Understanding temporal and spatial variations of viral disease in the US: The need for a one-health-based data collection and analysis approach

Evan O'Brien, Irene Xagoraraki*

Department of Civil and Environmental Engineering, Michigan State University, East Lansing, MI 48824, USA

ABSTRACT

Viral diseases exhibit spatial and temporal variation, and there are many factors that can affect their occurrence. The identification of these factors is critical in the efforts to predict and lessen viral disease burden. Because viral infection is able to spread to humans from the environment, animals, and other humans, the One-Health framework can be used to investigate the critical pathways through which viruses are transported and transmitted. A holistic approach, incorporating publicly available clinical data for human, livestock, and wildlife disease occurrence, together with environmental data reported in federal and state databases such as parameters related to land use, environmental quality, and weather, can enhance the understanding of variations in disease patterns, leading to the design and implementation of surveillance systems. An example analysis approach is presented for Michigan, United States, which is a state with large urban centers as well as a sizeable rural and agricultural population. Analysis of publicly available data from 2017 indicates that gastrointestinal (GI) and influenza-associated illnesses in Michigan may have been related with agricultural land use to a higher extent than with developed land use during that year. Meanwhile, hepatitis A virus appears to be most closely related with developed land use in dense population areas. GI illnesses may be related to precipitation, and this relationship is strongest in the springtime, although GI illnesses are most common in the winter months. Integration of human-related clinical data, animal disease data, and environmental data can ultimately be used for prioritization of the most critical locations and times for viral outbreaks in both urban and rural environments.

1. Introduction

The burden of viral disease is a global challenge, and the surveillance and reporting of viral disease is one way in which to manage and mitigate outbreaks. In the United States, the Centers for Disease Control (CDC) publish surveillance statistics regarding the rate and occurrence of disease for a number of human viruses, and annual summaries of these surveillance statistics are published in various forms. The Summary of Notifiable Diseases (SoND) is an annual report containing information on those diseases for which "regular, frequent, and timely information regarding individual cases is considered necessary for the prevention and control of the disease or condition", a list of which is updated regularly. The CDC also maintains the National Outbreak Reporting System (NORS), which includes information on the number of disease cases and outbreaks for a number of infectious agents, including certain viruses. Influenza statistics, meanwhile, are reported most frequently by the CDC via published FluView Weekly Influenza Surveillance Reports, documenting the number of cases of influenza and influenza-like illnesses in the United States. In assessing national viral disease burden, it is necessary to analyze data from all of these sources.

Fig. 1 presents the number of disease cases by month for influenza A as reported by FluView, West Nile virus and hepatitis A virus as reported by SoND, and norovirus, sapovirus, and rotavirus as reported by NORS from 2012 to 2016 [1–7]. Each of the six viruses exhibit different times of year in which disease cases are more prevalent. Insect-transmitted viruses such as West Nile virus are more common in the warmer months from July to September. Meanwhile, the waterborne viruses (norovirus, sapovirus, rotavirus, and hepatitis A virus) all exhibit different trends. Perhaps most notable is the distinction between norovirus, which is most common in the winter from January to March, and sapovirus, which is most common in autumn from September to November. Norovirus and sapovirus are closely related, both being members of the *Caliciviridae* family, yet they have strikingly different seasonal infection trends. Hepatitis A virus, on the other hand, does not show significant variation throughout the year. Rather, rates of infection are relatively constant from one month to the next.

In addition to temporal variations, virus outbreaks also exhibit spatial variations, with certain areas being more commonly affected than others. The aforementioned CDC sources also publish information regarding the disease cases for each individual state. Fig. 2 presents heatmaps of disease cases relative to state population for the six viruses mentioned above. West Nile virus appears to be more prevalent in the plains states of the central United States, while norovirus is most common in the Midwest and New England. Moreover, there is no significant spatial differentiation for hepatitis A virus from one region to

* Corresponding author at: A124 Engineering Research Complex, Michigan State University, East Lansing, MI 48824, USA.
  E-mail address: xagorara@msu.edu (I. Xagoraraki).

Fig. 1. Disease cases by month as reported by SoND (West Nile virus, Hepatitis A virus) NORS (norovirus, sapovirus, rotavirus) and FluView (influenza A) for 2012–2016 [1–7]. Data summarized by the authors.

another, mimicking its temporal trends. Rotavirus and sapovirus, meanwhile, tend to be concentrated in specific states, suggesting that outbreaks are the most common drivers of occurrence of these diseases. It is important to note, however, that these statistics are only a measure of reported cases, and that the actual incidence of viral disease could be significantly higher than the reported statistics indicate. For example, the CDC estimates that the rates of hepatitis A virus are approximately twice as high as reported incidence rates indicate [8].

Viral disease data have also been collected for the State of Michigan. Viral disease has been demonstrated to impact human, animal, and environmental health within the state of Michigan. Numerous human

outbreaks due to multiple viral agents have been reported. These outbreaks include coronavirus in Lenawee County in 1966 [9], norovirus in Macomb County in 1979 [10] and in Ottawa County in 2008 [11], hepatitis A virus in Calhoun and Saginaw Counties in 1997 [12], and West Nile virus in Kent County in 2002 [13]. Michigan has also been in the midst of an outbreak of hepatitis A virus since 2016 [14]. Illustrated in these examples is both the variety of human viral diseases that have impacted the state as well as that different areas of the state are subject to outbreaks.

Numerous governmental agencies publish data regarding clinical cases of disease both spatially and temporally. In addition to the
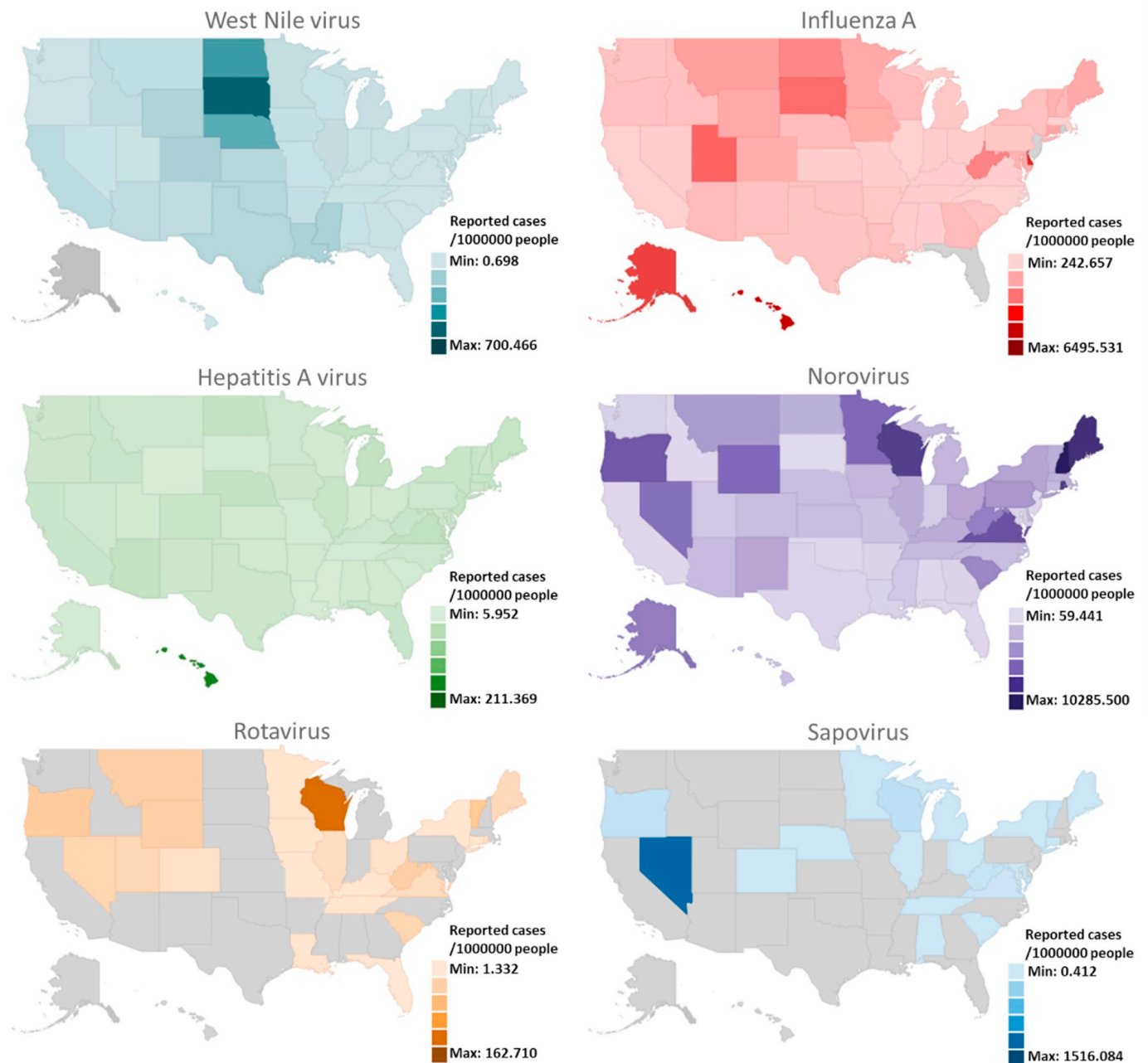
Fig. 2. Heatmaps of disease cases relative to population in the United States for 2012–2016 as reported by the CDC [1–7]. Data summarized by the authors.

