Adjusted, non-Euclidean cluster detection of *Vibrio parahaemolyticus* in the Chesapeake Bay, USA

Anton Kvit,[1] Benjamin Davis,[1] John Jacobs,[2] Frank C. Curriero[1]

[1]Spatial Science for Public Health Center, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; [2]Cooperative Oxford Lab, National Oceanic and Atmospheric Administration, Oxford, MD, USA

Abstract

Vibrio parahaemolyticus (*V. parahaemolyticus*) is a naturally-occurring bacterium found in estuaries, such as the Chesapeake Bay (USA), that can cause vibriosis, a food - and waterborne illness, in humans. Tracking the spatial and temporal distribution of *V. parahaemolyticus* in the Chesapeake Bay, which varies in part due to water temperature, salinity, and other environmental variables, can help identify areas and time periods of high risk. These observations can support interventions used to reduce the burden of vibriosis. Spatial and spatiotemporal clusters of high *V. parahaemolyticus* abundance were identified among surface water samples in the Chesapeake Bay between 2007 and 2010. While Euclidean distances between geographic points in spatial analyses are often used for cluster detection, non-Euclidean distances should be considered for cluster detection due to the complex nature of the Chesapeake Bay shoreline. Comparison of both methods consistently showed the non-Euclidean cluster detection providing unique and more reasonable clusters than the Euclidean approach. Residuals from univariate and multivariate models were used to identify how clusters changed after controlling for environmental variables. Most clusters tended to decrease in space, time, or significance after adjustment, suggesting these covariates contributed to the original formation of the clusters and as such are useful observation tools for vibriosis risk managers. Clusters that remained after adjustment suggest areas for further study and intervention. These findings reinforce the importance of using non-Euclidean distances when tracking the spatiotemporal variation of *V. parahaemolyticus* as well as the benefits of cluster detection methods for *V. parahaemolyticus* risk management in estuaries.

Correspondence: Frank C. Curriero, Spatial Science for Public Health Center, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Rm. E6541, Baltimore MD 21205, USA.
Tel. +1.410.614.5817 - Fax: +1.410.614.0902.
E-mail: fcurriero@jhu.edu

Acknowledgements: the authors thank DNR, VADEQ, and Old Dominion University for collecting and processing the water quality data.

Key words: *Vibrio parahaemolyticus*; Cluster detection; Chesapeake Bay; Non-Euclidean; Spatiotemporal.

Funding: AK, BD, FCC were supported by the National Institutes of Allergy and Infectious Diseases (grant no. 1R01AI123931-01A1, Frank Curriero, PI).

Received for publication: 23 April 2019.
Revision received: 22 August 2019.
Accepted for publication: 12 September 2019.

©*Copyright: the Author(s), 2019*
Licensee PAGEPress, Italy
Geospatial Health 2019; 14:783
doi:10.4081/gh.2019.783

This article is distributed under the terms of the Creative Commons Attribution Noncommercial License (CC BY-NC 4.0) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Introduction

Vibrio parahaemolyticus is a naturally-occurring bacterium found in estuaries that can cause vibriosis in humans, resulting in gastroenteritis and septicaemia (Weaver and Ehrenkranz, 1975). The bacterium is one of the leading causes of seafood-related gastroenteritis in the United States (Chen *et al.*, 2017), resulting in nearly 35,000 domestically acquired foodborne illnesses annually (Scallan *et al.*, 2011). Exposure commonly occurs from eating raw or undercooked shellfish such as oysters, which can accumulate *V. parahaemolyticus* in high concentrations due to their filter-feeding behaviour, but infection can also occur in recreational waters via open wounds (Newton *et al.*, 2012; Davis *et al.*, 2017). Therefore, the natural presence of *V. parahaemolyticus* in the Chesapeake Bay poses a serious public health risk. Oyster landings are expected to increase over time as Maryland continues to transition from public harvesting sites to oyster aquaculture (Maryland Sea Grant, 2019). This trend, coupled with evidence that *V. parahaemolyticus* abundance is likely to increase given global climate change (Muhling *et al.*, 2017), may further exacerbate the risk of vibriosis stemming from the Bay.

The distribution and abundance of *V. parahaemolyticus* varies significantly throughout the Chesapeake Bay both temporally and spatially due to the bacterium's ability to associate with aquatic organisms and surfaces (Lovell, 2017). The bacterium has rarely been isolated from water samples when surface temperatures fall below 15°C (Johnson *et al.*, 2012), and so is more common in warmer months while being absent from the water column in winter. *V. parahaemolyticus* persists in the sediment by entering a viable but nonculturable (VBNC) state during colder months,

Non-commercial use only

resuscitating and resuspending into the water column once temperatures rise (Oliver, 2010). *V. parahaemolyticus* abundance is further impacted by predatory bacteria (Richards *et al.*, 2012) and bacteriophages (Bastías *et al.*, 2010) present in the water column.

In a previous study in Chesapeake Bay, *V. parahaemolyticus* presence and abundance were positively associated with water temperature, negatively associated with water clarity, were found to require at least low levels of salinity, and to have additional complex interactions with other water quality measures such as dissolved oxygen and nutrients (Davis *et al.*, 2017). These water characteristics also vary temporally and spatially throughout the Bay, with higher salinity levels in the south near the mouth of the Bay and warmer temperatures in shallower waters. Determining instances of high *V. parahaemolyticus* concentrations in space and time is important for understanding and minimizing the risk of vibriosis and helping regulators implement effective control measures (FDA, 2015).

