navy blue) and homes without toilet facilities (yellow); the association overlap of both variables is represented in green; (b) census blocks with significant associations of GDI with infrastructure factors, particularly homes without water supply (orange) and homes without sewage (purple).

infrastructure. While an overall reduction in cases was observed during the analyzed periods (2007–2009 and 2010–2014), the positive association between GDI and socio-hydrological risk remains strong through time. Given our regional approach to the analyses, remarks and conclusions are only valid for the highlands and lowlands, and not at the level of individual census blocks.

The set of univariate and bivariate global spatial autocorrelation analyses conducted indicates a high diversity in the factors that statistically explain the heterogeneity in the spatial distribution of GDI between the center of the city and its periphery. In the city center, the lowlands, GDI strongly correlates with areas highly vulnerable to flooding. In the periphery, the highlands, more incidences were associated with lack of sanitation, infrastructure deficiencies and low income. While the observed relationship between GDI and flooding in the center of Mexico City is strong and lasting, the mechanisms behind this pattern are hard to determine. However, some hypotheses can be advanced. The area of high positive association has historically been the most prone to flooding. Yet it has been recognized as one of the cores of the city’s wealth (Romero Lankao 2010a), with only few areas with poor residents. Moreover, since pre-Columbian times, elaborate infrastructure and engineering investments have been made in the city center as the main means of reducing the risk of large scale floods (Ezcurra et al 1999, Tellman et al 2018). However, the aging of
Table 2. Regression model. Relative contribution of each analyzed factor in explaining the variance of GDI per analyzed period (2007–2009; 2010–2014) and city region (lowlands and highlands). The asterisks show the level of significance of factors in the model: ***, 99%; **, 95%; *, 90%. Note: given that the flooding database corresponds to probability of flooding, the correlation coefficients reported are divided by 10 to facilitate interpretation.

| Factor                        | Lowlands | Highlands | 2007–2009                  | Lowlands | Highlands | 2010–2014                  |
|-------------------------------|----------|-----------|-----------------------------|----------|-----------|-----------------------------|
|                               | All      | Parasitic | Bacterial | Undetermined | All      | Parasitic | Bacterial | Undetermined | All      | Parasitic | Bacterial | Undetermined |
| Intercept                     | −0.726   | 0.065***  | 0.003     | −0.704***     | 0.015    | 0.065     | 0.009     | 0.015     | 0.234*** | 0.002*** | −0.059    | 0.032     |
| Flooding                      | 21.043*  | 0.122     | 20.030**  | 1.75          | −1.088** | 15.399    | −0.063*   | 0.65      | 37.604***| 3.500**  | 0.163     | 0.001     |
| Homes with dirt floor         | −1.074   |           |           | −0.087        | 0.004    |           |           |           | −5.249   |           |           | −0.092    |
| Homes without water supply    |           |           |           |              |          |           |           |           | 3.183    |           |           | 0.662***  |
| Homes without toilet facilities| −0.899   |           | −0.869*   | −0.002        | 0.036    | 0.002*    | 0.001     |           | −1.247** |           | −1.237*   | 0.061**   |
| % people < 18 years old       | 0.069**  | 0.001*    | 0.066**   | −0.006        | 0.456*** | 0.003*    |          |           |          |           |           | 0.001     |
| Level of education            | 0.044    | 0.002     | 0.036     |              |          |           |           |           |          |           |           |           |
| Income Index                  | 0.706**  | 0.044**   | 0.456***  |              |          |           |           |           |          |           |           |           |
| Sewer drains                  | 0.004*   | 0.003*    |           |              |          |           |           |           |          |           |           |           |
| Street food stands            | 0.004*   | 0.318*    | 0.001     |              |          |           |           |           |          |           |           |           |
| Overcrowding                  | 0.326    | 0.318.    | 0.016     |              |          |           |           |           |          |           |           |           |
| rho                           | 0.587*** | 0.443***  | 0.529***  | 0.599***     | 0.445*** | 0.210***  | 0.169***  | 0.542***  |
infrastructure together with the growth of the city and its population over the last 50 years have increased the demand on the system to not only discharge storm water but also to cope with increased waste water. The city is constantly struggling with these problems of capacity that continue to plague the sewage system’s effectiveness (Romero Lankao 2010a). In fact, data shows that the absence of infrastructure in the city center (the lowlands) is not a problem, but other aspects such as failures in the infrastructure are (figure S4). The periphery of the city (the highlands), in contrast, presents deficits in water infrastructure (e.g. sewage coverage) that are well documented (figure S4). The peripheral areas correspond, in general, to neighborhoods with high levels of economic informality (Gilbert and De Jong 2015, Romero Lankao 2010b), which underscores the linkages between fast-growing, irregular and illegal urban settlements and health (Khan 2012). In these areas with fast growth and high informality, the lack of services compel people to find strategies to locally manage the distribution, storage and use of water (Eakin et al 2016). Household water management introduces many opportunities for contamination, thus taxing households not only with the costs of water storage and delivery, but also with the burden of disease (Cohen et al 2008).