published federal data, individual states publish disease surveillance statistics, such as the Michigan Department of Health and Human Services (MDHHS). MDHSS maintains the Michigan Disease Surveillance System (MDSS) which publishes weekly disease reports on a number of communicable diseases [15]. Data taken from MDSS reports show an increase in viral disease over the past five years, as shown in Fig. 3. For this paper, gastrointestinal (GI) illnesses, influenza-like illnesses, hepatitis A illnesses, and norovirus illnesses are selected. While some GI illnesses and influenza-like illnesses may be caused by bacterial pathogens, a large percentage of GI illnesses and influenza-like illnesses are expected to be of viral origin and all hepatitis and norovirus illnesses are of viral origin. These diseases have been selected for investigation since they have different exposure pathways [16]. Influenza illnesses may be zoonotic but are not waterborne. Hepatitis A illnesses are waterborne but are not zoonotic. Norovirus is commonly foodborne, it may be waterborne, and it is not typically zoonotic. GI illnesses may be both waterborne and zoonotic and may be caused by

viruses, bacteria, or paracites.

Viral disease outbreaks have also affected animals in Michigan, including viral diarrhea in cattle [17], eastern equine encephalitis virus in deer [18], and an outbreak of a novel calicivirus in rabbits [19]. According to the USDA report on death loss in U.S. cattle and calves (2015), 31.8% of non-predator cattle deaths and 42.3% of non-predator calf deaths were due to digestive or respiratory causes. These figures are amplified in the state of Michigan; the percentages are 37.8% and 66.3% respectively, equating to approximately 9027 cattle deaths and 27,926 calf deaths in the state during 2015 [20]. While the report does not specify the etiological nature of the deaths, a portion of these illnesses are due to viral causes, illustrating the potential burden of viral disease on animals. The Michigan Department of Agriculture & Rural Development (MDARD) also publishes annual statistics on reportable animal diseases. The MDARD report from 2017 includes many viral animal disease cases, including 373 cases of bovine leukemia virus, 160 cases of caprine arthritis encephalitis, 17 cases of porcine reproductive
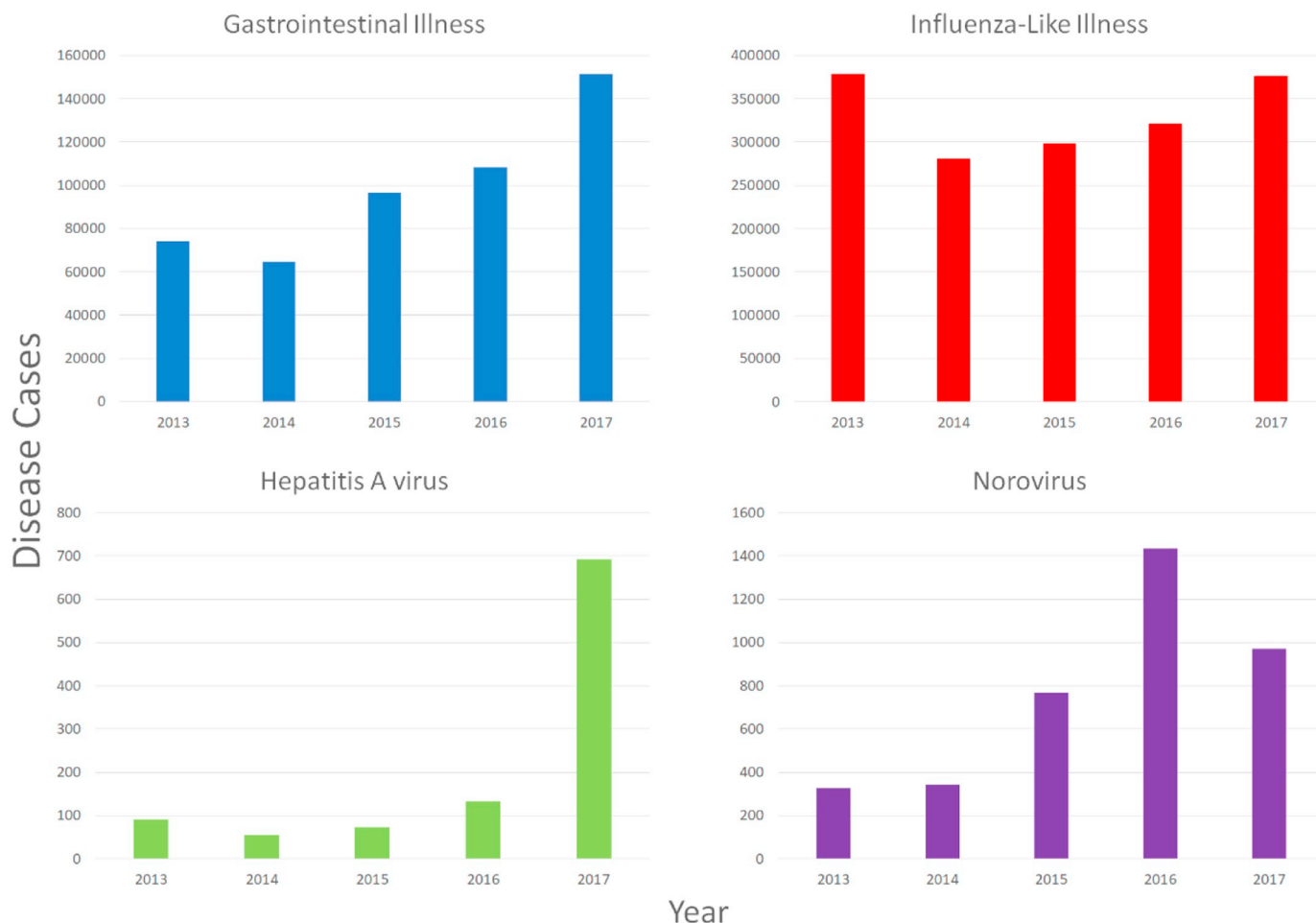
Fig. 3. Reported cases in Michigan over the past five years for GI illnesses, influenza-like illnesses, Hepatitis A virus, and norovirus as reported by MDSS [15]. Note: MDSS is a continually active system and reported numbers in the MDSS weekly reports are not final.

and respiratory syndrome virus, 7 cases of swine enteric coronavirus, 9 cases of canine influenza, 7 cases of eastern equine encephalitis, 10 cases of equine herpesvirus, and 15 cases of West Nile virus in equines [21].

Moreover, viruses have been detected in environmental samples in Michigan. Human enteric viruses have been detected in the effluent of multiple Michigan wastewater treatment plants, which is released into surrounding surface waters [25,26]. Adenovirus and other human viruses have also been detected at public recreational beaches in Michigan, leading to beach closures [27,28]. Numerous environmental factors may contribute to the likelihood of infectious disease in certain areas or time periods, including but not limited to land use [22], precipitation [23], and population density [24]. Land use is relevant to determine the environmental state of the area, and can be impactful during runoff events. Precipitation levels inform where these runoff events may occur. Population and population density can affect the spread of viral disease and can also be used to normalize disease levels from one county to another. Other factors can be used to further characterize land use, such as information related to agricultural activity. Variables such as livestock population can not only illustrate the level of agricultural activity in an area, but also illustrate the expected quality of nearby surface water after runoff events.