Spatiotemporal cluster detection is a popular approach to identify areas and time periods of elevated risk and has been used previously to identify clusters of infectious disease incidence, such as cholera (Carrel *et al.*, 2009) and tuberculosis (Touray *et al.*, 2010), as well as water contamination (Gonçalves and Alpuim, 2011). Generally, clusters are found by identifying groups of points in space and time that neighbour each other and are sufficiently different from the rest of the study area. Many spatial cluster detection methods presented in the literature rely on Euclidean distance when identifying neighbouring points, which often does not consider the geography of the study area or actual travel distance between points, and instead simply captures the shortest linear distance between them. *V. parahaemolyticus* has been found to attach to zooplankton (Kaneko and Colwell, 1975) and water column particles (Blackwell and Oliver, 2008), and is significantly spatially associated with water quality measures. Therefore, a non-Euclidean proximity measure, such as distances travelling by water only, rather than Euclidean distances that can cross both water and land, would more accurately represent the dynamics of the bacterium. This distinction is particularly important in the Chesapeake Bay due to its complex shoreline geography with many parallel tributaries (Davis and Curriero, 2019) (Figure 1). To observe the impact of this proximity metric disparity on cluster detection in marine environments, the following study compares the Euclidean and non-Euclidean distance-based spatial and spatiotemporal clusters of *V. parahaemolyticus* identified in the Chesapeake Bay between 2007 and 2010, and examines the impact of adjusting for well-known environmental determinants of the bacterium on the identified clusters.

Materials and Methods

The Chesapeake Bay is the largest estuary in the United States bordering the states of Maryland and Virginia. The Chesapeake Bay watershed includes over 150 rivers and streams, with the Potomac, James, and Susquehanna Rivers being the three largest rivers and providing the majority of fresh water to the Bay (Chesapeake Bay Program, 2019). The Bay experiences significant temporal and spatial variation in temperature, salinity, and other water quality measures given seasonal atmospheric variations coupled with the complex mixing of marine and fresh water. To assist in interpreting and communicating the results of this analysis, the Bay tidal waters were split into 19 regional aggregations of previously delineated segments (Chesapeake Bay Program, 2019) (Figure 1). Surface water samples at 148 monitoring locations throughout the Bay were collected by the Maryland Department of Natural Resources and the Virginia Department of Environmental Quality between 2007-2010 and analysed using quantitative PCR (qPCR) for genetic markers of *V. parahaemolyticus*. Samples were taken in April (spring), July (summer), and October (fall) of each year, resulting in 1,531 samples. As the focus of this work was to identify clusters of detectable *V. parahaemolyticus* rather than high abundance, each sample was assigned a binary indicator depending on whether *V. parahaemolyticus* measurements were above or below the qPCR limit of detection. Sampling methods and microbial analyses are described in more detail elsewhere (Jacobs *et al.*, 2014; Davis *et al.*, 2017).

Euclidean and non-Euclidean distances were calculated between all sample locations. Distances were restricted to surface tidal waters and thus were only calculated across two dimensions (Davis and Curriero, 2019). In order to calculate non-Euclidean distances, a polygon shapefile of the Chesapeake Bay was rasterized into 1km pixels. Cost distance analysis was used, assigning no cost between water pixel boundaries, but infinite cost between water-land boundaries. Least-cost distances were therefore the shortest distances traversed entirely over water. This procedure was performed using the gdistance package (van Etten, 2017) from the R (version 3.4.4.) statistical software (R Development Core Team, 2018) and has been previously described in more detail (Murphy *et al.*, 2015; Davis and Curriero, 2019).

Spatial and spatiotemporal clusters of *V. parahaemolyticus* presence (*i.e.* above the limit of detection) using Euclidean and non-Euclidean distances were calculated using the SaTScan (SaTScan, Information Management Services, Inc., version 9.4.4) software. SaTScan utilizes a moving window of variable size, centring at each sample location and recording the number of observed and expected *V. parahaemolyticus* detects inside the window. For Euclidean distances, the window is radial and increases by finding the closest neighbour to the original sample location, then the second closest neighbour, and so on, until reaching a predetermined maximum size. Non-Euclidean cluster detection was implemented by creating a neighbours file in the SaTScan software where neighbours were ranked by water distance and window size increased by the ranking.

The cluster detection analysis used a Bernoulli model, which implements a likelihood ratio test of the probability of samples within each potential cluster being a detect *vs* a non-detect (Kulldorff, 1997). For each window location and size, the null hypothesis is that the proportion of detects inside the window is equivalent to the proportion outside. The likelihood function is maximized, identifying the cluster that was least likely to occur by chance, and a P-value is obtained by repeating this process with Monte Carlo hypothesis testing. In this analysis, 999 Monte Carlo replications and the default SaTScan P-value calculation was used, along with circular spatial windows for the Euclidean approach (Kulldorff, 2018). Given the exploratory nature of this work, with the focus being more on changes in cluster location and shape rather than statistical significance, along with the bacterium dataset having a low proportion of detects (19.4%), α=0.1 was used for all analyses to determine whether identified clusters were statistically significant.

