Spatiotemporal Impacts of Ideology and Social Vulnerability on COVID-19 for the United States

Erich Seamon[1*], Jennifer Johnson-Leung[1,2], Craig R. Miller[1,3], and Benjamin J. Ridenhour[1,2]

[1]University of Idaho, Institute for Modeling, Collaboration, and Innovation, Moscow, 83843, USA
[2]University of Idaho, Department of Mathematics and Statistical Science, Moscow, 83843, USA
[3]University of Idaho, Department of Biological Sciences, Moscow, 83843, USA
[*]Corresponding author: erichs@uidaho.edu

ABSTRACT

In early 2020, the Coronavirus Disease 19 (COVID-19) rapidly spread across the United States, exhibiting significant geographic variability. While several studies have examined the predictive relationships of differing factors on COVID-19 deaths, few have looked at spatiotemporal variation at refined geographic scales. The objective of this analysis is to examine spatiotemporal variation of COVID-19 deaths in association with socioeconomic, health, demographic, and political factors, using regionalized multivariate regression as well as nationwide county-level geographically weighted random forest (GWRF) models. Analyses were performed on data from three sepearate timeframes: pandemic onset until May 2021, May 2021 through November 2021, and December 2021 until April 2022.Regionalized regression results across three time windows suggest that existing measures of social vulnerability for disaster preparedness (SVI) are associated with a higher degree of mortality from COVID-19. In comparison, GWRF models provide a more robust evaluation of feature importance and prediction, exposing the importance of local features, such as obesity, which is obscured by regional delineation. Overall, GWRF results indicate a more nuanced modeling strategy is useful for capturing the diverse spatial and temporal nature of the COVID-19 pandemic.

Introduction

The burden of COVID-19 in the United States has not been distributed evenly, either geographically or from a population distribution perspective[1,2]. As well as the elderly[5,6], groups most vulnerable to COVID-19 include those with co-morbid conditions such as obesity, diabetes, heart disease, and hypertension[3,4]. Social determinants of health have also been shown to correlate with increased risk and morbidity. Further disparities in COVID-19 hospitalizations and deaths have been observed across racial and ethnic groups[7], with differences varying by geographic region[8,9] and socioeconomic status[10,11]. Adoption of protective behaviors, including non-pharmaceutical interventions like masks and vaccine uptake, has been affected by a patchwork of federal, state, and local policies[12] as well as individual responses to these policies[13]. In this study, we analyze the importance of social vulnerability, demographic and health parameters, and political geography in predicting COVID-19 deaths for the United States using novel, spatially explicit modeling techniques which incorporate the spatial and temporal variation of these potential predictive factors.

Certain communities were identified early in the pandemic as being more prone to high disease burden. For example, rural communities were identified by the World Health Organization (WHO) as skewing towards having a larger population of residents 65 or older and having fewer healthcare resources[14]. Similarly, these communities have increased rates of poverty, with 15.4 percent of rural populations living below the federal poverty line, versus 11.9 percent in urban areas[15]. Such spatial divergence across economic and social factors can also be seen with insurance access[16], telecommunications/broadband access[17], housing[18], and political ideology[19]. An additional complication in understanding the relationship between these risk factors and local pandemic outcomes is the varying collective (via government policies) and individual responses as the pandemic progressed.

Political ideology in particular has been identified as an unexpected and problematic predictor for COVID-19 mortality[20]. The importance of this spatially heterogeneous variable may be explained in part by the enhanced the ability of ideological groups to recruit followers and refine messaging as a result of the wide adoption of social media via mobile devices and the increase of virtual/remote engagement for work, school, and social activities[21]. When social and political polarization combine with

the aforementioned spatial divergence across health and demographic factors, a magnifying effect can be seen in terms of health vulnerability. The study of cognitive dissonance in relationship to new and challenging scientific information[22] and self-justification of belief structures as a protective behavior[23] provide valuable context for understanding the influence of ideology on COVID-19 beliefs and behavioral responses which may help to explain the usefulness of political ideology as an epidemiological predictor. If group messaging reinforces actions counter to recommended health behavior, populations which are most at risk bear the brunt of these consequences.

Modeling predictive and correlative features for COVID-19 across space and time has proven to be a difficult problem. There have been a number of efforts to model COVID-19 by incorporating spatial characteristics: Anderson et al.[24] used cluster analysis to evaluate spatial patterns of transmissibility, and their relationships to social health parameters; similarly, Sun et al.[25] looked at spatial lag aspects of COVID-19 in the United States, compared to county-level demographic variables. Adding to model complexity are the effects of several COVID-19 waves, which differed by location and time. As such, a number of research efforts have been conducted to explore such spatial and temporal variability of COVID-19, including disease spread and associations with social determinants of health. Andersen et al.[26] utilized a cluster analysis, in combination with a three-stage regression technique, to explore impacts to mortality and morbidity. Similarly, Mollalo et al.[27,28] evaluated differing regression-based approaches to examine spatial heterogeneity, while Xie et al.[29] applied exploratory spatial analysis methods to analyze spatial and temporal differentiation in relationship to demographic and social variables. A common theme for all of the above research efforts were the difficulties in optimizing spatial interactions, in combination with time scales.

To address these concerns, we use a strategy (Figure 1) which compares more traditional linear modeling to geographically weighted random forest (GWRF) modeling. GWRF is a statistical model that combines the strengths of two well-established techniques: Geographically Weighted Regression (GWR) and Random Forest (RF). GWR is a spatial regression model that allows the estimation of regression coefficients that vary spatially. This permits the model to capture local patterns in the data, which is particularly useful when dealing with data that exhibit spatial heterogeneity[30]. However, GWR is limited in its ability to handle complex nonlinear relationships between variables. Comparatively, RF is an ensemble machine learning model that can handle complex nonlinear relationships between variables and has the ability to capture variable interactions[31], and is a popular choice for large data sets due to low computational costs. The GWRF modeling approach combines the strengths of GWR and RF by weighting the RF model locally, allowing it to capture local patterns in the data while also handling complex nonlinear relationships[32]. GWRF modeling additionally has the ability to handle spatial heterogeneity in data and to provide predictions that are locally accurate. The value of the GWRF approach lies in its ability to provide accurate and locally relevant predictions for complex spatial data, and is particularly useful for environmental and geographical applications where spatial heterogeneity is a major concern, such as the COVID-19 pandemic data[33].

Methods and Data

We initially performed exploratory data analyses of fatality rates and deaths across time, for the entire United States, as well as by region. Based on this review of the data, we developed a set of modeling frameworks for three time frames, approximately corresponding to the Alpha, Delta, and Omicron waves in the United States. For our analyses, we define: the Alpha wave as the period from the beginning of the pandemic through the universal availability of COVID-19 vaccines to adults in the United States (January 2020–April 2021); the Delta wave as the period from May 2021–November 2021; and lastly, the Omicron wave as the shortest period December 2021–April 2022. The response variable in all of our analyses is the cumulative US county-level COVID-19 deaths, adjusted for population, for the full time period of interest. Fifteen (15) predictor variables are evaluated as county-level measures: socioeconomic status, household composition and disability, minority status and language, housing type and transportation, voting percentage, vaccination rate estimates as of April 2022, population density, obesity, unemployment, uninsured adults, social associations, diabetes, food insecurity, broadband access, and the percentage of population over 65.

We first evaluate spatial autocorrelation of all variables, for all three time frames, using Moran's I. We then use two different modeling approaches in our analyses. The first approach uses multilinear Poisson regression for nine (9) regional areas of the United States, plus the entire United States as a singular model, for a total of ten (10) distinct county groupings. Regional models are based on Health and Human Services (HHS) divisions, which were selected as a generalized policy/funding separation (HHS, http://hhs.gov). Regions 1 and 2 are combined into one region because of the small number of states in Region 2. The second approach is a more novel geographically weighted random forest modeling technique for the United States as a whole. Each model is evaluated for each of the pandemic time periods, for a total of thirty-three (33) models (Fig.1). The full set of these analyses can be seen in the associated supplemental materials.