Because viral infection is able to spread to humans from the environment, animals, and other humans, the One-Health framework is ideal to investigate the critical pathways through which viruses are transported and transmitted [16]. Data collection related to human, animal, and environmental health is crucial to attain preliminary information for the identification of these critical pathways. This information can help to illuminate the parameters that affect the spatial and temporal patterns of disease. The goal of this study is to present an example preliminary data collection and analysis approach for the state of Michigan.

2. Methods

2.1. Data collection

Disease data was collected from weekly MDSS reports for 2017 from MDHHS, which reports the number of cases for each disease for each county for both the current week and year-to-date (note: MDSS is a continually active system and reported numbers in the MDSS weekly reports are not final) [15]. Year-to-date values were chosen as the values utilized in this data analysis as they were found to be more comprehensive compared to current week values; it is suspected this is because some cases for given weeks would not be reported until after those weeks' reports were published, thus they would only be reflected in the year-to-date values.

To adequately compare counties to one another, population data for each county was collected from U.S. Census data [29]. Population data was used to calculate the relative number of disease cases per capita for each county and each week. Weeks which contained days in more than one month were grouped into the month for which there were more days in that week (e.g., the week of 1/29–2/4 was designated as February as it contains four days in February compared to three days in January). The relative numbers of disease cases are expressed as "number of reported disease cases per 1000 people" and are considered

the dependent variables for the analysis that follows.

Land use data at the county level was collected from the United States Geological Survey (USGS) Land Cover Data Viewer [30]. Absolute land cover was collected as hectares and relative land cover was also calculated using total land area for each county. County-level agricultural data was collected from the USDA Census of Agriculture [31]. Precipitation information was collected from USDA as a 30-year average of monthly precipitation for each county in Michigan; annual values were also reported [32].

2.2. Exploratory data analysis

After data collection, exploratory data analysis was performed to investigate relationships between independent and dependent variables. Spatial distributions of variables were visualized with the creation of county-level heatmaps. Correlations were performed between variables to obtain correlation coefficients and determine which pairs of variables exhibited relationships with one another. Scatter plots were also created between independent and dependent variables to represent relationships between variables visually.

2.3. Statistical methods

Independent variables determined to have a potential correlation with disease levels were selected and utilized in the development of a preliminary statistical model. Spatial regression analysis was performed in R to assess the validity of the independent variables as predictors of the corresponding dependent variables. First, ordinary least squares (OLS) regression was performed to determine whether the collected independent variables were significantly related to the diseases studied. Independent variables were introduced into the OLS model based upon the prior exploratory data analysis; those with the highest correlations with disease levels were interpreted as the most likely predictors and were incorporated first, followed by the next highest correlation, and so on. The regression model was run each time a new variable was introduced. Those that did not exhibit a relationship with 85% confidence (i.e. p-value not < 0.15) were omitted from further consideration. This conservative level of confidence has been employed in prior studies performing spatial regression of environmental data [33]. Predictor variable collinearity was assessed using the calculation of variance inflation factor (VIF) scores; it was ensured that no predictor variable had a VIF score > 3.0 [34]. This analysis provided an initial model with which to assess the relationships between variables.

However, OLS regression does not account for spatial autocorrelation in the data, and other regression models that do account for this may be appropriate [33]. The degree of spatial autocorrelation was assessed in R and quantified with Moran's I and Lagrange multiplier diagnostics using k-nearest neighborhoods of different sizes. It was found that values of $k > 1$ provided appropriate results; a value of $k = 5$ was utilized in diagnostic tests to adequately account for spatial autocorrelation. These diagnostic tests found the existence of spatial autocorrelation in this dataset, and determined that a spatial lag model would be more appropriate. The spatial lag regression model was therefore performed in R to adjust the regression coefficients of the selected predictor variables. Akaike information criterion (AIC) values for each of the models were calculated to determine which model was of higher quality.

3. Results and discussion

3.1. Spatial and temporal distribution of viral disease in Michigan

Included in the MDSS reports are disease statistics by county for various viruses, and certain areas of the state are more commonly affected by viral disease than others. Fig. 4 shows heatmaps for cases of four diseases (GI illnesses, influenza-like illnesses, hepatitis A virus, and norovirus) for each Michigan county. Variation in spatial distribution of diseases can be observed in Michigan, with GI illnesses concentrated in the southwest portion of the state, whereas the eastern portion of the state is most affected by hepatitis A.

Because MDSS issues weekly reports on disease statistics, temporal trends can also be observed for the illnesses in question. Fig. 5 displays the number of disease cases by month for the state of Michigan in the year 2017 for GI illnesses, influenza-like illnesses, hepatitis A virus, and norovirus. GI illness and influenza norovirus are all more prevalent in the winter and spring months. Hepatitis A virus cases are more common in the latter half of the year, but there is relatively little annual variation as compared to the other diseases in question.

3.2. Spatial parameters of consideration

The primary spatial factor to consider in this case is land use. For each county in Michigan, correlations are calculated between the number of reported cases of disease (normalized to population) and the types of land use for that respective county as reported by USGS [30]. Table 1 presents the calculated correlation coefficients between these two variables.

The correlations of the diseases with agricultural vegetation presents a prominent contrast; the relationships of influenza-like illnesses and GI illnesses with agricultural land use (bolded in the table) are markedly stronger than those of hepatitis A virus and norovirus. This indicates the possibility that agricultural activity may have an impact on the transport of influenza-like illnesses and GI illnesses; the notion that agricultural land use can introduce pathogens to surrounding surface waters is supported by the literature [35–38]. This is an expected finding given that hepatitis A virus and norovirus are not thought to be zoonotic, whereas some influenza-like illnesses and GI illnesses, while uncommon, have the potential to be zoonotic. Similarly, the relationship of developed land with hepatitis A virus is much higher than with the other three diseases studied. This implies that more heavily populated areas may contribute to the incidence of hepatitis A virus. This relationship exists despite the fact that the number of disease cases for each county was normalized to that county's population, signifying that this relationship does not arise merely from a large number of reported cases in urban areas.

These relationships can be more plainly distinguished with the use of scatter plots. Fig. 6 displays scatter plots to visualize the correlations reported in Table 1 between agricultural vegetation and the four diseases investigated. A positive correlation is observable in the first two plots representing GI illness and influenza-like illness, especially when contrasted with hepatitis A virus, which shows no relationship between the two variables.

Agricultural data can assist in determining critical locations, as comparisons can also be made to agricultural trends. Fig. 7 displays heatmaps of farmland acreage, cattle population, swine population, and sheep population as reported by the USDA [31]. According to visual examination, the most commonly affected areas of viral disease appear to typically be contained within major watersheds, including the Grand River watershed for influenza-like illnesses. These illnesses also appear to correspond to areas with high cattle populations. With these observations in mind, particular attention could be paid to those factors when determining where to sample in these locations.

Fig. 8 presents heatmaps of average annual precipitation and population density for each county in Michigan as reported by the Agricultural Applied Climate Information System [32] and U.S. Census Bureau [29] respectively. Visual examination determines that precipitation levels appear highest in the western part of the state, similar to the areas most commonly affected by GI illness and influenza-like illness. Meanwhile, population density is highest near the Detroit area, which is the most area most affected by hepatitis A virus.