Euclidean and non-Euclidean spatial clusters were calculated for each season-year combination. Spatiotemporal clusters can also be identified by varying the moving window across time as

well as space. Spatiotemporal Euclidean and non-Euclidean clusters were calculated for the entire study period, with sampling seasons across the years defined as equidistant time points. Imputed *V. parahaemolyticus* values were included for the winters of 2008-2010 to avoid skipping a season between fall and spring. Winter measurements were classified as non-detects, assuming that all *V. parahaemolyticus* bacteria were in a VBNC state (Bates and Oliver, 2004). A maximum window size of 25% of total sample locations was used for all cluster detection analyses.

All spatial and spatiotemporal cluster analyses were also adjusted for water temperature, clarity (Secchi disk depth), salinity, and dissolved oxygen, each of which were recorded at the same time and place as *V. parahaemolyticus*. Additional details on these water quality parameters have been described previously (Jacobs *et al.*, 2014; Davis *et al.*, 2017). In order to identify adjusted clusters, logistic regression models were calculated in R, and the

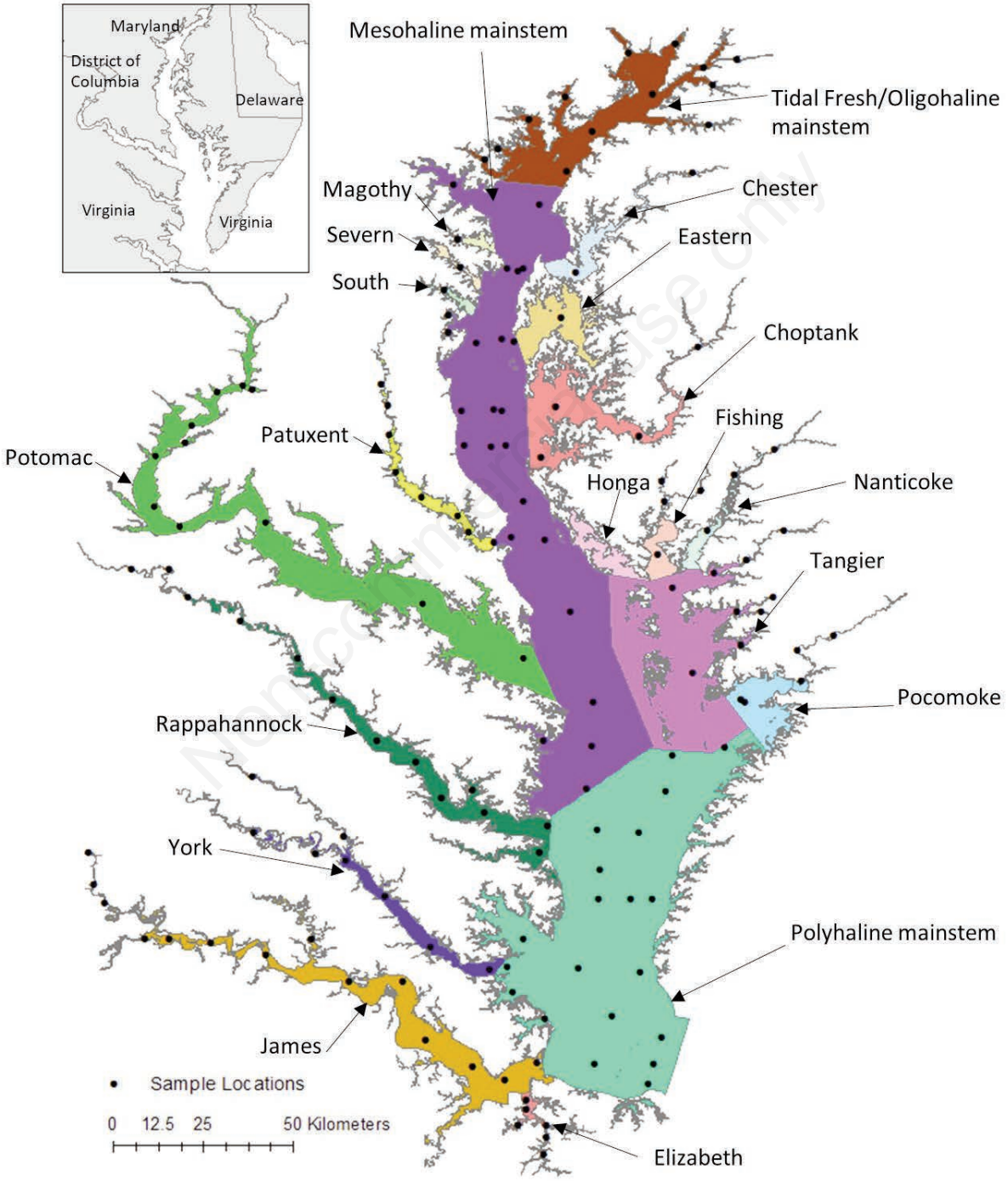

Figure 1. The Chesapeake Bay regional categories as well as sample locations used in the analysis.