The results of the LISA test and the regression analyses indicate that there is strong spatial association in the center of Mexico City between flooding and GDI. In other words, the incidence in one location increases significantly when the location is surrounded by high incidence neighborhoods. This is known as the ‘neighborhood effect’ (Witten et al 2017). According to (Ellen et al 2001, Ellen and Turner 1997) neighborhood conditions can influence health outcomes in the short and long term; in the short term, by influencing the behavior and attitude of the residents in the surrounding areas, and in the long term, through the reinforced and accumulated effect of the external conditions (e.g. infrastructure) that shape neighborhood heterogeneity. Interpreting our results under this view, it can be suggested that the latter factors (adverse environmental conditions associated with infrastructure), are influencing higher incidence in the center. Two pieces of evidence can be used to support this idea: first, the strong spatial autocorrelation of flooding event probability, and second, the fact that other socioeconomic factors were mostly non-significant in the results of the regression model. Nonetheless, the effect of behavior cannot be underestimated, as we found a strong correlation between GDI and the number of street food stands, which can be considered an indicator of a common daily consumption behavior of Mexico City residents. Further studies should focus in explicitly testing these hypotheses, as causality cannot be claimed based only in correlative results.

The spatial heterogeneity of GDI in Mexico City is similar to patterns observed in other fast growing urban areas of the developing world. For instance, Reiner and collaborators found that average incidences of cholera in the city center of Dhaka were two times higher than in the periphery (Reiner et al 2012). Similar patterns were observed in the same city for rotavirus infections (Martinez et al 2016). In Ahmedabad, India, Santos and collaborators found strong heterogeneity in malaria incidence related to socio-economic differences, with higher risk levels in the city center than in wards far from the historic center (Santos-Vega et al 2016b). While the pathogenic agent and therefore the transmission mechanisms of these diseases differ, all the studies point to the management of socio-hydrological risk factors and/or to differences in socio-economic conditions and densification patterns of core urban areas as explanatory factors. Moreover, it is important to highlight the fact that half of the explanatory variables (three out of six; see table 2) for the highlands change between the two analyzed periods (2007–2009 and 2010–2014). This heterogeneity of the explanatory variables for disease incidence could be due to higher heterogeneity of socio-hydrological risk factors in the highlands than in the lowlands. In fact, The F-test results in table 3 demonstrate a larger variance in the highlands than in the lowlands. This is consistent with what could be expected in terms of spatial heterogeneity in the highlands. On this basis, we may argue that such variance is, in fact, the result of socio-ecological processes that differ between lowlands and highlands. Nonetheless, we cannot rule out the possibility that our observations might be (at least in some part) a consequence of insufficient control of some of the relevant socio-hydrological variables affecting gastrointestinal disease incidence.

The significant association of flooding to GDI in the city center and of lack of water infrastructure with GDI in the urban periphery (highlands) underscores the need to consider the dynamic interaction between water management at the regional scale and control of waterborne diseases in public health planning. Cities are likely to be more vulnerable to hydrological hazards under climate change, as they simultaneously grow in area and population (Cutter et al 2012). Population growth in the periphery is increasing the demand for

| Factor                        | 2007–2009 | 2010–2014 |
|-------------------------------|-----------|-----------|
| All the pathogens             | 0.016***  | 0.013***  |
| Homes with dirt floor         | 1.0153    | 0.057***  |
| Homes without water supply    | 0.258***  | 0.241***  |
| Homes without toilet facilities| 0.035***  | 0.005***  |
| Homes without sewage          | 0.125***  | 0.061***  |
| % people < 18 years old       | 0.033***  | 0.063***  |
| Level of education            | 0.024***  | 0.035***  |
| Income Index                  | 0.009***  | 0.013***  |
| Sewer drains                  | 81.693    | 81.693    |
| Street food stands            | 12.337    | 2.4319    |
| Overcrowding                  | 1.0024    | 1.1759    |
drinking water, and is in turn diminishing the capacity of the sewage system. Moreover, an increase in extreme rainfall over time has been associated with the process of urbanization and with the drainage of Mexico City’s lakes (Benson-Lira et al. 2016). And as urban growth continues in the city and its surrounding regions, the quantity of runoff into the sewage system is expected to increase, thus overwhelming a system already operating over its capacity, hampered by weathering and subsidence (Romero Lankao 2010b). If policy and health programs are to be improved under such scenarios, there will be an increasing need to work with dynamic disease transmission models that couple social factors and ecological determinants of disease, as well as with traditional epidemiological models that determine the causal paths of exposure.