Table 2 presents county-level correlations between the four diseases investigated and the aforementioned variables. As is suggested by the
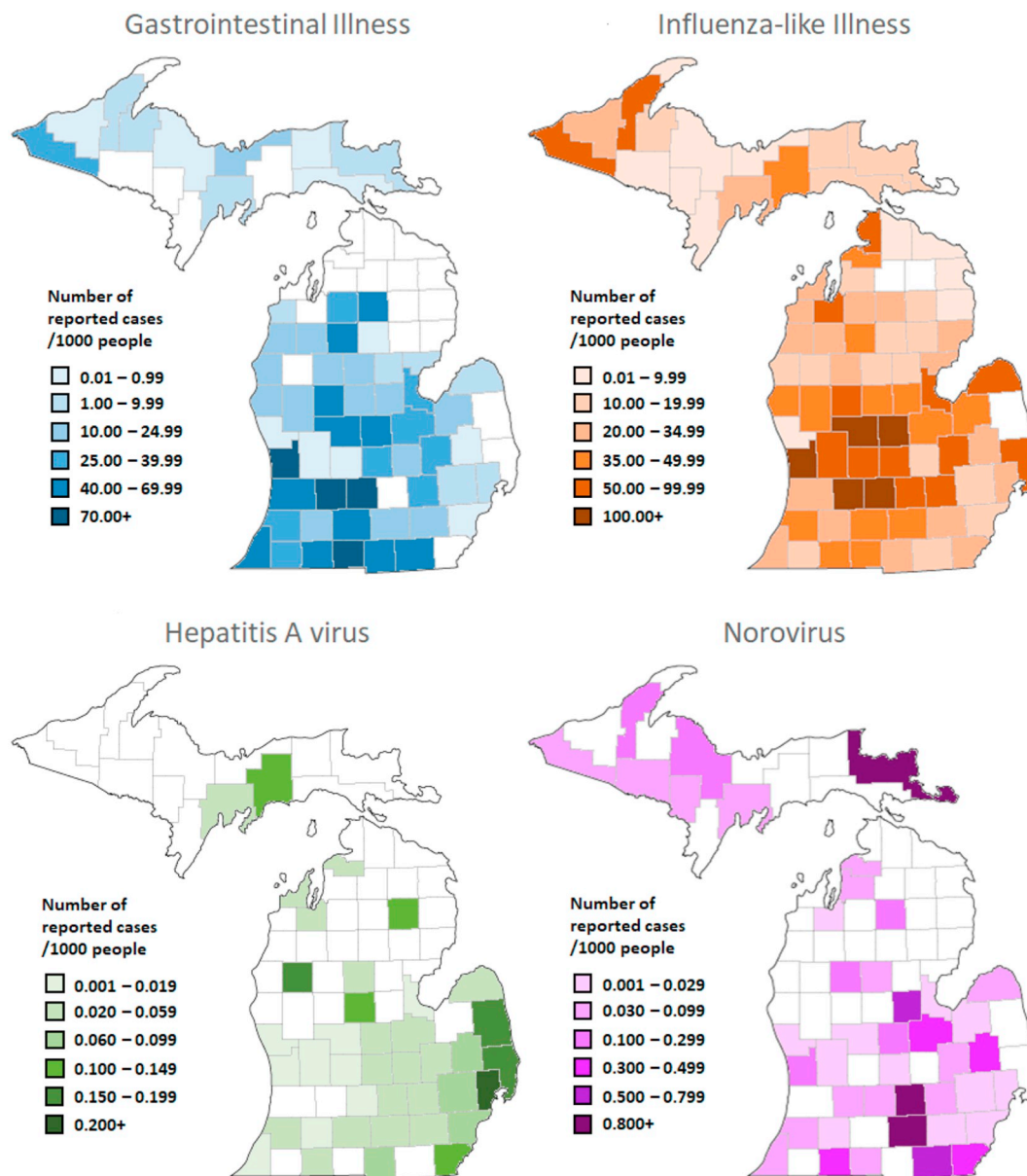
Fig. 4. Heatmaps of disease cases relative to population for Michigan counties for the year 2017 as reported by MDSS. (number of cases divided by population for county multiplied by 1000) [15]. Maps prepared by the authors. Note: MDSS is a continually active system and reported numbers in the MDSS weekly reports are not final.

heatmaps, precipitation has a high degree of correlation with GI illness, and a slight correlation with influenza-like illness, while showing no substantial relationship with the other two diseases. Meanwhile, population density has a high correlation with hepatitis A virus; this is an understandable result given the established correlation with developed land use. Livestock inventory is also seen to have stronger correlations with influenza-like illness and GI illness than with the other two diseases studied, but these correlations are not as strong with that of agricultural land use as seen in Table 1. This suggests that the potential relationship between agricultural activity and GI illness/influenza-like illness may be a result of particular agricultural practices, such as the land application of biosolids and manure as fertilizer, rather than the livestock animals themselves.

3.3. Temporal factors of consideration

The temporal variation of factors such as precipitation and surface water runoff may also help to explain viral disease occurrence. Surface water discharge, such as the flow rates of specific rivers in the state, can also provide valuable information about the status of a watershed over time. Temperature also assists in determining when runoff and first-flush events will occur. The timing of sanitary sewer overflow (SSO) and combined sewer overflow (CSO) events can give an idea of the times in which certain areas are most at-risk for pathogen exposure [39]. Similarly, comparison of the timing of manure application with the timing of runoff events can help to determine the impact of land application of biosolids on environmental water quality [40].

Comparisons can be made between the temporal distribution of disease cases and these temporal factors. One such comparison can assess the relationship between reported monthly disease cases and monthly precipitation. Accurate county-wide monthly precipitation measurements are not readily available for every county in Michigan during the year 2017, but the Agricultural Applied Climate Information System reports the 30-year average monthly precipitation levels for Michigan counties in addition to annual figs. [32]. These precipitation levels in each county can be correlated with reported diseases cases in
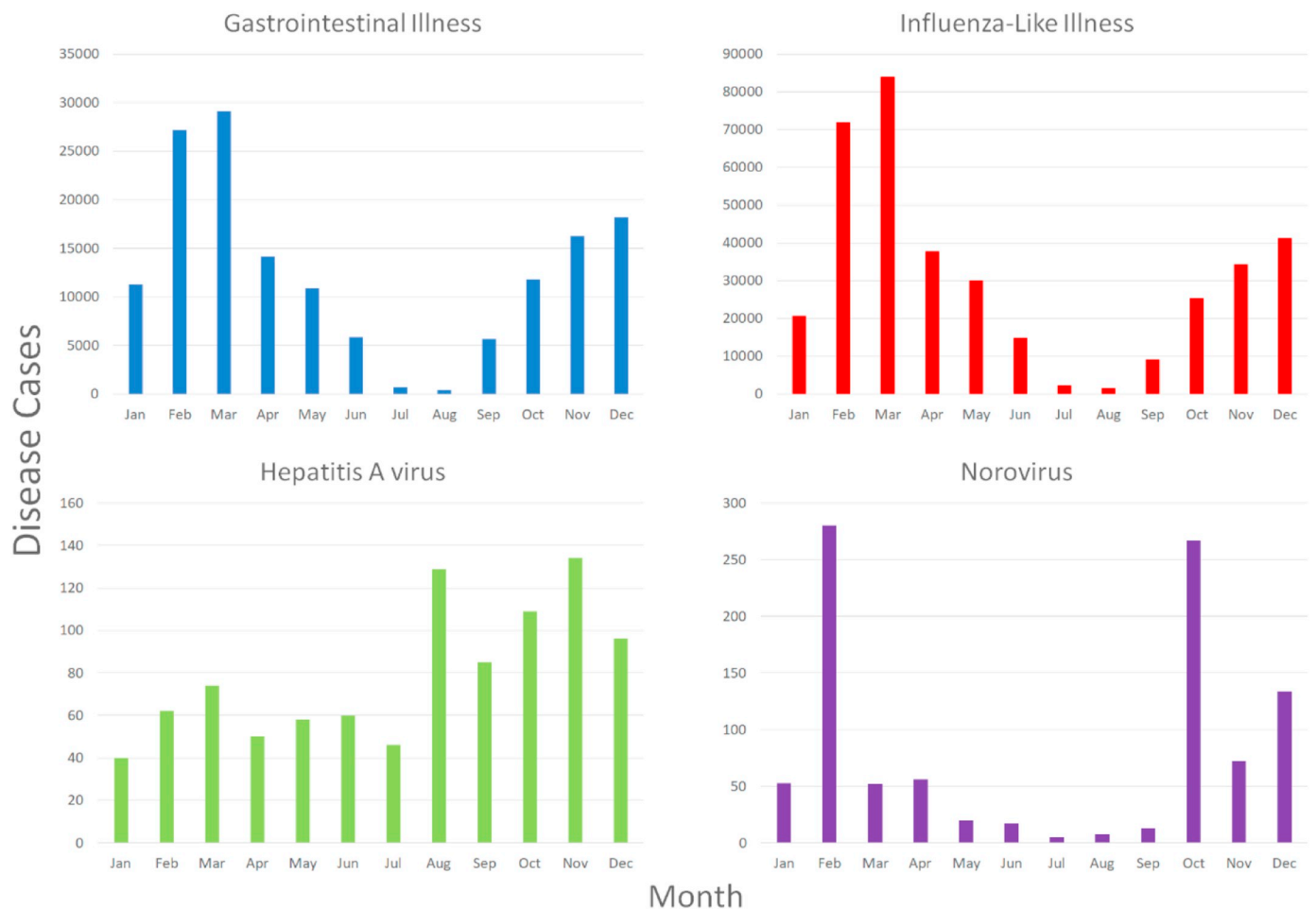
Fig. 5. Disease cases by month in Michigan for the year 2017 as reported by MDSS [15]. Note: MDSS is a continually active system and reported numbers in the MDSS weekly reports are not final.