Non-commercial use only

resulting deviance residuals were imported into SaTScan. Clusters of these residuals were analysed using the Gaussian (*i.e.* normal) model, which is similar to the Bernoulli model described above, but rather than comparing the probability of samples inside and outside the cluster, the average continuous deviance residuals are compared. Since water temperature is frequently associated with *V. parahaemolyticus* abundance (Johnson, 2015), and is often required for Vibrio Control Plans in shellfish harvesting waters (FDA, 2015), regressions only adjusted for temperature were also estimated. Finally, to determine whether specific clusters identified during the spring, summer, or fall seasons persisted in the same locations over multiple years, data were stratified by season, and spatiotemporal cluster detection was run with each year set as the temporal interval and with a minimum window size of 3 years. No internal SaTScan spatial, temporal, or spatiotemporal adjustments were used, since no known temporal trends or relative risks were modelled, and external covariate adjustment was imported into SaTScan, as described above.

Results

A comparison of Euclidean and Non-Euclidean water distances between these Chesapeake Bay monitoring stations has been described previously (Davis and Curriero, 2019). A total of nine statistically significant spatial Euclidean clusters were identified across the entire sampling period, compared to ten significant non-Euclidean clusters (Table 1). Note that several seasons contained no statistically significant Euclidean or non-Euclidean clusters. No clusters were identified during the winter season by design. Differences between Euclidean and non-Euclidean clusters were observed at every season-year where significant clusters were identified, though the extent of the differences varied. In three of the seasons, Euclidean clusters that were found to be significant had non-Euclidean non-significant clusters identified in similar locations (fall 2007, fall 2008, summer 2009), in four seasons the opposite was the case (fall 2007, summer 2008, spring 2010, fall 2010), and significant Euclidean and non-Euclidean clusters were found in similar locations during three other seasons (summer

Table 1. Number of statistically significant spatial Euclidean distance and non-Euclidean water distance clusters at α=0.1 for spring, summer, and fall seasons between 2007 and 2010.

Year	Season	Euclidean	Non-Euclidean
2007	Spring	0	0
	Summer	1	1
	Fall	2	1
2008	Spring	0	0
	Summer	1	3
	Fall	1	0
2009	Spring	1	1
	Summer	1	0
	Fall	0	0
2010	Spring	0	1
	Summer	2	2
	Fall	0	1

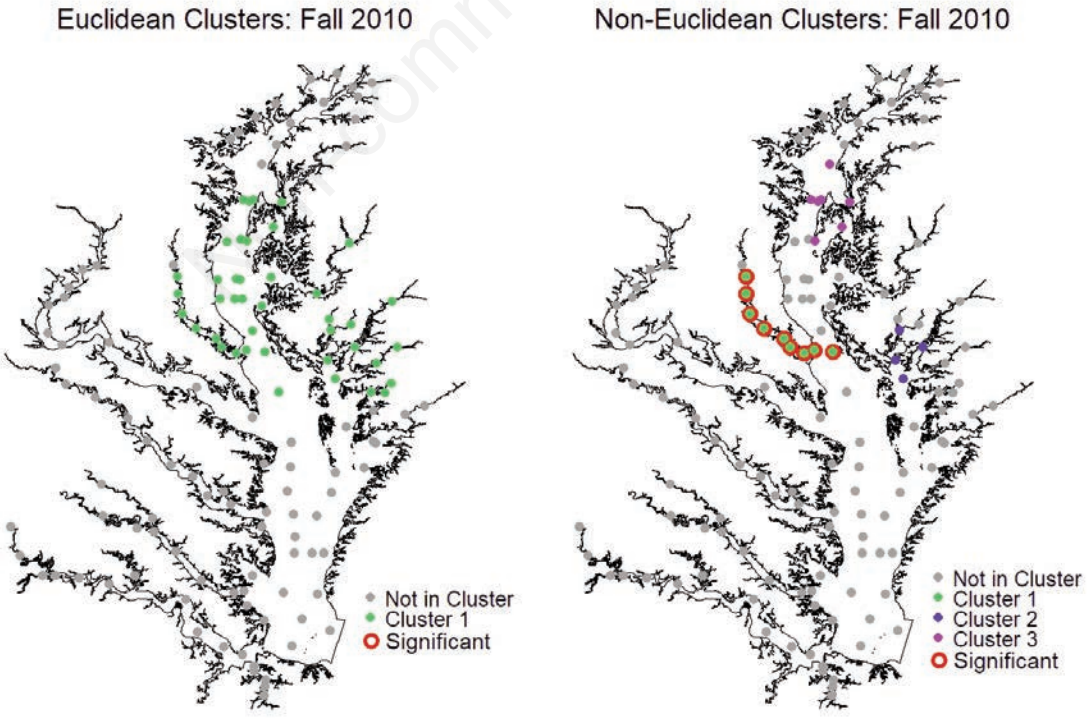

Figure 2. Euclidean distance and non-Euclidean water distance spatial clusters in the fall of 2010. The significant cluster (P<0.1) is highlighted in red.

Non-commercial use only

2007, spring 2009, summer 2010). The differences between Euclidean and non-Euclidean clusters were all consistent with the expected behaviour arising from distance measurements that included both land and water *vs* water only. For example, in the fall of 2010, the non-Euclidean method identified a significant ($P<0.05$) cluster covering all samples in the Patuxent River (Figure 2). In contrast, the Euclidean method identified a much larger, non-significant ($P=0.33$) cluster that included the Patuxent River, as well as large portions of the main stem Mesohaline region, Chester, Choptank, and Eastern Rivers, and a portion of the Tangier Sound.