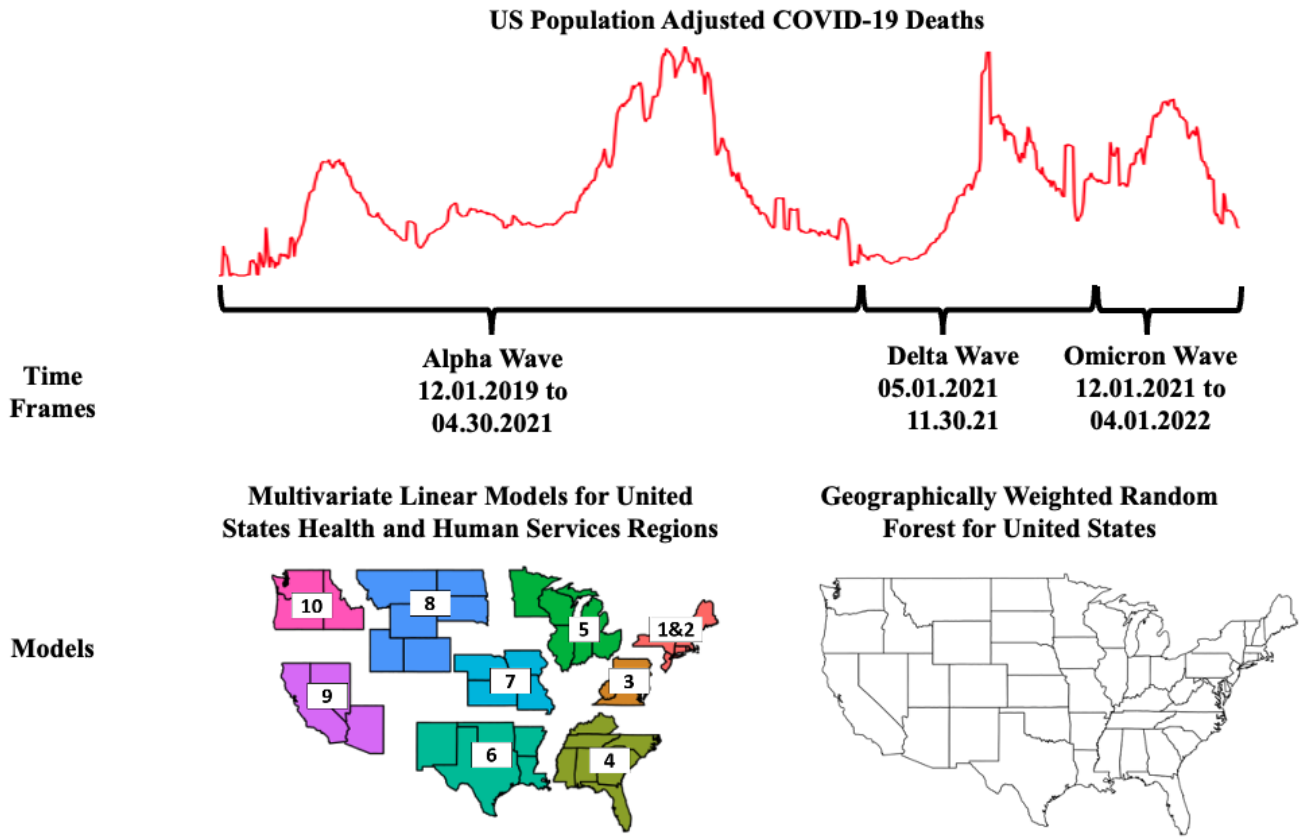


Figure 1. Analysis scheme. Our approach examines fifteen (15) different county-level independent variables, in relationship to cumulative population adjusted deaths during a pandemic wave. We regress our variables at the regional level, and subsequently use them for spatially weighted random forest models. In total, thirty-three (33) models were run (a national regression model, nine regional regression models, and a national GWRF model for each of the three time periods). We additionally assess spatial heterogeneity at a county level by using local and global Moran's I values, with Monte Carlo simulations performed to minimize error[34,35].

Data Assembly

Predictor data are collected from a number of sources: variable descriptions and sources are listed in Table 1. All data were normalized based on a 0 to 1 scaling structure. Figure 2 shows the spatial distributions of a select set of independent variables. Of particular note are issues of missing data related to vaccinations. Early in the pandemic, a number of states (Texas, New Hampshire) failed to report vaccination data: as such, assessing vaccination rates early on (2020) is not possible for the entire United States. Because of this, we use county-level vaccination rates (defined as the percentage of individuals receiving at least one dose of a vaccine) as of April 2022; thus our vaccination measure reflects the portion of the population willing to (eventually) seek out a vaccine, rather than the portion of the population actually immunized at a particular time. Additionally, a number of sparsely populated counties (<20) have reporting errors with regards to deaths and case counts. In order to appropriately fit our models, these counties are excluded from our analyses.

Variable Name	Description		Source
Deaths	Cumulative deaths in time period per population		New York Times[36]
Socioeconomic Status	Employment, Income, Education	Standardized composite measures calculated at the census tract level for emergency management	CDC SVI 2018[37]
Household Composition & Disability	Age, Disability, Single Parent		
Minority Status & Language	Race, English Proficiency		
Housing Type & Transportation	Mobile Homes, Multi-Unit No Vehicle, Crowding, Group		
Obesity	Incidence per population via CDC Behavioral Risk Factors Surviellance System (BRFSS)		U Wisconsin Population Institute County Health Ranking Model[38]
Diabetes			
Unemployment	Incidence per population via Bureau Labor & Statistics		
Uninsured	Incidence per population via Census Small Area Health Insurance Estimates Program (SAHIE)		
Social Associations	Incidence per population of membership in clubs, churches, etc.		
Food Insecurity	Relative index via Feeding America		
Broadband Access	Incidence per population with no access via Census American Community Survey 2020		
Population Density	Relative population per unit area, normalized		2020 US Census[39]
Voting	Percent 2020 Presidential vote for Biden		MIT Election Lab[40]
Vaccination	Incidence per population of at least one vaccination by April 2022		CDC[41]
Age Over 65	County level population over age 65		US Census[39]