More studies need to be run which explicitly test for the causes of the observed spatial patterns of GDI, including routes of exposure to waterborne pathogens. However, our results underscore the need for a more integrated governance of water resource infrastructure and public health. Today little anticipatory dialogue exists between sector agencies in relation to proactive risk management, yet they operate in similar spatial domains and consider similar socio-economic factors to assess risk and vulnerability. The agencies that govern and operate water resources and infrastructure rarely consider health factors when defining actions and plans. Public health institutions, on the other hand, tend to be more interested in looking at infrastructure deficiencies and failures as foci of exposure to waterborne pathogens, but the lack of information and communication between sectors can lead to overlooking biophysical and infrastructural conditions that influence disease burden in urban areas (Parkes and Horwitz 2009). While it is unrealistic to expect any sector agency to expand its mandate to address the issues and demands of other agencies, coordinated action could produce synergistic outcomes of benefit for multiple sector objectives. Our research illustrates the spatial association between flood risk location and disease burden. While these relationships are often subject to speculation, our analysis of all the available data at the city scale demonstrates that the spatial correlations exist, and that the factors contributing to disease need to be investigated in order to definitively identify routes of exposure and behavioral aspects that contribute to increased risk of GDI. As we come closer to confronting the challenges of the Anthropocene, improved governance will be needed in order to address the complex interactions between sector dynamics and megacity hazards.

Acknowledgments

We would like to thank the National Institute of Statistics and Geography (INEGI), the Secretary of Health of Mexico City and the Mexico City water operator (SACMEX) for the data provided. Also, we want to thank Bertha Hernandez, and Sergio Bourguet for their technical assistance and Esmeralda Osejo Brito for professional English grammar copy editing. This research was funded by the National Science Foundation (Grant No. 1414052) and the Inter-American Institute for Global Change Research (Project number: CRN3108).

Appendix

A1. Spatial dependence

The concept of spatial dependence, or spatial autocorrelation, has its roots in Tobler’s First Law of Geography, which states that ‘everything is related to everything else, but near things are more related than distant things’ (Tobler 1979). In this sense, geographic units that are closer in space must have the propensity to influence each other, and must consequently have similar attributes. In other words, spatial dependence implies that the values of a certain variable X are determined by the values of the same variable in other locations nearby. This relationship can be formally expressed as

$$\text{cov}(x_i, x_j) = E(x_i, x_j) - E(x_i) \cdot E(x_j) \neq 0,$$

where $i \neq j$.

where $i = 1, \ldots, N$ and $j = 1, \ldots, N$, represent the two locations, while $x_i$ and $x_j$ are the values of the variable of interest location $i$ and $j$ respectively (Anselin 1988). $\text{cov}(x_i, x_j)$ can take positive and negative values. Positive values indicate that high values of a variable in $i$ are associated with high values in $j$. Finally, when the values of a particular variable in two locations are distributed randomly it is considered a non-dependence.

A2. Spatial autocorrelation index

To calculate global spatial dependence we use the Moran’s I ($I$), which is a non-parametric index of spatial dependence across the entire landscape. This index uses the local covariance weighted by the elements of the contiguity matrix, $w_{ij}$. The mathematical definition of Moran’s I is found in the following equation (Anselin 1995, Anselin 1996):

$$I = \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum (x_i - \bar{x})^2}$$

where $n$ is the number of spatial locations, $w_{ij}$ are the elements of the contiguity matrix (see A.2.1), and $(x_i - \bar{x})$ and $(x_j - \bar{x})$ are the normalized values of variable X in location $i$ and $j$.