The Euclidean *vs* non-Euclidean difference pattern generally persisted after adjusting for temperature alone, as well as for temperature, salinity, clarity, and dissolved oxygen. After repeating the purely spatial cluster analyses while adjusting for temperature alone, overall results indicated that spatial clusters were not substantially different from both Euclidean and non-Euclidean unadjusted clusters. While the exact shape of the clusters changed slightly in some season-years, the extent of the change was small, and no trend in the change was identified. Purely spatial Euclidean and non-Euclidean clusters that were adjusted for water temperature, salinity, clarity, and dissolved oxygen were also mostly similar to their unadjusted counterparts, though a general trend of clusters becoming less significant could be seen in several season-years. Spatiotemporal Euclidean analysis identified three significant clusters (Figure 3). Two occurred between the summer and fall of 2007, the first located in the Tangier Bay and surrounding tributaries and the other located in northern Mesohaline of the main stem as well as Magothy, Severn, and Patuxent rivers. The third cluster, located in the lower Polyhaline mainstem as well as James, York, and Elizabeth rivers occurred between the summer and fall of 2010. The unadjusted non-Euclidean approach identified two significant clusters, both occurring between the summer and fall of 2010. One was located in a similar area to the southern Euclidean cluster, while the other spanned the Mesohaline region of the main stem along with Tangier Bay and Patuxent river (Figure 3A). Adjusting for temperature as well as for multiple environmental variables did not alter the Euclidean clusters substantially, though the southernmost cluster shifted from the Polyhaline mainstem further west into the James, York, and Rappahannock rivers (Figure 3C). Adjusting for temperature split the same southernmost non-Euclidean cluster into two, located in James and York rivers (Figure 3B). Further adjustment left only the York river non-Euclidean cluster significant (Figure 3C). In contrast, the non-Euclidean cluster centred in the Mesohaline of the main stem shrunk only slightly after adjusting for covariates.

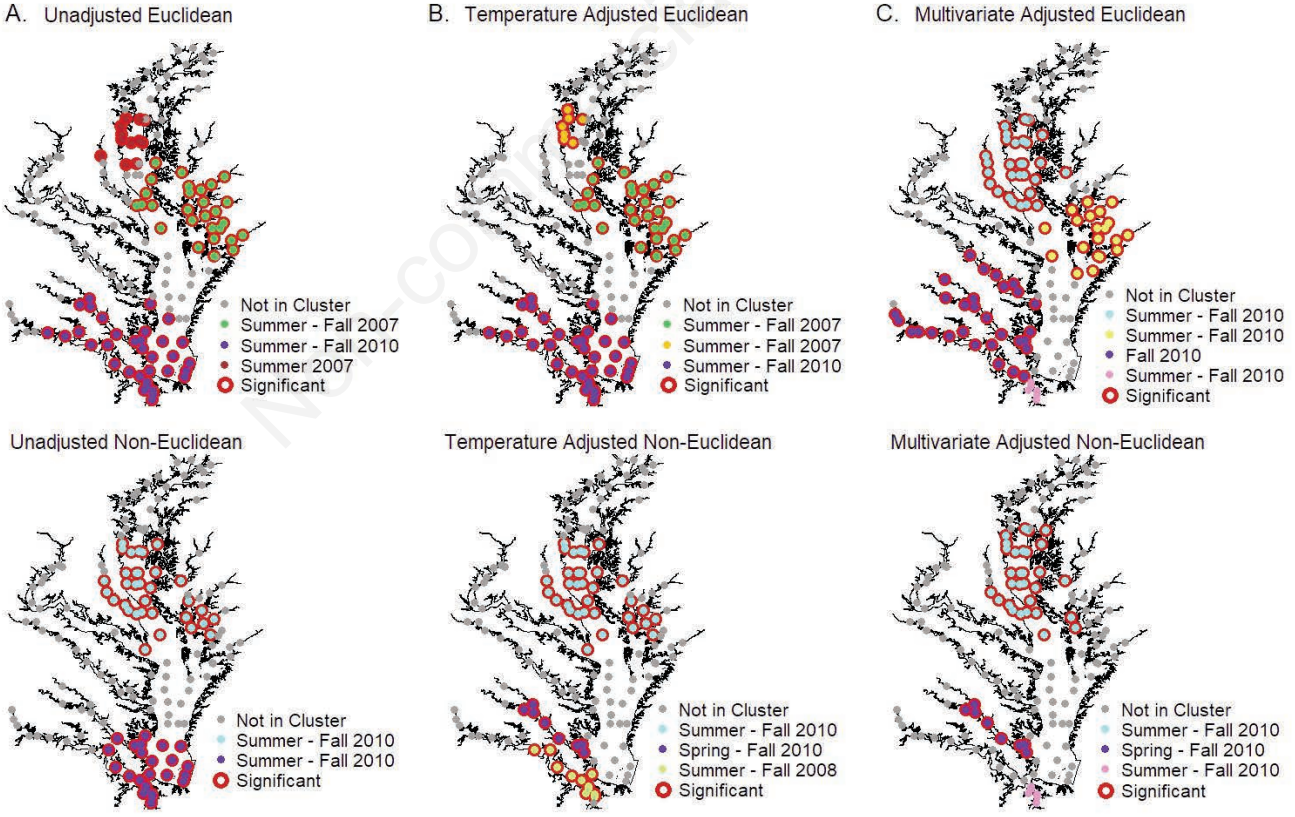

Figure 3. Spatiotemporal clusters identified between 2007 and 2010 not adjusting for any covariates, adjusting for temperature alone, and adjusting for temperature, salinity, clarity, and dissolved oxygen. Significant clusters ($P<0.1$) are highlighted in red. A) Unadjusted Euclidean and Unadjusted Non-Euclidean cluster; B) Temp Adjusted Euclidean and Temp Adjusted Non-Euclidean cluster; C) Mult Adjusted Euclidean and Mult Adjusted Non-Euclidean cluster.