Table 1. Description and Sources of Model Variables

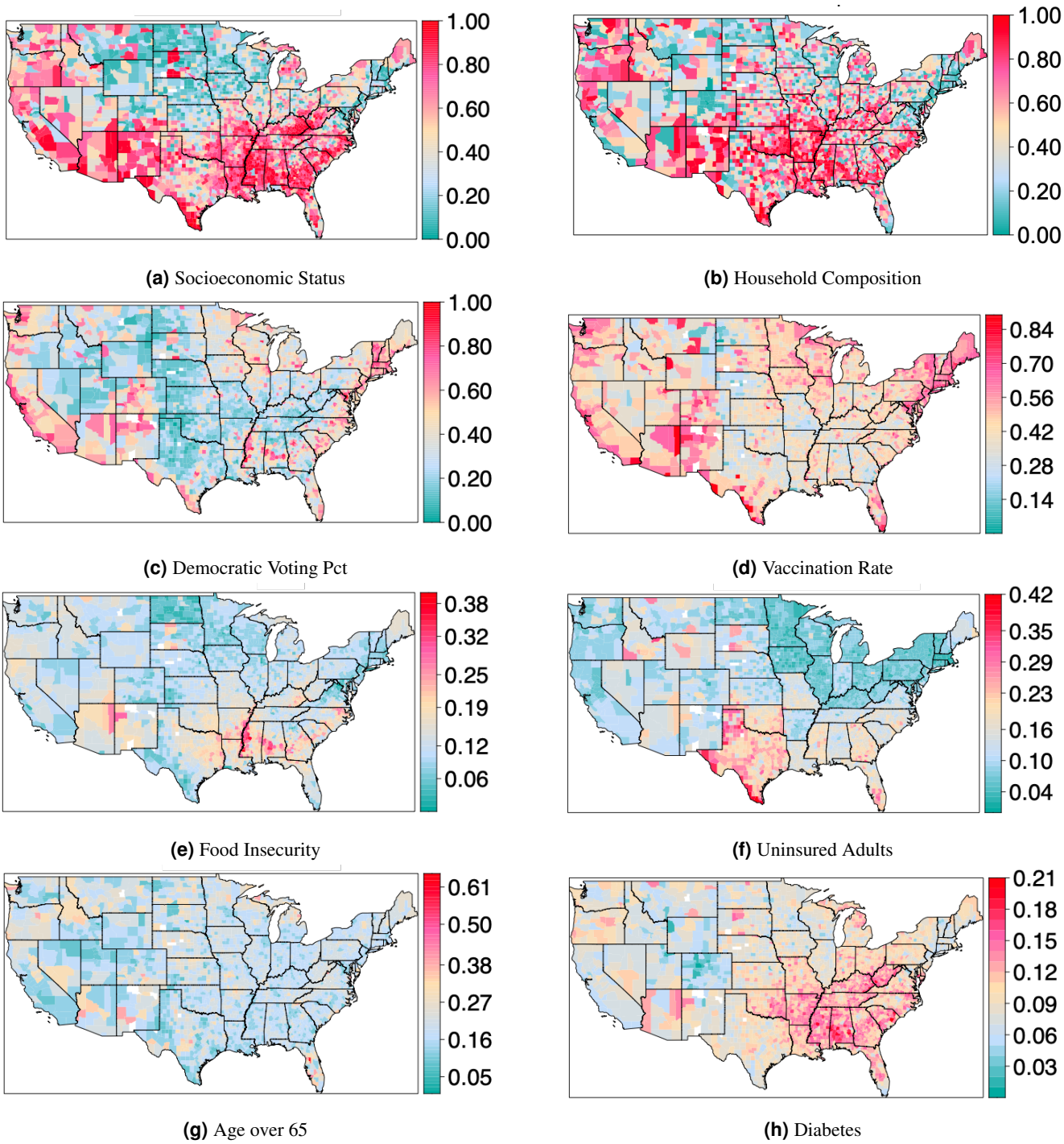


Figure 2. Select predictor variables mapped for the entire United States. a) Socioeconomic status: SVI representing income, poverty, employment, and education. b) Household composition: SVI representing levels of single parent households, disabilities, or those with children or the elderly. c) Democratic voting percentage: Percentage of democratic vote from the 2020 general presidential election. d) Vaccination rate: Percentage of individuals in a county with at least one vaccine dose, as of April 1, 2022. e) Food insecurity: Percentage of households with insufficient access to food, or food of an adequate quality. f) Uninsured adults: Percentage of adults that are uninsured. g) Age over 65: Percentage of individuals over age 65. h) Diabetes. Percentage of adults that have diabetes. Maps for all fifteen predictor variables can be found in the associated supplemental materials.

Spatial Autocorrelation

The Moran's I statistic,

$$I = \frac{n}{W}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right)\right)\left(\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2\right)^{-1}$$

measures spatial autocorrelation[34,35,42]. Here n is the number of spatial units, w_{ij} are spatial weights, x is the variable being tested for autocorrelation with mean $\bar{x}$, and $W = \sum_{i,j} w_{ij}$. The Moran's I weight matrix specifies the degree of spatial proximity between pairs of spatial units, and can be calculated using contiguity-based weighting, network-based weighting, or distance-based weighting. Here we use contiguity-based weighting, which can be described as a n × n positive symmetric matrix W, where the element of this matrix is w_{ij} at location i, j for n locations. There are a number of contiguity weight approaches, including linear contiguity, rook contiguity, bishop contiguity, and queen contiguity[43]. Here we use queen contiguity, which creates a neighborhood list based on a common edge or a common vertex. Queen contiguity is recommended when irregular polygons (e.g. counties) are used[44]. This provides a method to evaluate spatial patterns in terms of positive (clustering), negative (dispersion), or neutral (random) spatial autocorrelation, with a range from -1 to 1. By examining Local Moran's I values for each observation (in this instance, an individual county), we can look at clustering in combination with other factors (i.e. voting). Such comparisons across differing time frames give insight into how COVID-19 deaths may be spatially correlating with covariates[45].

Regression Analysis and Geographically Weighted Random Forest

We construct regionalized Poisson regression models (with population as an offset) for each of the three time windows (Alpha, Delta, and Omicron), using all fifteen variables, with cumulative county-level death counts as the dependent variable. Normality assumptions are evaluated, including residual and partial residual/component residual plots, as well as standardized beta coefficients.

To address some of the limitations of the regression analysis, a modified novel approach of ensembled regression decision trees (random forest) were used. Regression decision trees are a method of constructing a set of decision rules on a predictor variable[46–48] that is continuous (versus categorical). These rules are constructed by recursively partitioning the data into successively smaller groups with binary splits based on a single predictor variable, with the goal of encapsulating the training data in the smallest possible tree[49]. Random forest, or ensembled decision trees, are a combination of many decision tree predictors, where each tree depends on the values of a random vector, sampled independently and with the same distribution for all trees in the forest[31]. Random forest modeling reduces the potential for over-fitting through the use of bootstrap aggregation, averaging across many trees, and provides a level of feature importance for assessing predictor power[46,50].

Geographically weighted random forest (GWRF) is a modified version of the classic random forest approach, which incorporates spatial non-stationarity, as part of spatially weighted regression (SWR) techniques[32]. GWRF fits a sub-model for each observation in space, taking into account neighboring observations (in this instance, observations are represented as counties), and is based on Fotheringham, Yang and Kang's[30] work on spatially-explicit coefficient modeling. The main difference between a traditional (linear) SWR and GWRF is that we can model non-stationarity within a non-linear model, which minimizes over-fitting and thus relaxes the assumptions of traditional Gaussian statistics. As part of our GWRF construction, we utilized 10-fold cross validation, a model validation technique used to assess the generalizability of the model. Model construction for this analysis used the recursive partitioning and regression trees package (*rpart*), as well as the *gfr* package within R[46,50]. For the three sets of models developed (Alpha, Delta, and Omicron wave time periods), external cross validation was performed, using an adaptive kernel distance weighting function across a range of bandwidth values (number of surrounding counties, which ranged from 3 to 30).

The adaptive kernel function as part of the GWRF model is defined as:

$$W_{ij} = exp\left(-\frac{|d_{ij}|}{bw}\right)$$

where w_{ij} is the weight assigned to the observation j for the estimation of i, and d_{ij} is the distance between j and i, and bw is the bandwidth (number of surrounding counties). Each variant model was run for progressively larger bandwidths, in combination with inversely varying global and local model weighting (local weighting at .25, .50, .75, and 1, combined with global weighting at 1, .75, .50, .25), to evaluate both root mean square error (RMSE) as well as mean absolute errors (MAE). Using this optimization method, we were able to select the model parameters which minimized error for each variant window (Supplemental Figures S23, S27, S31).