Table 1
Correlation coefficients between disease cases normalized to population for each MI county and relative land cover for different types (hectares of type in county per total hectares in county).

Relative land cover	Disease Cases (normalized to population)			
	Influenza-Like Illness	Gastrointestinal Illness	Hepatitis A virus	Norovirus
Forest & woodland	−0.424	−0.423	−0.321	−0.146
Shrubland & grassland	−0.015	−0.057	0.217	0.270
Agricultural vegetation	**0.454**	**0.489**	0.183	0.191
Developed & other human use	0.113	0.030	**0.441**	0.007
Recently disturbed or modified	−0.218	−0.199	−0.200	−0.219
Open water	−0.266	−0.273	−0.181	−0.122

Bolded values indicate potentially significant relationships.

each county by month, taking the spatial analysis from above and introducing a more specific temporal element. A summary of these correlation coefficients is presented in Table 3.

This analysis reveals that the spatial correlations (represented by the annual figures) fluctuate at different points throughout the year. For example, as mentioned, GI illnesses have a correlation of 0.474 with annual precipitation on the spatial level, but this relationship is strongest in the month of May, when it reaches a correlation coefficient of 0.532. Moreover, the correlation coefficients between GI disease and precipitation increase in magnitude from February to May. This finding is interesting because the spring months are the times in which land application of fertilizers and manure are most common, as it is the

beginning of the growing season. This relationship with precipitation, combined with the aforementioned relationship between GI disease and agricultural land use, strengthens the possibility that agricultural runoff could be a critical pathway for GI diseases in Michigan. Stronger relationships are also observed between the months of August and November and GI illnesses as well as the months of September and November and influenza-like illnesses. These correlations indicate that these months, in addition to the aforementioned spring months, could also be critical times at which runoff is an important pathway for the studied illnesses.

Other independent variables that have not been collected could be utilized as data becomes available. In addition to spatial and temporal
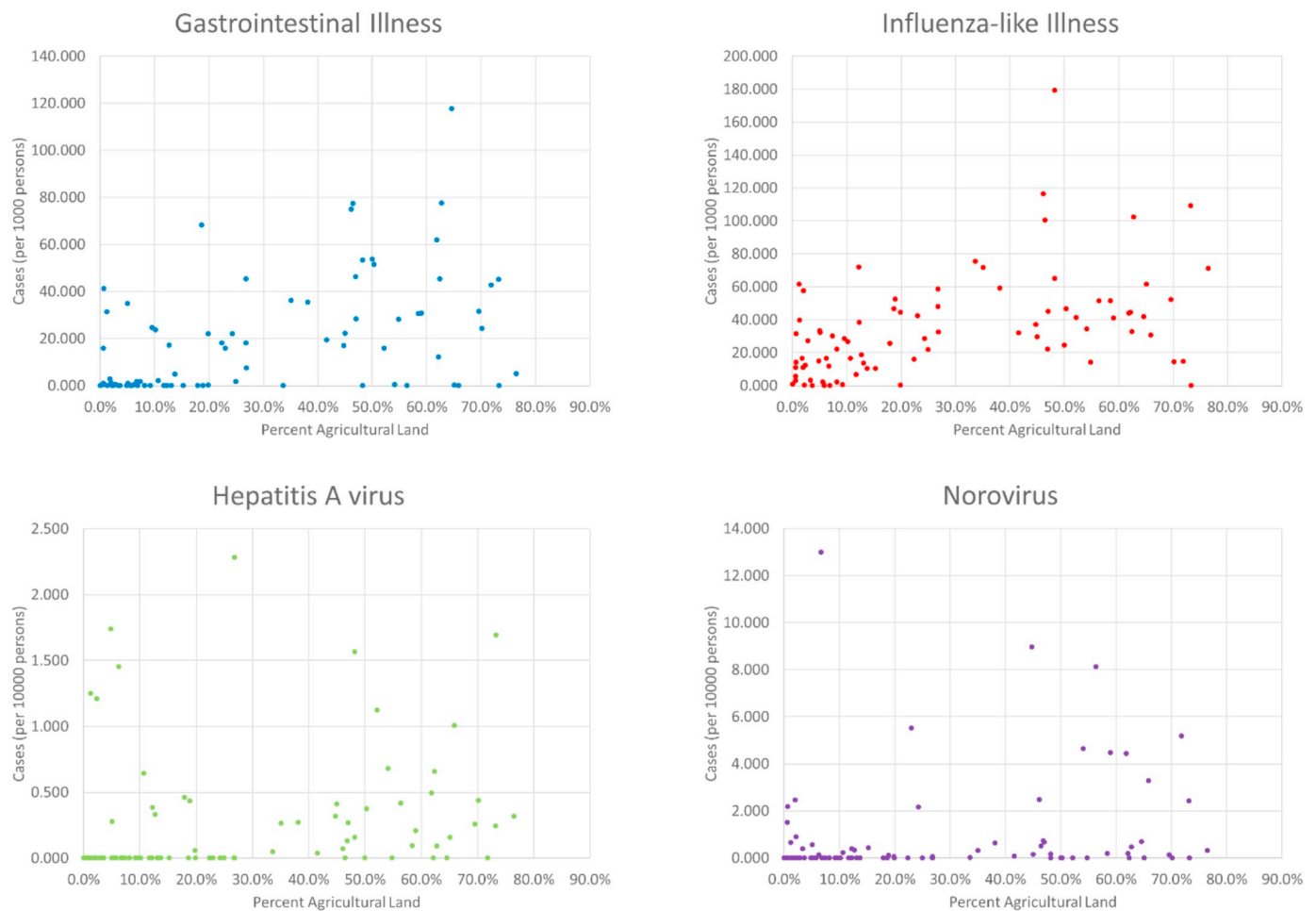
Fig. 6. Scatter plots displaying correlation between relative agricultural land cover in each county with reported disease cases (normalized to population) in each county for gastrointestinal illness, influenza-like illness, hepatitis A virus, and norovirus.

distribution of publicly available human disease data, livestock and wildlife disease data would be very useful. However, governmental agencies do not collect or provide such data to the same detail as human disease data.

3.4. Spatial regression modeling

Based on exploratory data analysis, relative agricultural land use and annual precipitation were determined to be independent variables of interest for GI illness and influenza-like illness. The OLS regression model performed showed that the initial inferences drawn from exploratory data analysis were appropriate, as all other variables (other types of land use, livestock information) did not meet the threshold of confidence (p-value not < 0.15) for further consideration. A summary of the results is listed in Table 4. As shown, both variables display a relationship with GI illness with a high degree of confidence ($p < .001$). Of the other diseases, none were related to precipitation with a high degree of confidence, and while all three meet the threshold for consideration (p-value $< .15$), influenza-like illness was found to be related to agricultural land use with a higher degree of confidence (p-value $< .001$). Additionally, the regression coefficients for agricultural land use for both hepatitis A virus and norovirus were much smaller than those for GI illness and influenza-like illness, suggesting that the relationship is not nearly as strong as with the two latter diseases.

The spatial lag model adjusted the regression coefficients for the GI model to 28.87805 for agricultural land use (P-value = .005703) and 2.22179 for precipitation (P-value = .003851). The regression coefficients for the spatial lag model are less than those for the OLS model, as the spatial lag model accounts for spatial autocorrelation, lessening the influence of the predictor variables. The AIC values for each of the models also determined that the spatial lag model was of higher quality than the OLS model, validating the use of spatial regression.

This analysis is one rudimentary example of the types of statistical techniques that can be employed to assess the relationships between collected disease data and other independent variables. Moreover, the variables assembled in these analyses is a non-exhaustive list of the potential environmental factors that can impact viral disease. As more data becomes available, more relationships of interest may be observed in exploratory data analysis, and new predictor variables could be incorporated in the above spatial regression analysis. Additionally, this regression analysis only accounts for spatial interactions between variables, and exploratory data analysis revealed that precipitation is more strongly related to GI illness in certain months of the year. Temporal data could therefore also be incorporated into future regression analyses to further pinpoint the critical times and locations for GI illness.