Non-commercial use only

For season-specific spatiotemporal clusters, a similar trend of clusters slightly decreasing in size and significance after adjusting for temperature alone and decreasing further after adjusting for all covariates was identified. During the summer seasons, the unadjusted model identified significant non-Euclidean clusters in the James and York Rivers, as well as the Tangier Sound and Nanticoke River area (Figure 4). After adjusting for temperature, only the James River and Tangier Sound clusters remained significant. After adjusting for all the covariates, only the Tangier Sound cluster remained significant. Clusters identified in the spring and fall also showed a pattern of decreased significance after adjustment (data not shown).

In this study, unadjusted clusters identified via a Bernoulli model were compared to adjusted clusters identified via a Gaussian model. As a sensitivity analysis, residuals from intercept-only (*i.e.* no covariates) logistic regression models were also calculated and considered for unadjusted cluster detection using the Gaussian model in order to compare clusters identified via the same model. The clusters identified using the intercept-only regression residuals were almost identical to the ones identified via the Bernoulli model, supporting the approach used in this work.

Discussion

This study compared multiple spatial and spatiotemporal cluster detection analyses using Euclidean and non-Euclidean distance metrics on *V. parahaemolyticus* surface water samples in the Chesapeake Bay. Findings indicate that the difference in the metrics was noticeable, often resulting in clusters being identified in different locations or substantial changes in statistical significance. Adjusting cluster detection analyses by water quality measures presented only slight changes in identified clusters, often decreasing the size or the statistical significance of unadjusted clusters. These results overall indicate that the use of an appropriate proximity metric is necessary for cluster detection of *Vibrio* bacteria, particularly in geographically complex marine environments. Furthermore, providing adjusted clusters to risk managers who are already accounting for known environmental determinants of the *V. parahaemolyticus* will limit the size and scope of clusters that warrant additional investigation.

The observed discrepancies between Euclidean distance and non-Euclidean water distance spatial and spatiotemporal clusters are all consistent with the differences expected when measuring the shortest linear distance between sample points *vs* water distance. For instance, in the fall of 2010 the non-Euclidean approach identified a significant cluster in the relatively isolated Patuxent River, while the Euclidean approach included those samples in a much larger non-significant cluster by ignoring the peninsula that separates the River from the main stem of the Bay (Figure 2). The non-Euclidean water distance-based approach thus better captured the complex geography of the Bay's coastline and ultimately identified a significant localized cluster that was overlooked by the Euclidean approach. The identified non-Euclidean clusters are consistently more representative of how water-suspended particles, or bacteria traveling via the waterways, would behave when compared to the Euclidean clusters. While it might be possible that Euclidean cluster detection methods for *V. parahaemolyticus* are more appropriate in some settings, the suspension of the bacterium in the water column, as well as its strong association with estuarine parameters, lead one to expect that spatial variation of the bacterium to be a function of a non-Euclidean water distance.

Both Euclidean distance and non-Euclidean water distance results for purely spatial clusters did not substantially change when adjusting for water quality measures, and there was no difference in the nature of the change between the two approaches. A slight decrease in cluster significance when adjusting for all water quality measures is consistent with previous studies that found these variables to be associated with *V. parahaemolyticus* abundance (Davis *et al.*, 2017), and thus help explain the formation of the clusters. More substantial changes after adjusting for temperature