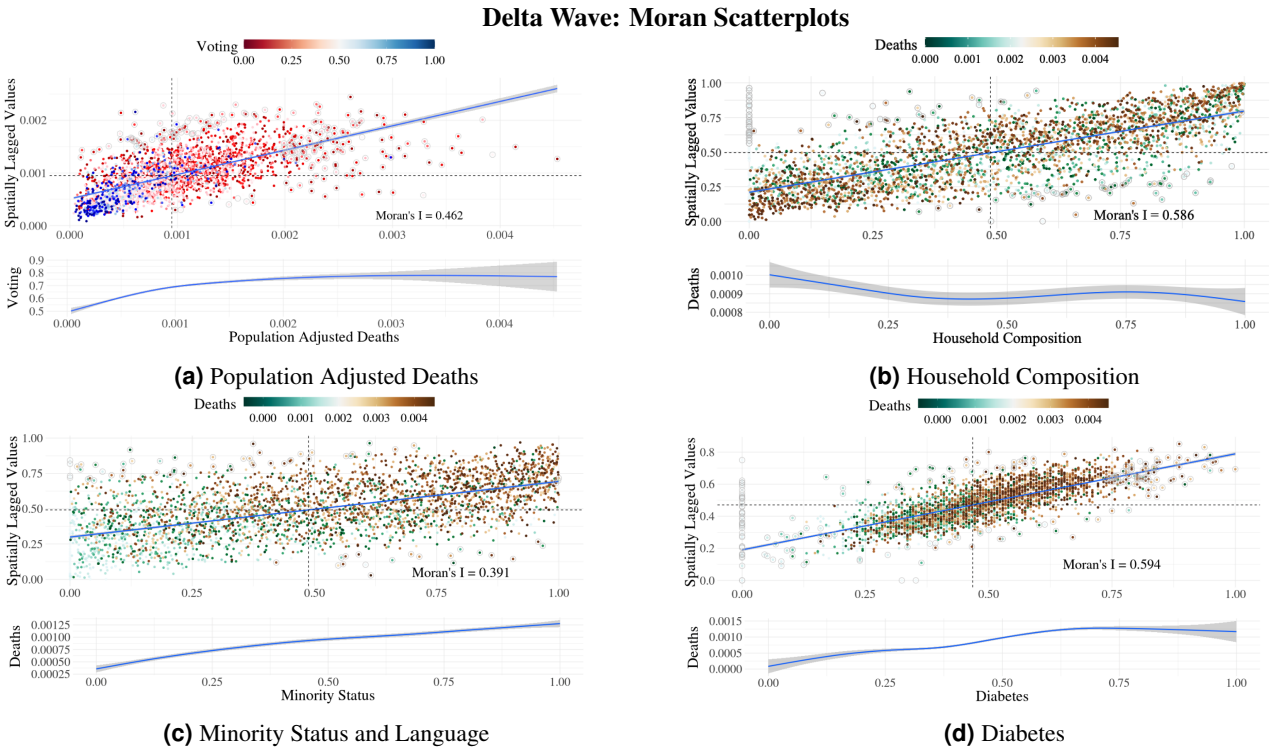
Results

All timeframes (Alpha, Delta and Omicron) indicated an overall positive spatial clustering, with moderate outliers, for COVID-19 deaths (localized Moran's I, M = 0.462). By plotting deaths vs spatially lagged values (Fig. 3a) we see a progression from negative spatial autocorrelation for more liberal counties (blue counties, lower left quadrant) to positive spatial autocorrelation for more conservative counties (red counties, upper right quadrant). Outliers were overwhelmingly conservative-voting counties, which suggests that conservative counties are having a stronger influence over model fit. In addition, Monte Carlo simulations confirmed strong positive spatial autocorrelation (spatial clustering), with a density peak to the left of mean cumulative deaths values, for all three time windows. When examining the clustering behavior for independent variables, we see varying degrees of positive spatial clustering, as measured by the Moran's I statistic. Fig. 3b shows that household composition is spatially correlated and that increased death rates were observed in poorer counties. Similar effects can be seen for the other SVI measures (Fig. 3c) as well as known health risk factors, such as diabetes (Fig. 3d) A number of variables—uninsured adults, social associations, unemployment—show no consistent pattern of deaths in relation to spatial clustering. See Part 4 of the supplementary materials for Moran's scatterplots for each of the variables in all three waves.

Regression analysis results were performed for the coterminous United States, as well as by region, with full results found in the provided supplemental materials. The R^2 for regional US regression models varied across the differing variant windows (Alpha $R^2 = 0.41$ (Region 7) to $R^2 = 0.83$ (Region 9); Delta $R^2 = 0.65$ (Region 6) to R^2=0.90 (Region 3); Omega R^2=0.47 (Region 7) to $R^2 = 0.90$ (Region 9). Overall, regional R^2 were moderately higher than national regression model values (US Alpha $R^2 = 0.416$; US Delta $R^2 = 0.675$; US Omicron $R^2 = 0.505$). Across all variant windows, region 10, regions 1 and 2 and region 9 consistently performed best, with region 4, region 8 and region 7 having the lowest R^2 values (Fig. 4). Which predictors were significant varied considerably by region and time window, but several patterns were seen. For example, uninsured adults and housing composition were significant for all time windows for region 6; diabetes was significant across all time windows for region 4; voting was significant for singular time windows for a number of regions, with regions 9 and 10 having two time windows that were significant; and vaccination rate, as a proxy for protective health behaviors, became non-significant across most regions as the pandemic progressed, except for region 5.

Geographically weighted random forest outputs for the three models in question (Alpha, Delta, Omicron), optimized for bandwidth selection using local/global model root mean square errors (RMSE), performed well in tracking general spatial trends, while overestimating extreme values (Fig. 5). Global out of bag (OOB) R^2 for the three models were 0.40 (Alpha), 0.45 (Delta), and 0.34 (Omicron), while localized R^2 by county indicated a clear variation in predictive power, with values as high as 0.90. The R^2 values in this instance are generated by dividing the vector of mean square errors by the variance of y, then subtracting from 1. Given that the GWRF model approach permutes across a large number (500+) of decision trees based on an adaptive kernel function, there is the possibility of models performing worse than a random outcome (resulting in some R^2 values below zero)[51]. Predictive strength for early in the pandemic (Alpha) was strongest in coastal Eastern regions, the Southwest, as well as Pacific Northwest coastal regions. We see a shift of model performance during the Delta wave, with the west coast regions, the Midwest, and Colorado being highest in terms of predictive power. For Omicron, predictive strength had a more varied spatial distribution, with clusters in Southern California and Arizona. Optimized bandwidth selection (number of counties included in individual county RF model runs) ranged from as high as 24 (Alpha), to 12 (Omicron).

For GWRF, mapped feature importance values (Fig. 6) provide spatial patterns for predictor variables across all time variant models. For this analysis, feature importance is defined as the average increase in squared residuals of the test dataset, as variables are randomly permuted[46]. Higher values indicate a greater importance of the variable on the model performance. Given this random permutation, there can be instances where the mean square error of a random variable may be higher than a selected variable, generating a negative value. Model results indicate that diabetes feature importances were highly



(a) Population Adjusted Deaths

(b) Household Composition

(c) Minority Status and Language

(d) Diabetes

Figure 3. Select spatial autocorrelation Moran's scatterplots for the Delta wave time window. Each point represents an individual county value. a) COVID-19 Population adjusted deaths Moran's scatterplot. Color range depicts voting. b) Household Composition Moran's scatterplot. Color range depicts COVID-19 population adjusted deaths. c) Minority Status and Language Moran's scatterplot. Color range depicts COVID-19 population adjusted deaths. d) Diabetes Moran's scatterplot. Color range depicts COVID-19 population adjusted deaths. Below each Moran plot is a loess plot of the main variable (x-axis) in comparison to the categorized variable (y-axis). Moran's plots for all variables are available in the supplemental materials.

impactful for predicting COVID-19 death in the southwestern portion of the United States across all three time windows; similarly, vaccination rates were impactful in the northern Utah region, across both Alpha and Delta time windows. Household composition influence were high in the Utah/Colorado region for Alpha and Delta time windows as well, while voting was variable across all three variants, with small hotspots in many rural communities. Obesity also provided interesting spatial patterns, with uniquely high feature importance values in the northeastern portions of the United States (Maine). All feature importance plots can be found in the attached supplemental materials (Supplementary Figures S26, S30, and S34).