The index is calculated under the null hypothesis of no spatial dependence, and assuming normal distribution, but it is important to remark that this is only an approximation, since the empirical distribution of this statistic is unknown.
A2.1. Contiguity matrix
The contiguity matrix, \( W \), also known as the spatial weight index, is a binary matrix that identifies the area of influence of each spatial unit. Thus, when two spatial units, \( i \) and \( j \), are considered neighbors, the value of cell \( w_{ij} = 1 \), and zero in other cases (Anselin 1996, Anselin 2001). The way in which the spatial weights are assigned depends on the type of contiguity that is used. For this work we use a Queen contiguity, in which its units are considered neighbors if they share boundaries and corners.

\[
I_{X,Y} = \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_j (y_j - \bar{y})^2}} \tag{A-3}
\]

where \((y_j - \bar{y})\) are the normalized values of variable \( Y \) in location \( i \).

The values of \( I_{X,Y} \) must be interpreted similarly to those obtained by the univariate Moran Index. That is, positive values reflect a positive relationship between variables \( X \) and \( Y \).

A3. Bivariate moran spatial autocorrelation statistic
The bivariate Moran quantifies the global spatial dependence between two variables, and it is calculated based on the multivariable spatial correlation (Wartenberg 1985):

\[
Y_i = \rho W Y_i + \beta X_i + \epsilon_i \tag{A-5}
\]

was used to incorporate the full set of predictors and the spatial dependence observed in the data. \( Y_i \) is an \( N \times 1 \) vector of observations of the dependent variable, with one observation for every unit in the sample, \( X_i \) is a \( N \times K \) matrix of exogenous explanatory variables, \( \beta \) is a \( K \times 1 \) vector with unknown parameters to be estimated, \( \epsilon_i \) and is a \( N \times 1 \) vector of disturbance terms, where \( \epsilon_i \) is taken to be independent and identically distributed for all \( i \), with zero mean and variance \( \sigma^2 \). In order to capture the spatial dependence observed in the incidence data, the model incorporates an additional regressor in the form of a spatially lagged variable, \( W Y_i \) (Anselin 2001). This variable captures cross-section dependencies, and has a different covariance structure in each spatial unit (Anselin 1988, Anselin 2001). The term \( \rho \) is the unknown spatial lag coefficient, and \( W \) is the \( N \times N \) contiguity matrix.

It is important to point out that in the case of the spatial model (the model with the spatial lag parameter) the standard errors estimator takes the form of:

\[
\hat{\sigma} = \sqrt{\frac{\hat{\epsilon}_0 - \rho \hat{\epsilon}_0 - \rho \hat{\epsilon}_L}{N}}
\]

where \( \epsilon_0 \) is the no-spatial regression residuals term and \( \epsilon_L \) is the spatial regression residuals term; the term \( \rho \) is the unknown spatial lag coefficient.

References
Ahern J 2013 Urban landscape sustainability and resilience: the promise and challenges of integrating ecology with urban planning and design Landscape Ecol. 28 1203–12
Ahern M et al 2005 Global health impacts of floods: epidemiologic evidence Epidemiol. Rev. 27 36–46
Anselin L 2001 An Introduction to Spatial Regression Analysis in R (Urbana, Champaign: University of Illinois) p 25
Anselin L 1995 Local indicators of spatial association—LISA Geogr. Anal. 27 93–115
Anselin L 2001 Spatial econometrics A Companion to Theoretical Econometrics pp 310–30
Anselin L 1988 Spatial Econometrics: Methods and Models (Dordrecht: Springer)
Benson-Lira V et al 2016 Loss of a lake system in a megacity: the impact of urban expansion on seasonal meteorology in Mexico City J. Geophys. Res. Atmos. 121 3079–99
Cohen S A et al 2008 The SEEDs of two gastrointestinal diseases: Socioeconomic, environmental, and demographic factors related to cryptosporidiosis and giardiasis in Massachusetts Environ. Res. 108 185–91
CONEVAL (Consejo Nacional de Evaluaci ´on de la Pol ´ıtica de Desarrollo Social) 2013 Nota técnica del Rezago Social en las AGEB urbanas en México pp 1–7
Cutter S et al 2012 Managing the risks from climate extremes at the local level Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation (Cambridge: Cambridge University Press) pp 291–338
Dye C 2008 Health and urban living Science 319 766–9
Eakin H et al 2016 Adapting to risk and perpetuating poverty: household’s strategies for managing flood risk and water scarcity in Mexico City Environ. Sci. Policy 66 324–33
Eakin H et al 2017 Opinion: Urban resilience efforts must consider social and political forces Proc. Natl Acad. Sci. USA 114 186–9
Elhorst J P 2014 Spatial Econometrics From Cross-Sectional Data to Spatial Panels (Heidelberg: Springer)
Ellen I G, Mijanovich T and Dillman K-N 2001 Neighborhood effects on health: exploring the links and assessing the evidence J. Urban Aff. 23 391–408
Ellen I G and Turner M A 1997 Does neighborhood matter? Assessing recent evidence Hous. Policy Debate 8 833–66
Ezcurra E et al 1999 The Basin of Mexico: Critical Environmental Issues and Sustainability (New York: United Nations University Press)
Gilbert L and De Jong F 2015 Entanglements of periphery and informality in Mexico City Int. J. Urban Reg. Res. 39 518–32