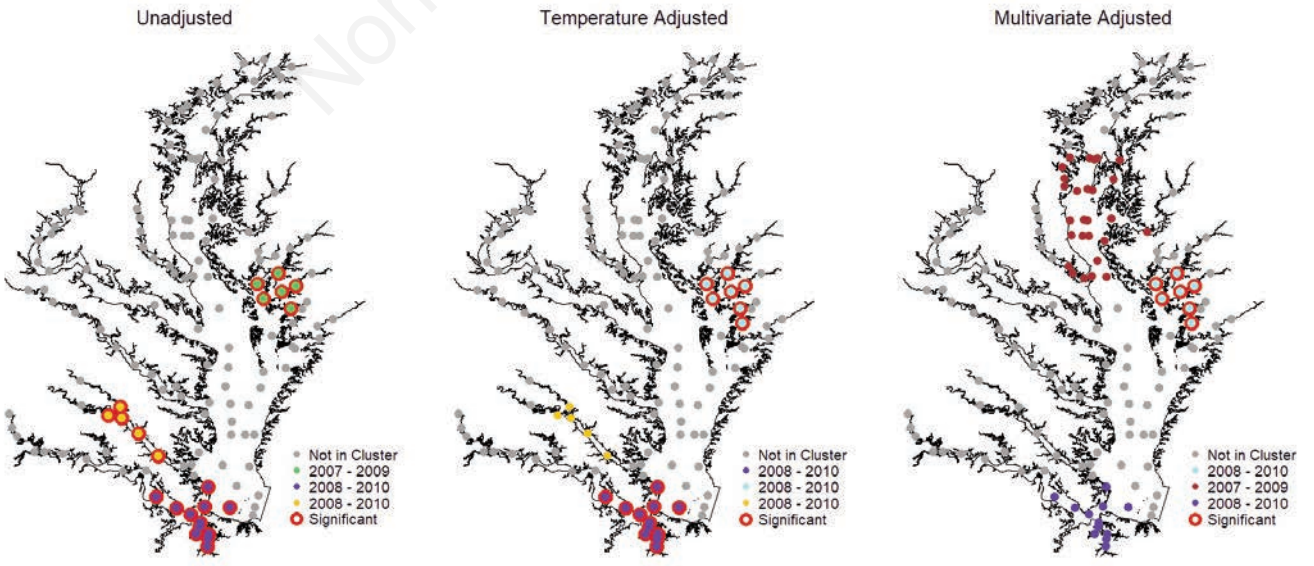

Figure 4. Spatiotemporal unadjusted, temperature adjusted, and multivariate adjusted non-Euclidean water distance clusters, identified during the summers of 2007 to 2010. Significant clusters (P<0.1) are highlighted in red.

Non-commercial use only

alone, as well as for all four of the water quality measures, were observed for the spatiotemporal clusters. This may be due in part to the larger sample sizes for the spatiotemporal analyses, as more statistical power was available to distinguish between significant and non-significant clusters. The approach of identifying clusters adjusted for known predictor variables can be useful for researchers and regulators to identify areas of high *V. parahaemolyticus* abundance that would not be expected based on previously known associations. The existence of clusters that occur during a particular season and persist in the same area year after year are also of interest since they might represent areas that require heightened attention and opportunities for intervention efforts by regulators. While the full spatiotemporal analyses (across all seasons and years) were unable to identify sustained clusters given the seasonal variation of the bacterium, such analyses could still prove useful in predicting clusters that may occur in the fall based upon clusters observed in the preceding summer.

A limitation of this work is that fixed, geostatistical sample locations were used to identify clusters, with each sample assigned a binary indicator based on the quantitative PCR limit of *V. parahaemolyticus* detection. However, the fixed sampling locations are distributed around the Chesapeake Bay fairly uniformly and thus identified clusters are not likely due to a disproportionate number of samples in a given area. The monitoring stations used in this study are also frequently measured to evaluate water quality in the Bay and therefore are useful locations to perform future cluster detections for infectious bacteria or other water quality conditions. Future applications of non-Euclidean cluster detection could also circumvent this limitation by being applied to interpolated surfaces of *V. parahaemolyticus* and related water quality measures. Efforts to use kriging in the Chesapeake Bay based upon non-Euclidean distances have already been investigated (Davis and Curriero, 2019), and future work could expand on this effort by performing space-time kriging in tidal waters. Such statistical interpolations could also be informed by mechanistic models that can account for flow and the hydrodynamics of the estuary based upon seasonal trends and morphometric differences in Bay regions.

Rather than making a binary indicator based on the qPCR limit of detection, the originally left-censored continuous values of *V. parahaemolyticus* abundance could have been used to identify clusters. Due to the very low proportion of detects with high *V. parahaemolyticus* abundance in the current dataset, the binary approach was deemed to be most appropriate for this study. However, in areas such as the Puget Sound in Washington state or the Gulf of Mexico, where high *V. parahaemolyticus* abundance samples are more prevalent, a censored regression approach should be considered.

While clusters in this analysis could be compared to each other, there is no standard to determine whether non-Euclidean water distance-based clusters of *V. parahaemolyticus* water samples are genuinely more representative and accurate in identifying areas that have increased vibriosis risks. Including information from previous vibriosis cases or outbreaks or measuring *V. parahaemolyticus* abundance in oysters could help better evaluate the public health significance of the difference between Euclidean and non-Euclidean clusters in future studies. Surveillance systems intending to identify only clusters of public health concern could also consider higher binary cutoff thresholds than the qPCR limit of detection used in this analysis, ideally based on a risk of illness that is deemed acceptable.

Conclusions

The findings of the spatiotemporal cluster detection analysis are consistent with the overall inter-annual trends of *V. parahaemolyticus* abundance in the Chesapeake Bay that have been identified previously (Davis *et al.*, 2017). Future sampling efforts for the bacterium will result in more long-term data, making spatiotemporal cluster detection methods even more useful for identifying persistently high abundance in an estuarine body of water. Future work will also apply non-Euclidean cluster detection onto interpolated surfaces informed by mechanistic models to help account for the complex spatiotemporal dynamics of *V. parahaemolyticus*, further improving the quality of detected clusters.

Given the results presented, the non-Euclidean water distance cluster detection approach should be considered for other studies that include water-based bacteria or water-suspended particles, or in other settings where locations of samples are connected by geographically constrained pathways (roads, pipes, *etc.*) More accurate identification of *V. parahaemolyticus* abundance in space and time is necessary to better understand the spatiotemporal distribution of vibriosis risk, and for reducing the burden of this illness.

References

Bastías R, Higuera G, Sierralta W, Espejo RT, 2010. A new group of cosmopolitan bacteriophages induce a carrier state in the pandemic strain of *Vibrio parahaemolyticus*. Environ Microbiol 12:990-1000.

Bates TC, Oliver JD, 2004. The viable but nonculturable state of Kanagawa positive and negative strains of *Vibrio parahaemolyticus*. J Microbiol 42:74-9.