Comparisons of regression results with random forest outputs suggest that regional boundaries in the linear model may be limiting predictive capabilities, given its artificial geographic structure. While a number of the regional outputs indicate moderately performing models (region 10 (Pacific Northwest), region 1 and 2 (Northeast), region 9 (West)), a majority of the regional regression models had moderate to poor predictive power. Temporal patterns of model performance show higher R^2 values in the Alpha and Delta waves. This suggests that early in the pandemic, typical sociological, economic, political, and health parameters had a greater effect on predicting deaths—yet over time, external effects became more influential. Such effects might include: varying policy response, changing economic factors, social media misinformation, and increasing population immunity either through exposure to the virus, vaccination, or a combination. Conversely, our customized GWRF nationwide model performed considerably better, with feature importance at a county level indicating clustering patterns which do not conform to HHS regional boundaries. Global model results had R^2 values between 0.88 and 0.90. Out of bag R^2 values for GWRF outputs ranged from 0.34 to 0.45, suggesting that, while general trends can be seen, the the model is not able to accurately predict the cumulative deaths for counties in the tails of the distribution. Feature importance rankings displayed interesting spatial patterns (See Supplementary Figs. S26, S30 and S34). Urbanized centers showed higher model influence for minority status and language across all three time windows; socioeconomic measures had the most effect on model performance in rural areas, particularly in the western portions of the United States. Voting showed considerable heterogeneity in terms of model influence, with small pockets of increase in rural regions as well.

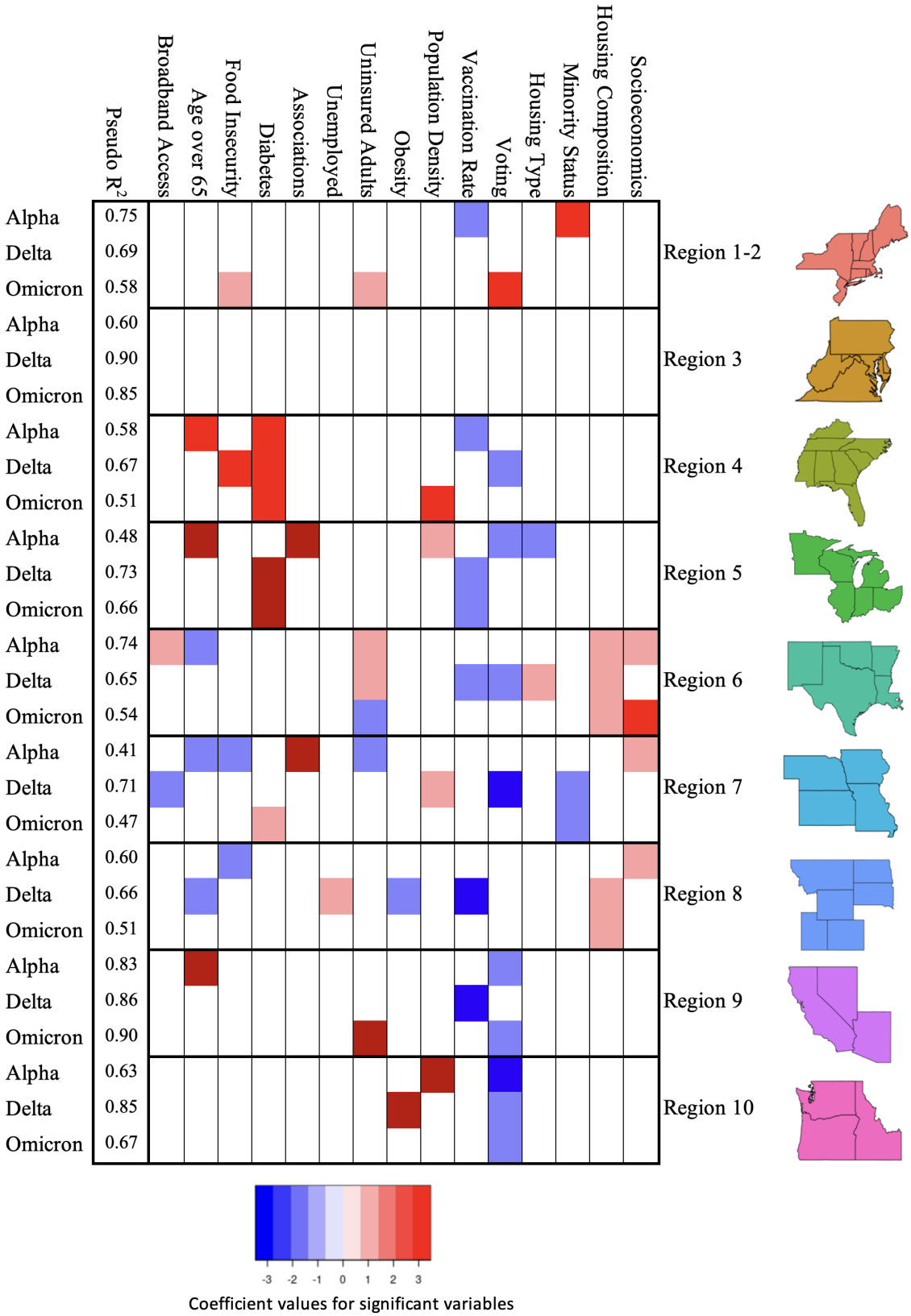


Figure 4. Significant variables for each regional model, as well as R^2 values, per variant time window.