ORCID iDs
Andrés Baeza  https://orcid.org/0000-0003-0776-951X
Ana E Escalante  https://orcid.org/0000-0001-8147-4598

IOP Publishing  Environ. Res. Lett. 13 (2018) 064016

10
Gujarati D N and Porter D 2008 Basic Econometrics (McGraw-Hill) p 922
Hunter P R 1997 Waterborne Disease: Epidemiology and Ecology (New York: Wiley) (Accessed: 28 July 2017)
INEGI 2011 Censo Nacional de Población y Vivienda 2010 (Mexico City: Instituto Nacional de Estadistica Geografía e Informatica)
IPCC 2012 Managing the risks of extreme events and disasters to advance climate change adaptation A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change ed C B Field et al (Cambridge: Cambridge University Press)
Jacobi P et al 2010 The Citizens at Risk: from Urban Sanitation to Sustainable Cities (Routledge)
Khan M H 2012 Urban health in megacities of developing countries Pub. Health Forum 20 29–30
Martinez P P et al 2016 Differential and enhanced response to climate forcing in diarrheal disease due to rotavirus across a megacity of the developing world Proc. Natl Acad. Sci. USA 113 4092–7
Reiner R C et al 2012 Highly localized sensitivity to climate forcing drives endemic cholera in a megacity Proc. Natl Acad. Sci. USA 109 2033–6
Romero Lankao P 2010a Water in Mexico City: what will climate change bring to its history of water-related hazards and vulnerabilities? Environ. Urban. 22 157–78
Romero Lankao P 2010b Water in Mexico City: what will climate change bring to its history of water-related hazards and vulnerabilities? Environ. Urban. 22 157–78
Santos-Vega M et al 2016a Population density, climate variables and poverty synergistically structure spatial risk in urban malaria in India PLoS Negl. Trop. Dis. 10 e0005155
Santos-Vega M, Martinez P P and Pascal M 2016b Climate forcing and infectious disease transmission in urban landscapes: integrating demographic and socioeconomic heterogeneity Ann. NY Acad. Sci. 1382 44–55
Sivapalan M, Savenije H H G and Bloschl G 2012 Socio-hydrology: a new science of people and water HydroL. Process. 26 1270–76
Sivapalan M, Konar M, Srinivasan V, Chhatre A, Wutich A, Scott C A, Wescot J L and Rodriguez-Iiturbe I 2014 Socio-hydrology: use-inspired water sustainability science for the Anthropocene Earth’s Futur. 2 225–30
Tellman B et al 2018 Adaptive pathways and coupled infrastructure: seven centuries of adaptation to water risk and the production of vulnerability in Mexico City Ecol. Soc. 23 1
Tobler W R 1979 Cellular geography Phil. Geogr. 20 379–86
Turner B L et al 2003 Illustrating the coupled human-environment system for vulnerability analysis: three case studies Proc. Natl Acad. Sci. USA 100 8080–5
UN-HABITAT 2017 New Urban Agenda United Nations, Department of Economic and Social Affairs, Population Division 2016 The World’s Cities in 2016—Data Booklet (ST/ESA/SER.A/392) (www.un.org/en/development/desa/population/publications/pdf/urbanization/the_worlds_cities_in_2016_data_booklet.pdf)
Vlahov D et al 2005 Perspectives on urban conditions and population health Cadernos de Saúde Pública 21 949–57
Vlahov D et al 2007 Urban as a determinant of health J. Urban Health Bull. NY Acad. Med. 84 16–26
Wartenberg D 1985 Multivariate spatial correlation: a method for exploratory Geogr. Anal. 17 263–83
Witten K et al 2017 Neighborhoods and Health International Encyclopedia of Geography: People, the Earth, Environment and Technology (Oxford: Wiley) pp 1–4
Zhou W, Pickett S T A and Cadenasso M L 2017 Shifting concepts of urban spatial heterogeneity and their implications for sustainability Landscape Ecol. 32 15–30