Additionally, disease data such as the data used in this analysis contains many zero values (counties and times at which no cases were reported). In this case, zero-inflated linear regression techniques, such as the zero-inflated Poisson model could be of use. Furthermore, more robust reporting of clinical data would be valuable in this analysis. One potential reason for the inconclusive relationships between norovirus and the investigated independent variables could be that norovirus is not as widely reported as GI illnesses or influenza-like illnesses, making it more difficult to observe correlations between variables.
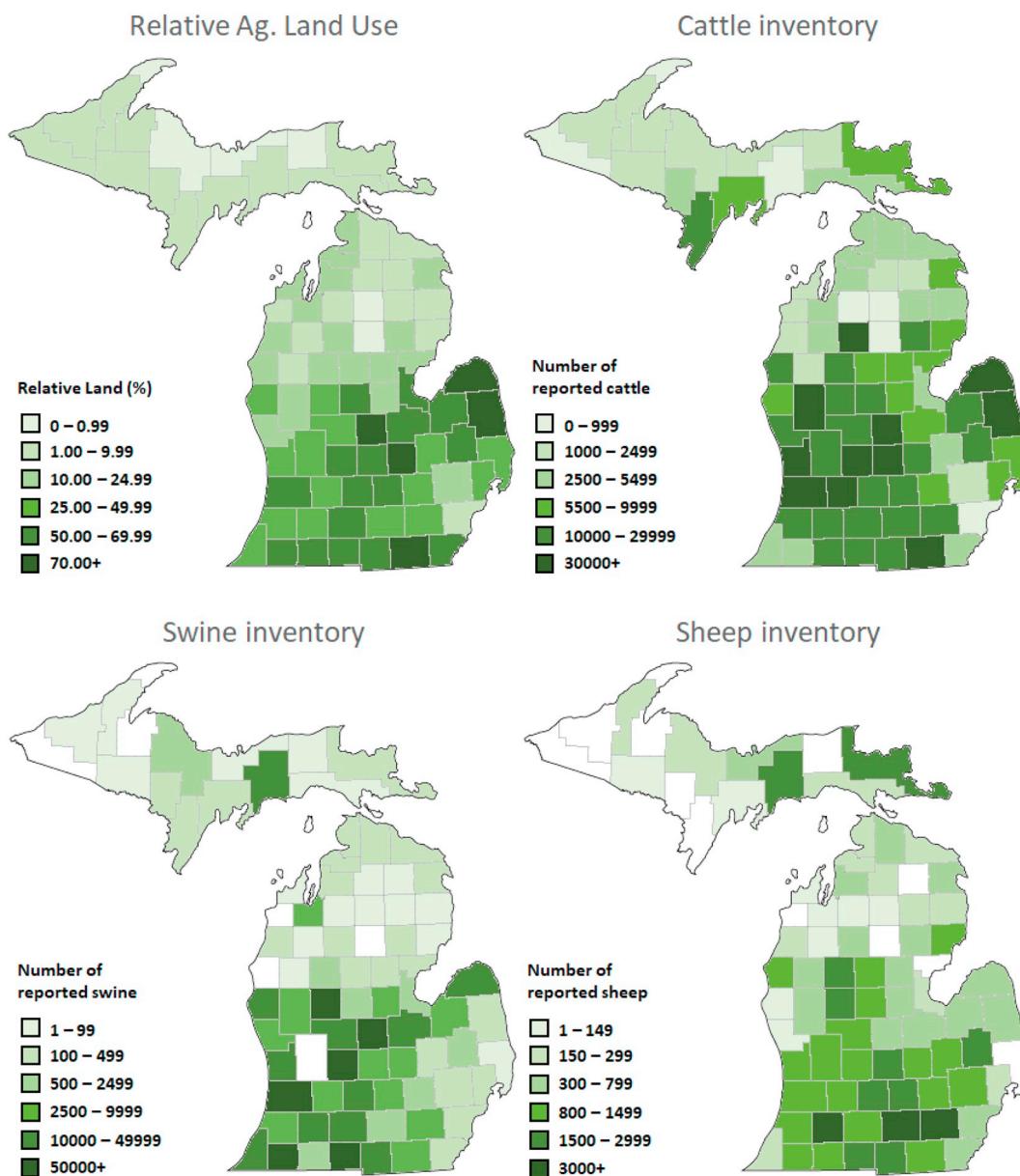
Fig. 7. Heatmap of agricultural data by county for the state of Michigan as reported by USDA (2012). Top-left: farmland acreage, top-right: cattle inventory, bottom-left: swine inventory, bottom-right: sheep inventory [31]. Maps prepared by the authors.

4. Discussion

To summarize, there are two potential findings from the above analyses. Influenza and particularly GI illnesses may be related to agricultural land use and precipitation, and this relationship with precipitation is strongest in the springtime, although GI illnesses are most common in the winter months. Meanwhile, hepatitis A virus appears to be most closely related with developed land use, and is more common in the later months of the year in autumn and winter. GI and influenza-related disease cases are observed to be relatively high in counties located in the Grand River watershed.

As the methodology used in this study relies on the reporting of disease cases, it is critical that the extent to which patients and physicians report disease cases is quantified. The analysis contained within this study is reliant on the assumption that diseases are reported at similar rates regardless of location or time of occurrence. However, this may not be the case, and reporting bias may exist. For example, certain communities may have different access to local clinics than others;

highly populated areas may be underserved by not having the necessary facilities to handle the population, or sparsely populated areas may not be within close enough proximity to accessible medical care. In future analysis this issue could be addressed by obtaining data on the number of clinics and medical facilities in each county and the number of visits to these facilities and normalizing this data to population.

In addition to the data sources utilized in this paper, there are numerous other databases that can be used to obtain relevant information to determine and predict disease variability. Environmental quality data are important factors that may be considered and results from environmental sampling and analysis may correlate with disease occurrence. For example, wastewater treatment plants and concentrated animal feeding operations (CAFOs) can be valuable sampling points, since sampling and characterizing community wastewater and livestock manure represents a snapshot of the status of community human and animal health. The Michigan Water Environment Association maintains a list of wastewater facilities in the state of Michigan [41]. Sampling can also take place at other locations, such as storm drains, agricultural
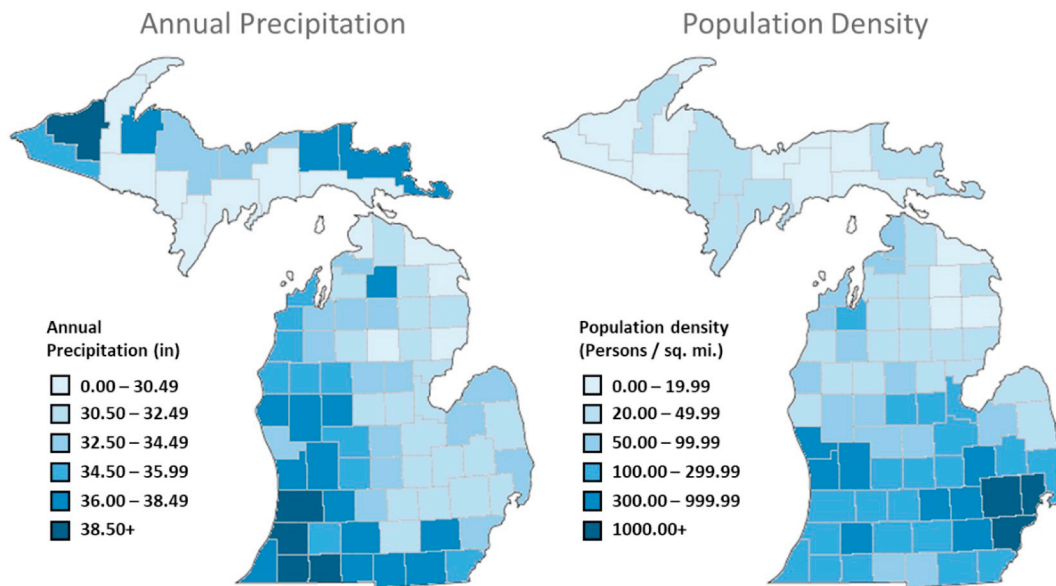
Fig. 8. Left: average annual precipitation by county for Michigan for the years 1981–2010. Right: Population density in persons per square mile by county for Michigan. Maps prepared by the authors.