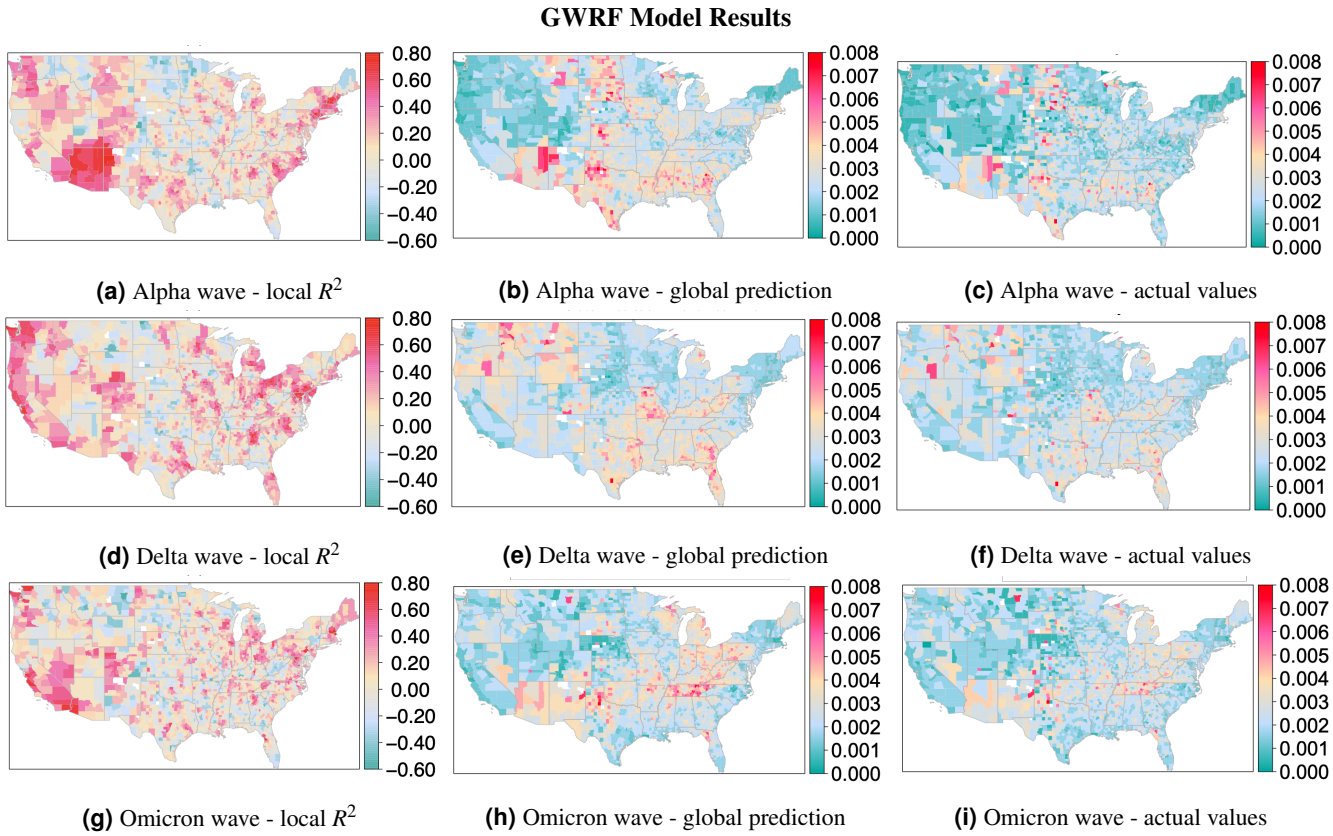


Figure 5. Geographically weighted random forest model results for Alpha, Delta, and Omicron wave time windows. Each panel indicates localized R^2 values, global prediction results, and observed cumulative deaths by county (adjusted for population). a) thru c): Panel plot of Alpha wave results. d) thru f): Panel plot of Delta wave results. g) thru i): Panel plot of Omicron wave results. Full model results for all waves can be seen as part of the supplemental materials.

Discussion

The United States recorded over 1.1 million excess deaths during the period of the COVID-19 pandemic from January 2020 through March 2022[52], with some communities experiencing much greater losses than others. While our regional regression models were able to explain a large portion of the variation in many cases, this did not always result in the identification of significant features. For example, in Region 3 not a single variable was significant despite the high R^2 values (Figure 4). Applying spatial modeling methods exposes relationships among predictor variables which are not evident through more traditional modeling techniques. Geographic weighting approaches applied with non-parametric random forest methods better capture variation corresponding with geographic heterogeneity than our regional regression models.

Political ideology was identified early in the pandemic as an atypical predictor of deaths due to COVID-19[53]. Our work supports the finding that political ideology, as measured by voting data, serves as an important feature for explaining variation in COVID-19 deaths in some areas and at certain time intervals, such as the western United States during the Delta wave (Figure 6). Population adjusted deaths for all three time frames show positive spatial autocorrelation—with conservative voting from the 2020 general election aligned with higher deaths, and democratic voting aligned with lower deaths (Fig. 3a). This reinforces previous research which indicates that stronger exposure to conservatism increases age-standardized COVID-19 mortality rates[54]. Results of our GWRF modeling additionally indicate progressively increasing importance of voting on model performance, with Delta and Omicron waves showing strong regionalized influences in the western United States as well as the northeast, indicating a increasing interaction between political ideology and the changing COVID-19 climate over time.

We expected that measures of obesity and diabetes would be important for predicting deaths, as these health factors are known to be strongly associated with increase risk of morbidity and mortality from COVID-19[55]. Indeed, spatial trends for both obesity and diabetes in all waves indicate that increasing spatial autocorrelation corresponds to a higher COVID-19 death rate, as seen for the Delta wave in Figure 3d. However, the only region for which one or both of these variables was significant

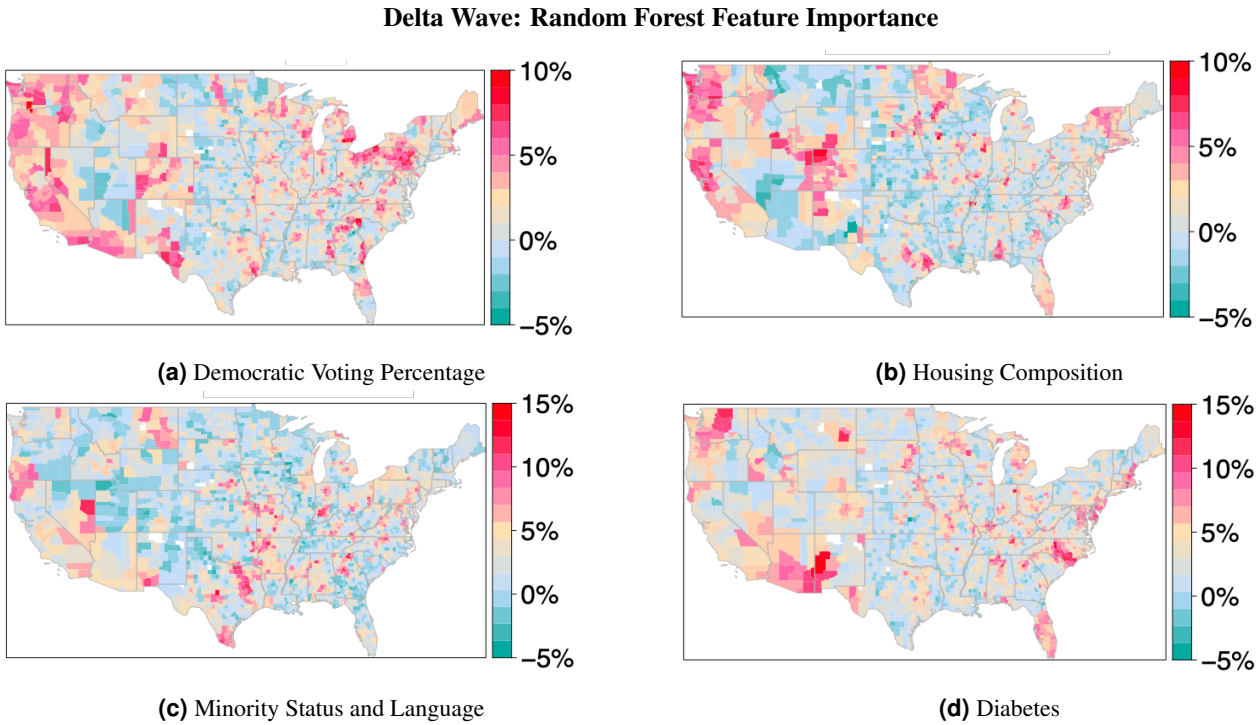


Figure 6. Delta wave random forest feature importance for housing composition, diabetes, minority status and language, and democratic voting percentages (2020 US presidential election). Other feature importance maps can be found in the associated supplemental materials.

in all waves was the southeast (Region 4). This demonstrates that important vulnerabilities are not captured by the regional regression but are detected by our GWRF feature importance plots.

All of the composite SVI measures, which were constructed for use by FEMA for responding to disasters[56], showed a positive association with deaths in the Moran's I plots in the first two waves. This relationship was less marked for the Omicron period, with only the measures corresponding to socioeconomic status and minority populations continuing the pattern. This supports previous findings[57,58] that these measures are also useful for predicting clusters of counties with high vulnerability to pandemic threat. Additionally, spatial autocorrelation of both dependent and independent variables show clustering patterns which suggest that particular factors may have differing spatial influences on deaths on whole. For example, for vaccination rates we see decreasing spatial autocorrelation with increasing deaths. Overall, spatial lag variations between independent variables is reflective of local socioeconomic and cultural views which can dampen (or exacerbate) the effects of COVID-19 associated factors (e.g. deaths)[59].

Select regional analyses provide insights into the value of a spatially explicit modeling approach. For example, variable importance values for obesity across the Delta and Omicron waves show a unique hotspot for the state of Maine. Given Maine's demographics as older, rural, yet politically progressive, model results suggest that obesity has a uniquely strong influence on predicting COVID-19 associated deaths, which have associations with statewide policy or regulatory components[60]. Comparatively, older populations (over 65) had a moderately stronger influence on model performance in southeastern rural areas (northeastern Georgia, northern Mississippi) across all three variant time windows. This pattern was similar to other regions of the United States, where older populations in rural regions showed higher levels of model influence (central Texas, northeastern Pennsylvania). Such patterns reinforce understandings of rurality on health outcomes, which may be interrelated to poverty and preventative care access[6].