Blackwell KD, Oliver JD, 2008. The ecology of *Vibrio vulnificus*, *Vibrio cholerae*, and *Vibrio parahaemolyticus* in North Carolina estuaries. J Microbiol 46:146-53.

Carrel M, Emch M, Streatfield PK, Yunus M, 2009. Spatio-temporal clustering of cholera: the impact of flood control in Matlab, Bangladesh, 1983-2003. Health Place 15:741-52.

Chen AJ, Hasan NA, Haley BJ, Taviani E, Tarnowski M, Brohawn K, Johnson CN, Colwell RR, Huq A, 2017. Characterization of pathogenic *Vibrio parahaemolyticus* from the Chesapeake Bay, Maryland. Front Microbiol 8:2460.

Chesapeake Bay Program, 2019. Facts and figures. Available from: https://www.chesapeakebay.net/discover/facts Accessed: March 26, 2019.

Davis BJK, Curriero FC, 2019. Development and evaluation of geostatistical methods for non-Euclidean-based spatial covariance matrices. Math Geosci 51:767-91.

Davis BJK, Jacobs JM, Davis MF, Schwab KJ, DePaola A, Curriero FC, 2017. Environmental determinants of *Vibrio parahaemolyticus* in the Chesapeake Bay. Appl Environ Microbiol 83:e01147-17.

FDA, 2015. National shellfish sanitation program guide for the control of molluscan shellfish. Food and Drug Administration, Silver Spring, MD, USA. Available from: http://www.fda.gov/Food/GuidanceRegulation/FederalStateFoodPrograms/ucm2006754.htm

Gonçalves AM, Alpuim T, 2011. Water quality monitoring using cluster analysis and linear models. Environmetrics 22:933-45.

Jacobs JM, Rhodes M, Brown CW, Hood RR, Leight A, Long W,

Non-commercial use only

Wood R, 2014. Modeling and forecasting the distribution of *Vibrio vulnificus* in Chesapeake Bay. J Appl Microbiol 117:1312-27.
Johnson CN, 2015. Influence of environmental factors on *Vibrio* spp. in coastal ecosystems. Microbiol Spectr 3.
Johnson CN, Bowers JC, Griffitt KJ, Molina V, Clostio RW, Pei S, Laws E, Paranjpye RN, Strom MS, Chen A, Hasan NA, Huq A, Noriea NF 3rd, Grimes DJ, Colwell RR, 2012. Ecology of *Vibrio parahaemolyticus* and *Vibrio vulnificus* in the coastal and estuarine waters of Louisiana, Maryland, Mississippi, and Washington (United States). Appl Environ Microbiol 78:7249-57.
Kaneko T, Colwell RR, 1975. Adsorption of *Vibrio parahaemolyticus* onto chitin and copepods. Appl Microbiol 29:269-74.
Kulldorff M, 1997. Bernoulli, discrete Poisson and continuous Poisson models: a spatial scan statistic. Commun Stat: Theor M 26:1481-96.
Kulldorff M, 2018. SaTScan™ user guide. Available from: http://www.satscan.org/
Lovell CR, 2017. Ecological fitness and virulence features of *Vibrio parahaemolyticus* in estuarine environments. Appl Microbiol Biotechnol 101:1781-94.
Maryland Sea Grant, 2019. Oyster aquaculture and restoration. Maryland Sea Grant, College Park, MD, USA. Available from: https://www.mdsg.umd.edu/topics/oysters/oyster-aquaculture-and-restoration
Muhling BA, Jacobs J, Stock CA, Gaitan CF, Saba VS, 2017. Projections of the future occurrence, distribution, and seasonality of three *Vibrio* species in the Chesapeake Bay under a high-emission climate change scenario. GeoHealth 1:278-96.
Murphy RR, Perlman E, Ball WP, Curriero FC, 2015. Water-distance-based kriging in Chesapeake Bay. J Hydrol Eng 20:05014034.
Newton A, Kendall M, Vugia DJ, Henao OL, Mahon BE, 2012. Increasing rates of vibriosis in the United States, 1996-2010: review of surveillance data from 2 systems. Clin Infect Dis 54:S391-5.
Oliver JD, 2010. Recent findings on the viable but nonculturable state in pathogenic bacteria. FEMS Microbiol Rev 34:415-25.
R Development Core Team, 2018. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Richards GP, Fay JP, Dickens KA, Parent MA, Soroka DS, Boyd EF, 2012. Predatory bacteria as natural modulators of *Vibrio parahaemolyticus* and *Vibrio vulnificus* in seawater and oysters. Appl Environ Microbiol 78:7455-66.
Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM, 2011. Foodborne illness acquired in the United States - major pathogens. Emerg Infect Dis 17:7-15.
Touray K, Adetifa IM, Jallow A, Rigby J, Jeffries D, Cheung YB, Donkor S, Adegbola RA, Hill PC, 2010. Spatial analysis of tuberculosis in an urban West African setting: is there evidence of clustering? Trop Med Int Health 15:664-72.
van Etten J, 2017. R package gdistance: distances and routes on geographical grids. J Stat Software 76:1-21.
Weaver RE, Ehrenkranz JN, 1975. *Vibrio parahaemolyticus* septicemia. Arch Int Med 135:197.

Non-commercial use only