Table 2
Correlation coefficients between disease cases normalized to population for each MI county and agricultural data, precipitation data, and population density for each MI county.

Parameter	Disease Cases (normalized to population)			
	Influenza-Like Illness	Gastrointestinal Illness	Hepatitis A virus	Norovirus
Cattle inventory	0.351	0.219	0.063	0.009
Hog inventory	0.136	0.359	−0.122	−0.063
Sheep inventory	0.095	0.149	0.016	0.336
Annual precipitation	0.230	**0.474**	0.041	0.024
Population density	0.050	−0.043	**0.470**	−0.001

Bolded values indicate potentially significant relationships.

Table 3
Summary of correlation coefficients for the relationship between average 30-year precipitation in the county with reported disease cases (normalized to population) in the county for each month.

Month	Influenza-Like Illness	Gastrointestinal Illness	Hepatitis A virus	Norovirus
Jan	0.031	0.084	−0.007	0.245
Feb	0.297	**0.317**	0.125	−0.064
Mar	0.125	**0.356**	0.119	0.131
Apr	0.271	**0.442**	0.164	−0.032
May	0.290	**0.532**	0.012	−0.073
Jun	−0.017	0.115	0.104	−0.173
Jul	−0.097	0.071	0.020	0.080
Aug	−0.002	**0.316**	−0.189	−0.104
Sep	0.205	0.083	0.002	0.047
Oct	−0.132	−0.034	−0.285	−0.095
Nov	**0.363**	**0.351**	−0.093	−0.041
Dec	0.198	0.210	0.138	−0.011
Annual	0.230	**0.474**	0.041	0.024

Bolded values indicate potentially significant relationships.

field runoff drains, and areas that have recently experienced combined sewer overflows. Other agricultural data could also be valuable in determining sampling points in rural areas of the watershed, such as amount of fertilizer purchased per week per county and location of

CAFOs. The Sierra Club maintains a readily available map of CAFOs throughout the United States, including in Michigan [42]. CAFO locations can help to determine where livestock populations are most abundant, heightening the risk for both animal disease and zoonotic disease. Beyond agricultural data, information on surface water quality can also be of use. For example, the Michigan Department of Environmental Quality (MDEQ) reports figures for public beach closures, which occur when surface water contamination is detected during regular screening for pathogen indicators. MDEQ also summarizes sanitary sewer overflow (SSO) and combined sewer overflow (CSO) events, which occur when wastewater levels in municipal sewer systems exceed the systems' capacity, resulting in untreated wastewater discharging into nearby surface waters.

Sampling times, meanwhile, can be determined by other factors. Using Michigan as an example, the relationship between precipitation and GI illness was strongest in the spring months, coinciding with land application of agricultural fertilizers. Additionally, these relationships were strongest in the area of the Grand River watershed, a large watershed in southwest Michigan encompassing Grand Rapids, Lansing, and surrounding agricultural areas. Therefore, sampling within this watershed in spring (from March to May) would be ideal to assess the impact of agricultural runoff on the occurrence of GI illness. This runoff would be at its peak when the flow rate of the Grand River would be highest. Therefore, examining the discharge of the Grand River can aid in determining the most critical times at which to sample. The USGS reports discharge data and water quality data for many surface waters and groundwaters in the United States, including the Grand River [44]. Animal disease data can also be collected and analyzed in this manner, but there remains a need for integrated human-animal disease surveillance to assess zoonotic disease occurrence [43].

One goal of integrative human, animal and environmental data analysis and subsequent environmental monitoring is to develop a system of prioritization for potential occurrence of disease for each county in Michigan per month. Analysis of the collected spatial and temporal data can determine the factors that are most critical in specific places and at specific times. This can lead to the determination of which locations and time periods are of the greatest concern for each disease investigated. This will, in turn, lead to higher levels of preparedness to combat viral disease outbreaks, as these critical times and locations can be surveyed before the disease develops.

Table 4
Summary of OLS regression results for each disease investigated.

Disease	Independent variable	Regression coefficient	Standard error	Test statistic	P-value
Gastrointestinal illness	Agricultural Vegetation	**35.5565**	9.665	3.679	**0.000423**
	Annual Precipitation	**2.6549**	0.7713	3.442	**0.000921**
Influenza-like illness	Agricultural Vegetation	**53.7628**	13.4585	3.995	**0.000143**
	Annual Precipitation	0.6955	1.074	0.648	0.519096
Hepatitis A virus	Agricultural Vegetation	0.046447	0.022812	2.036	0.045
	Annual Precipitation	−0.0023	0.00182	−1.262	0.211
Norovirus	Agricultural Vegetation	0.189612	0.106758	1.776	0.0795
	Annual Precipitation	−0.00365	0.00852	−0.429	0.6691

Bolded values indicate significant relationships.

5. Conclusions

The One Health framework can be readily applied to the investigation of viral disease, and determination of critical environmental factors is an important part of this process. This study shows the existence of significant relationships between clinically reported human viral infections and environmental factors such as land use and precipitation. The identification of these relationships can assist in the determination of the most critical times and locations for which humans and animals are most at risk of viral infection. Once these times and locations are determined, surveillance systems can be implemented and interventions can be introduced to mitigate potential viral outbreaks. While Michigan was used as an example in this study, these concepts are still relevant regardless of where and when this methodology is implemented. The utilization of the One Health framework in its full capacity can better help to identify, predict, and prevent viral disease outbreaks.

Declaration of Competing Interest

The authors declare that they have no competing interest or any conflict of interest in this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.onehlt.2019.100105.

References

[1] D.A. Adams, Summary of Notifiable infectious diseases and conditions — United States, 2015, MMWR Morb. Mortal. Wkly Rep. 64 (2017), https://doi.org/10.15585/mmwr.mm6453a1.

[2] D.A. Adams, R.A. Jajosky, U. Ajani, J. Kriseman, P. Sharp, D.H. Onwen, A.W. Schley, W.J. Anderson, A. Grigoryan, A.E. Aranas, Summary of notifiable diseases–United States, 2012, MMWR Morb. Mortal. Wkly. Rep. 61 (2014) 1–121.

[3] D. Adams, K. Fullerton, R. Jajosky, P. Sharp, D. Onweh, A. Schley, W. Anderson, A. Faulkner, K. Kugeler, Summary of notifiable infectious diseases and conditions—United States, 2013, MMWR Morb. Mortal. Wkly Rep. 62 (2015) 1–122.

[4] D.A. Adams, Summary of notifiable infectious diseases and conditions—United States, 2014, MMWR Morb. Mortal. Wkly Rep. 63 (2016).

[5] Annual statistics from the National Notifiable Diseases Surveillance System (NNDSS), https://wonder.cdc.gov/nndss/nndss_annual_tables_menu.asp, (2019) (accessed June 19, 2018).

[6] CDC, National Outbreak Reporting System (NORS) Dashboard, https://wwwn.cdc.gov/norsdashboard/, (2018) , Accessed date: 19 June 2018.

[7] CDC, GIS FluView, https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html, (2018) , Accessed date: 19 June 2018.

[8] CDC, U.S., Surveillance Data for Viral Hepatitis | Statistics & Surveillance | Division of Viral Hepatitis, CDC, 2014 2017 https://www.cdc.gov/hepatitis/statistics/2015surveillance/index.htm (accessed March 6, 2018).

[9] J.J. Cavallaro, A.S. Monto, Community-wide outbreak of infection with a 229E-like coronavirus in Tecumseh, Michigan, J. Infect. Dis. 122 (1970) 272–279.

[10] J.S. Koopman, E.A. Eckert, H.B. Greenberg, B.C. Strohm, R.E. Isaacson, A.S. Monto, Norwalk virus enteric illness acquired by swimming exposure, Am. J. Epidemiol. 115 (1982) 173–177, https://doi.org/10.1093/oxfordjournals.aje.a113288.

[11] C.M. Roberts, J. Archer, T. Renner, P.A. Heidel, D.L. VandeBunte, B.M. Brennan, C. Croker, R. Reporeter, S. Nakagawa-Ota, A.J. Hall, Norovirus outbreaks on three college campuses - California, Michigan, and Wisconsin 2008, Morb. Mortal. Wkly. Rep. 58 (2009) 1095–1100.