One limitation of the GWRF models appears in their limited ability to predict high and low levels of deaths. This may be due to confounding variables which are not included in the model which could include, but are not limited to, health parameters (hypertension, cardiovascular disease) and hospitalization effectiveness. Limiting model time windows to the three main variant waves could also be a concern. Future analyses could address these issues by incorporating autoreggressive integrated moving average (ARIMA) techniques, in combination with spatially-focused random forest, or potentially other spatiotemporal algorithms (or an ensemble of multiple algorithmic techniques)[61].

Early in the pandemic, deaths were generally higher in more liberal, population dense regions of the United States. Over time, we see a shift in deaths to more conservative, rural communities, which is likely a combination of vaccination rates and the overall spread of COVID-19 to lesser populated regions. We further see focused regions within the United States (southern Mississippi Delta region, southwestern portions of Arizona, Colorado, and New Mexico) that show variable clustering as well as high levels of model predictive power (Fig. 5). These clusters span multiple HHS boundaries, which helps to explain the weaker regression modeling performance for sections of the country. Given spatial and temporal variation of deaths and the inability for linear modeling techniques to effectively perform, policy decisions for future pandemic responses should consider spatially and temporally sensitive modeling efforts to assess public need[33]. In particular, collaboration across fixed boundaries for health regions, which may be used for funding allocations and policy decision, may be necessary for successful interventions. Our GWRF results indicate a considerable difference in spatial feature importance patterns between all three wave events, which align with more qualitative examinations of the pandemic response in the United States.

Acknowledgements

Research reported in this publication was supported by the National Institute Of General Medical Sciences of the National Institutes of Health under Award Number P20GM104420. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We would additionally like to thank the pandemic modeling group at the Institute for Modeling Collaboration and Innovation (IMCI) at the University of Idaho for help working on and thinking about COVID-19 related issues. Finally, we thank Dr. Holly Wichman for her invaluable research leadership at IMCI.

Author contributions

E.S., J.J.L. and C.R.M. conceived the analyses, E.S. conducted the analyses, and E.S., J.J.L. and B.J.R. analysed the results. All authors reviewed the manuscript.

Data Availability

A supplementary appendix as well as associated datasets can be found at: https://doi.org/10.5061/dryad.4j0zpc8j1

Competing Interests

The authors declare no competing interests.

References

1. Cuadros, D. F., Branscum, A. J., Mukandavire, Z., Miller, F. D. W. & MacKinnon, N. Dynamics of the COVID-19 epidemic in urban and rural areas in the United States. *Annals Epidemiol.* **59**, 16–20, DOI: 10.1016/j.annepidem.2021.04.007 (2021).

2. Dasgupta, S. *et al.* Differences in rapid increases in county-level COVID-19 incidence by implementation of statewide closures and mask mandates — United States, June 1–September 30, 2020. *Annals Epidemiol.* **57**, 46–53, DOI: 10.1016/j.annepidem.2021.02.006 (2021).

3. Zheng, Z. *et al.* Risk factors of critical and mortal COVID-19 cases: A systematic literature review and meta-analysis. *J. Infect.* **81**, e16–e25, DOI: 10.1016/j.jinf.2020.04.021 (2020).

4. Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269, DOI: 10.1038/s41586-020-2008-3 (2020).

5. Daoust, J. F. Elderly people and responses to COVID-19 in 27 Countries. *PLoS ONE* **15**, 1–13, DOI: 10.1371/journal.pone.0235590 (2020).

6. Mueller, A. L., McNamara, M. S. & Sinclair, D. A. Why does COVID-19 disproportionately affect older people? *Aging* **12**, 9959–9981 (2020).

7. Karaca-Mandic, P., Georgiou, A. & Sen, S. Assessment of COVID-19 Hospitalizations by Race/Ethnicity in 12 States. *JAMA Intern. Medicine* **181**, 131–134, DOI: 10.1023/A (2021).

8. Sze, S. *et al.* Ethnicity and clinical outcomes in COVID-19: A systematic review and meta-analysis. *EClinicalMedicine* **29-30**, 100630, DOI: 10.1016/j.eclinm.2020.100630 (2020).

9. Rostila, M. *et al.* Disparities in Coronavirus Disease 2019 Mortality by Country of Birth in Stockholm, Sweden: A Total-Population-Based Cohort Study. *Am. journal epidemiology* **190**, 1510–1518, DOI: 10.1093/aje/kwab057 (2021).

10. Agarwal, R. *et al.* Socioeconomic privilege and political ideology are associated with racial disparity in COVID-19 vaccination. *PNAS* **118**, 2021, DOI: 10.1073/pnas.2107873118/-/DCSupplemental (2021).

11. Moxley, T. A., Johnson-leung, J., Seamon, E., Williams, C. & Ridenhour, B. J. Application of Elastic Net Regression for Modeling COVID-19 Sociodemographic Risk Factors. *medRxiv* **Preprint**, 1–19 (2023).

12. Yang, B. *et al.* Effect of specific non-pharmaceutical intervention policies on SARS-CoV-2 transmission in the counties of the United States. *Nat. Commun.* **12**, 4–13, DOI: 10.1038/s41467-021-23865-8 (2021).

13. Ridenhour, B. J. *et al.* Effects of trust, risk perception, and health behavior on COVID-19 disease burden: Evidence from a multi-state US survey. *PLoS ONE* **17**, 1–16, DOI: 10.1371/journal.pone.0268302 (2022).

14. World Health Organization. Novel Coronavirus (2019-nCOV). Tech. Rep. 6, World Health Organization (2020).

15. Dobis, E. *et al.* Rural America at a Glance. Tech. Rep. 230, Ecnomic Research Service (ERS) (2021). DOI: 10.2139/ssrn.2202862.

16. Wray, C. M., Khare, M. & Keyhani, S. Access to Care, Cost of Care, and Satisfaction with Care among Adults with Private and Public Health Insurance in the US. *JAMA Netw. Open* **4**, 1–12, DOI: 10.1001/jamanetworkopen.2021.10275 (2021).

17. Reddick, C. G., Enriquez, R., Harris, R. J. & Sharma, B. Determinants of broadband access and affordability: An analysis of a community survey on the digital divide. *Cities* **106** (2020).

18. Narine, L. & Shobe, M. A. Making sense of housing disparities research: A review of health and economic inequities. *Soc. Work. Public Heal.* **29**, 35–41, DOI: 10.1080/19371918.2011.619454 (2014).

19. Cakanlar, A., Trudel, R. & White, K. Political Ideology and the Perceived Impact of Coronavirus Prevention Behaviors for the Self and Others. *J. Assoc. for Consumer Res.* **7**, 36–44, DOI: 10.1086/711834 (2022).

20. Sehgal, N. J., Yue, D., Pope, E., Wang, R. H. & Roby, D. H. The Association Between COVID-19 Mortality And The County-Level Partisan Divide In The United States. *Heal. Aff.* **41**, DOI: https://doi.org/10.1377/hlthaff.2022.00085 (2022).

21. Volkmer, I. Social Media and COVID-19: A global study of digital crisis interaction among Gen Z and millennials. Tech. Rep., World Health Organization (WHO) (2021).

22. Festinger, L. *A Theory of Cognitive Dissonance* (Stanford University Press, 1957).

23. Tavris, C. & Aronson, E. *Mistakes were made (but not by me): Why we justify foolish beliefs bad decisions, and hurtful acts* (Harper Collins, 2007), harcourt edn.

24. Anderson, R. M., Heesterbeek, H., Klinkenberg, D. & Hollingsworth, T. D. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet* 2–5 (2020).

25. Sun, F., Matthews, S. A., Yang, T. C. & Hu, M. H. A spatial analysis of the COVID-19 period prevalence in U.S. counties through June 28, 2020: where geography matters? *Annals Epidemiol.* **52**, 54–59.e1, DOI: 10.1016/j.annepidem.2020.07.014 (2020).

26. Andersen, L. M., Harden, S. R., Sugg, M. M., Runkle, J. D. & Lundquist, T. E. Analyzing the spatial determinants of local Covid-19 transmission in the United States. *Sci. Total. Environ.* **754**, 142396, DOI: 10.1016/j.scitotenv.2020.142396 (2021).

27. Mollalo, A., Vahedi, B. & Rivera, K. M. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Sci. Total. Environ.* **728**, 138884, DOI: 10.1016/j.scitotenv.2020.138884 (2020).

28. Mollalo, A., Rivera, K. M. & Vahabi, N. Spatial statistical analysis of pre-existing mortalities of 20 diseases with COVID-19 mortalities in the continental United States. *Sustain. Cities Soc.* **67** (2021).

29. Xie, Z. *et al.* Spatial and temporal differentiation of COVID-19 epidemic spread in mainland China and its influencing factors. *Sci. Total. Environ.* **744**, 140929, DOI: 10.1016/j.scitotenv.2020.140929 (2020).

30. Fotheringham, A. S., Yang, W. & Kang, W. Multiscale Geographically Weighted Regression (MGWR). *Annals Am. Assoc. Geogr.* **107**, 1247–1265, DOI: 10.1080/24694452.2017.1352480 (2017).

31. Breiman, L. Statistical Modeling: The Two Cultures. *Stat. Sci.* **16**, 199–231, DOI: 10.2307/2676681 (2001). 0010.

32. Georganos, S. *et al.* Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto Int.* **36**, 121–136, DOI: 10.1080/10106049.2019.1595177 (2021).

33. Franch-Pardo, I., Napoletano, B. M., Rosete-Verges, F. & Billa, L. Spatial analysis and GIS in the study of COVID-19. A review. *Sci. Total. Environ.* **739**, 140033, DOI: 10.1016/j.scitotenv.2020.140033 (2020).

34. Moran, P. Notes on Continuous Stochastic Phenomena Published by : Biometrika Trust Stable URL : http://www.jstor.org/stable/2332142. *Biometrika* **37**, 17–23 (1950).

35. Li, H., Calder, C. A. & Cressie, N. Beyond Moran's I: Testing for spatial dependence based on the spatial autoregressive model. *Geogr. Analysis* **39**, 357–375, DOI: 10.1111/j.1538-4632.2007.00708.x (2007).

36. The New York Times. Coronavirus (Covid-19) Data in the United States (2021).

37. Flanagan, B. E., Hallisey, E. J., Adams, E. & Lavery, A. Measuring Community Vulnerability to Natural and Anthropogenic Hazards: The Centers for Disease Control and Prevention's Social Vulnerability Index. *J. J Environ Heal.* **80**, 34–36 (2018).

38. University of Wisconsin Population Health Insti. 2022 County Health Rankings National Findings Report. Tech. Rep., University of Wisconsin School of Medicine and Public Health (2022).

39. U.S. Census Bureau. US Population, 2016-2020 American Community Survey 5-year estimates (2020).

40. Massachusetts Institute of Technology (MIT) Elections Lab. County Presidential Election Returns 2000-2020, DOI: 10.7910/DVN/VOQCHQ (2018).

41. Centers for Disease Control and Prevention (CDC). COVID Data Tracker (2022).

42. Griffith, D. A. *Spatial Autocorrelation* (Elsevier, 2009).

43. Suryowati, K., Bekti, R. D. & Faradila, A. A Comparison of Weights Matrices on Computation of Dengue Spatial Autocorrelation. *IOP Conf. Series: Mater. Sci. Eng.* **335**, DOI: 10.1088/1757-899X/335/1/012052 (2018).

44. Chen, Y. On the four types of weight functions for spatial contiguity matrix. *Lett. Spatial Resour. Sci.* **5**, 65–72, DOI: 10.1007/s12076-011-0076-6 (2012).

45. Anselin, L. Local Indicators of Spatial Association—LISA. *Geogr. Analysis* **27**, 93–115, DOI: 10.1111/j.1538-4632.1995.tb00338.x (1995).

46. Breiman, L., Friedman, J., Stone, C. J. & Olshen, R. A. *Classification and Regression Trees*. The Wadsworth and Brooks-Cole statistics-probability series (Taylor and Francis, 1984).

47. Verbyla, D. Classification Trees: a new discrimination tool. *Can. J. For. Resour.* (1987).

48. Clark, L., Pregibon, D. Tree-Based Models. In Chambers, J., H. T. (ed.) *Statistical Models in S* (Routledge., 1992).

49. Prasad, A. M., Iverson, L. R. & Liaw, A. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems* **9**, 181–199, DOI: 10.1007/s10021-005-0054-1 (2006).

50. Atkinson, E. J. & Therneau, T. M. An Introduction to Recursive Partitioning Using the RPART Routines Basic steps needed to use rpart Rpart model options Plotting options. Tech. Rep. October, Mayo Foundation (2000).

51. Liaw, A. & Wiener, M. Classification and Regression by randomForest. *R News* **2**, 18–22 (2002).

52. Centers for Disease Control and Prevention (CDC). Excess Deaths Associated with COVID-19 (2023).

53. Gao, J. & Radford, B. J. Death by political party: The relationship between COVID-19 deaths and political party affiliation in the United States. *World Med. Heal. Policy* **13**, 224–249, DOI: 10.1002/wmh3.435 (2021).

54. Krieger, N., Testa, C., Chen, J. T., Hanage, W. P. & McGregor, A. J. Relationship of political ideology of US federal and state elected officials and key COVID pandemic outcomes following vaccine rollout to adults: April 2021–March 2022. *Lancet Reg. Heal. - Am.* **16**, 100384, DOI: 10.1016/j.lana.2022.100384 (2022).

55. Al-Sabah, S., Al-Haddad, M., Al-Youha, S., Jamal, M. & Almazeedi, S. COVID-19: Impact of obesity and diabetes on disease severity. *Clin. Obes.* **10**, DOI: 10.1111/cob.12414 (2020).

56. Flanagan, B. E., Gregory, E. W., Hallisey, E. J., Heitgerd, J. L. & Lewis, B. A Social Vulnerability Index for Disaster Management. *J. Homel. Secur. Emerg. Manag.* **8**, DOI: 10.2202/1547-7355.1792 (2011).

57. Chen, J. T. & Krieger, N. Revealing the unequal burden of COVID-19 by income, race/ethnicity, and household crowding: US county versus zip code analyses. *J. Public Heal. Manag. Pract.* **27**, S46–S56, DOI: 10.1097/PHH.0000000000001263 (2021).

58. Dalton, A. F. *et al.* Relationships Between Social Vulnerability and Coronavirus Disease 2019 Vaccination Coverage and Vaccine Effectiveness. *Clin. Infect. Dis.* **76**, 1615–1625, DOI: 10.1093/cid/ciad003 (2023).

59. Santos, F., Graw, V. & Bonilla, S. A geographically weighted random forest approach for evaluate forest change drivers in the Northern Ecuadorian Amazon. *PLoS ONE* **14**, 1–37, DOI: 10.1371/journal.pone.0226224 (2019).

60. Robinson, K. N. & Saber, A. Obesity : Policy and Practice Recommendations for High-Risk Populations In fl uenced by the COVID-19 Pandemic. *mSystems* **7, No.3** (2022).

61. Kane, M. J., Price, N., Scotch, M. & Rabinowitz, P. Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. *BMC Bioinforma.* **15**, DOI: 10.1186/1471-2105-15-276 (2014).