[12] Y.J.F. Hutin, V. Pool, E.H. Cramer, O.V. Nainan, J. Weth, I.T. Williams, S.T. Goldstein, K.F. Gensheimer, B.P. Bell, C.N. Shapiro, M.J. Alter, H.S. Margolis, A multistate, foodborne outbreak of hepatitis a, N. Engl. J. Med. 340 (1999) 595–602, https://doi.org/10.1056/NEJM199902253400802.

[13] M. Emig, D.J. Apple, Severe West Nile virus disease in healthy adults, Clin. Infect. Dis. 38 (2004) 289–292, https://doi.org/10.1086/380458.

[14] MDHHS, Michigan Hepatitis a Outbreak, https://www.michigan.gov/mdhhs/0,5885,7-339-71550_2955_2976_82305_82310-447907–,00.html, (2018) , Accessed date: 21 October 2018.

[15] MDHHS, Michigan Disease Surveillance System, https://www.michigan.gov/mdhhs/0,5885,7-339-71550_5104_31274—,00.html, (2018) , Accessed date: 19 June 2018.

[16] E. O'Brien, I. Xagoraraki, A water-focused one-health approach for early detection and prevention of viral outbreaks, One Health. 7 (2019) 100094, , https://doi.org/10.1016/j.onehlt.2019.100094.

[17] H. Houe, J.C. Baker, R.K. Maes, H. Wuryastuti, R. Wasito, P.L. Ruegg, J.W. Lloyd, Prevalence of cattle persistently infected with bovine viral Diarrhea virus in 20 dairy herds in two counties in Central Michigan and comparison of prevalence of antibody-positive cattle among herds with different infection and vaccination status, J. Vet. Diagn. Investig. 7 (1995) 321–326, https://doi.org/10.1177/104063879500700304.

[18] S.M. Schmitt, T.M. Cooley, S.D. Fitzgerald, S.R. Bolin, A. Lim, S.M. Schaefer, M. Kiupel, R.K. Maes, S.A. Hogle, D.J. O'Brien, An outbreak of eastern equine encephalitis virus in free-ranging white-tailed deer in Michigan, J. Wildl. Dis. 43 (2007) 635–644, https://doi.org/10.7589/0090-3558-43.4.635.

[19] I.L. Bergin, A.G. Wise, S.R. Bolin, T.P. Mullaney, M. Kiupel, R.K. Maes, Novel Calicivirus identified in rabbits, Michigan, USA, Emerg. Infect. Dis. 15 (2009) 1955–1962, https://doi.org/10.3201/eid1512.090839.

[20] USDA, Death Loss in U.S. Cattle and Calves Due to Predator and Nonpredator Causes, 2017 https://www.aphis.usda.gov/animal_health/nahms/general/downloads/cattle_calves_deathloss_2015.pdf, (2015) (accessed July 11, 2018).

[21] M. (MDARD) Sprague, Reportable diseases, (2017), (2017).

[22] N.L. Gottdenker, D.G. Streicker, C.L. Faust, C.R. Carroll, Anthropogenic land use change and infectious diseases: a review of the evidence, EcoHealth. 11 (2014) 619–632, https://doi.org/10.1007/s10393-014-0941-z.

[23] F.C. Curriero, J.A. Patz, J.B. Rose, S. Lele, The association between extreme precipitation and waterborne disease outbreaks in the United States, 1948–1994, Am. J. Public Health 91 (2001) 1194–1199, https://doi.org/10.2105/AJPH.91.8.1194.

[24] K.E. Jones, N.G. Patel, M.A. Levy, A. Storeygard, D. Balk, J.L. Gittleman, P. Daszak, Global trends in emerging infectious diseases, Nature. 451 (2008) 990–993, https://doi.org/10.1038/nature06536.

[25] F.J. Simmons, I. Xagoraraki, Release of infectious human enteric viruses by full-scale wastewater utilities, Water Res. 45 (2011) 3590–3598, https://doi.org/10.1016/j.watres.2011.04.001.

[26] E. O'Brien, M. Munir, T. Marsh, M. Heran, G. Lesage, V.V. Tarabara, I. Xagoraraki, Diversity of DNA viruses in effluents of membrane bioreactors in Traverse City, MI (USA) and La Grande Motte (France), Water Res. 111 (2017) 338–345, https://doi.org/10.1016/j.watres.2017.01.014.

[27] I. Xagoraraki, D.H.-W. Kuo, K. Wong, M. Wong, J.B. Rose, Occurrence of human adenoviruses at two recreational beaches of the Great Lakes, Appl. Environ. Microbiol. 73 (2007) 7874–7881, https://doi.org/10.1128/AEM.01239-07.

[28] A. Oun, Z. Yin, M. Munir, I. Xagoraraki, Microbial pollution characterization of water and sediment at two beaches in Saginaw Bay, Michigan, J. Gt. Lakes Res. 43 (2017) 64–72, https://doi.org/10.1016/j.jglr.2017.01.014.

[29] SFA - Revenue & Economic Data, http://www.senate.michigan.gov/sfa/RevenueData/RevenueData.html, (2018) , Accessed date: 15 November 2018.

[30] USGS, Land Cover Viewer - Map, https://gis1.usgs.gov/csas/gap/viewer/land_cover/Map.aspx, (2018) , Accessed date: 24 August 2018.

[31] USDA, NASS, Census of Agriculture - Publications, n.d. https://www.agcensus.usda.gov/Publications/2012/, (2012) (accessed June 19, 2018).

[32] Field Office Technical Guide, n.d. https://efotg.sc.egov.usda.gov/ (accessed November 15, 2018).

[33] S.L. Hession, N. Moore, A spatial regression analysis of the influence of topography on monthly rainfall in East Africa, Int. J. Climatol. 31 (2011) 1440–1456, https://doi.org/10.1002/joc.2174.

[34] A.F. Zuur, E.N. Ieno, C.S. Elphick, A protocol for data exploration to avoid common

statistical problems, Methods Ecol. Evol. 1 (2010) 3–14, https://doi.org/10.1111/j. 2041-210X.2009.00001.x.

[35] J. Crowther, D. Kay, M.D. Wyer, Faecal-indicator concentrations in waters draining lowland pastoral catchments in the UK: relationships with land use and farming practices, Water Res. 36 (2002) 1725–1734, https://doi.org/10.1016/S0043-1354(01)00394-3.

[36] S.J. Gaffield, R.L. Goo, L.A. Richards, R.J. Jackson, Public health effects of in-adequately managed Stormwater runoff, Am. J. Public Health 93 (2003) 1527–1533, https://doi.org/10.2105/AJPH.93.9.1527.

[37] J.A. Patz, S.J. Vavrus, C.K. Uejio, S.L. McLellan, Climate change and waterborne disease risk in the Great Lakes region of the U.S, Am. J. Prev. Med. 35 (2008) 451–458, https://doi.org/10.1016/j.amepre.2008.08.026.

[38] J.B. Rose, P.R. Epstein, E.K. Lipp, B.H. Sherman, S.M. Bernard, J.A. Patz, Climate variability and change in the United States: potential impacts on water- and food-borne diseases caused by microbiologic agents, Environ. Health Perspect. 109 (2001) 211–221.

[39] T.-T. Fong, M.S. Phanikumar, I. Xagoraraki, J.B. Rose, Quantitative detection of human adenoviruses in wastewater and combined sewer overflows influencing a Michigan River, Appl. Environ. Microbiol. 76 (2010) 715–723, https://doi.org/10. 1128/AEM.01316-09.

[40] J. Venglovsky, J. Martinez, I. Placha, Hygienic and ecological risks connected with utilization of animal manures and biosolids in agriculture, Livest. Sci. 102 (2006) 197–203, https://doi.org/10.1016/j.livsci.2006.03.017.

[41] Michigan Water Environment Association, Facilities List, https://www.mi-wea.org/facilities_list.php, (2018) , Accessed date: 29 November 2018.

[42] Sierra Club, CAFO Map, Sierra Club, https://www.sierraclub.org/michigan/cafo-map, (2015) (accessed November 29, 2018).

[43] W.A. Mauer, J.B. Kaneene, Integrated human-animal disease surveillance, Emerg. Infect. Dis. 11 (2005) 1490–1491, https://doi.org/10.3201/eid1109.050180.

[44] USGS, Water Data for the Nation, https://waterdata.usgs.gov/nwis, (2019) , Accessed date: 27 May 2019.