



VCU

Virginia Commonwealth University
VCU Scholars Compass

Theses and Dissertations

Graduate School

2023

Building a macrosystems ecology framework to identify links between environmental and human health

Felisha N. Walls

Follow this and additional works at: <https://scholarscompass.vcu.edu/etd>



Part of the [Environmental Studies Commons](#)

© Felisha N. Walls

Downloaded from

<https://scholarscompass.vcu.edu/etd/7416>

This Dissertation is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

Building a macrosystems ecology framework to identify links between environmental and human health

By: Felisha Nicole Marie Walls (PhD, MS)

A Dissertation

Submitted in Partial Fulfillment

Of the Requirements for the Degree of

Doctor of Philosophy in

Integrative Life Sciences

Virginia Commonwealth University

Richmond, VA

May 2023

Daniel J. McGarvey, PhD. Chair

Lesley Bulluck, PhD

Ryan A. Hill, PhD

Derek M. Johnson, PhD

David C. Wheeler, PhD

Acknowledgements

This dissertation is dedicated to my son, Joel T. Bridgeman, Jr., who provided me with the extra motivation to finish my Dissertation. You have been an amazing son throughout this process. Thank you for helping me visualize the future I aim to build for the both of us. To my parents, Bill and Alberta Walls, who have sacrificed so much for me to achieve my goals. Thank you for believing in me and for your prayers. To my siblings, LaShonda, Sherry, Erica, Vinnie, TaNesha and Bill Walls IV, who have cheered me on. Thank you for being role models and for your words of encouragement. To LeShawn, who has uplifted me, celebrated me, and created so many joyous moments with me. Thank you for your love, support, and reminding me of what I am capable of. To my advisor, Dr. Daniel McGarvey, and family, you all have gone above and beyond to ensure that my academic journey was a success. Thank you for allowing me to conduct research on topics that are important to me and allowing me to pursue my dreams. Your guidance and support have been invaluable and greatly appreciated. To my committee members, Dr. Lesley Bulluck, Dr. Ryan Hill, Dr. Derek Johnson and Dr. David Wheeler. Thank you for your academic and professional guidance. To my family and friends for being my village. Thank you for inviting Joel and I over for holidays, game nights and Sunday dinners. You all have created safe spaces for us and reminded me to focus on the bigger picture. Financial support was provided by the U.S. National Science Foundation (grant DEB-1553111), the Southern Regional Education Board, and the Integrative Life Sciences PhD program.

Table of Contents

List of Tables.....	v
List of Figures.....	vii
Abstract.....	x
Introduction.....	1
Chapter 1 Prelude.....	8
Building a macrosystems ecology framework to identify links between environmental and human health: A random forest modelling approach.....	9
Chapter 2 Prelude.....	24
A systems-level model of direct and indirect links between environmental health, socioeconomic factors, and human mortality.....	25
Chapter 3: Systems-level models accounting for racial and regional differences in links between the environment, socioeconomic factors, and human mortality.....	35

List of Tables

Chapter 1

Table 1. Standardized path coefficients for all paths shown in the final path analysis model (Fig. 3). For each path, directionality moves from the 'Predictor' variable to the 'Response' variable. Coefficients are mean averages from 100 spatial filtering iterations, with 1 s.d. shown in parentheses. Separate results are included for the three spatial filtering distances (100, 150 and 200 km). All paths are significant at $p \leq 0.05$ (mean p -value among 100 iterations) unless their respective coefficients are highlighted in italic text; for these paths, the mean $p > 0.05$.

Table 2. Summary statistics for the final path analysis model (Fig. 3). Results include the number of counties (n) selected by the random spatial filtering process, three indices of model fit (SRMR, TLI, and CFI), and coefficients of determination (R^2) for the model response variables BMMI and AAMR. Values shown are mean averages from 100 spatial filtering iterations with 1 s.d. shown in parentheses. Separate results are included for the three spatial filtering distances (100, 150 and 200 km).

Table S1. List of the 85 covariates and 2 response variables included in the six path analysis models. Covariate descriptions, classes (geologic, climatic, hydrologic, land cover, socioeconomic), units of measure, and original data source citations (see main text for complete references) are shown.

Chapter 2

Table A1. List of the 77 covariates and 2 response variables included in the path analysis model. Covariate descriptions, classes (geologic, climatic, hydrologic, land cover, socioeconomic), units of measure, national-scale ranges (minimum; median; maximum), statistical transformations (to improve normality), and original data source citations (see main text for complete references) are shown.

Chapter 3

Table 1. Summary statistics (minimum, median, maximum, and transformation) for select variables in each of the three subregions. Variables are organized by class (Human health, Geology/Morphology, Climate, Land cover/Land use, and Socioeconomic) with units of measure shown in parentheses. Separate age-adjusted mortality rate (AAMR) statistics are shown for Black and White residents. The 20 environmental and socioeconomic variables listed here were used in the subregional path analysis models of AAMR.

Table 2. Summary statistics for the six final path analysis models (see Figs. 4-6). Results include the number of counties (n) selected by the 'majority rule' spatial overlay process, three indices of model fit (SRMR, TLI, and CFI), and coefficients of determination (R^2) for the model response variable AAMR. Separate results are included for the six demographic subgroup (Black and White) \times subregions (Northeast, Southeast, and Southern Plains).

Table S1. List of the 85 covariates and 2 response variables included in the six path analysis models. Covariate descriptions, classes (geologic, climatic, hydrologic, land cover, socioeconomic), units of measure, and original data source citations (see main text for complete references) are shown.

Table S2. Results from Moran's *I* tests of spatial autocorrelation. Separate results are listed for White and Black age-adjusted morality rate in each of the three study regions.

List of Figures

Introduction

Figure i Modeling age-adjusted mortality rate as a function of socioeconomic and environmental predictors. A map of mortality rate in U.S. counties is shown at center and seven classes of predictor variables are shown as concentric rings, with the strongest predictors closest to the center.

Chapter 1

Figure 1 Spatial dimensions of the digital stream network data. Panel (A) illustrates the spatial extents of the StreamCat ‘watershed’ and ‘catchment’ scales, relative to a discrete stream segment (indicated by the heavy black line at lower right). Panel (B) illustrates the spatial querying of digital stream and river segments within a given county. In this example, all stream segments that intersect the focal county (gray polygon) were selected in the spatial query (heavy black lines) and included in county-level summary statistics for the focal county.

Figure 2 Workflow diagram of the 2-stage process used to select a subset of covariates and build the random forest model of age-adjusted mortality rate.

Figure 3 Summary plots of observed and model-predicted AAMR. Panel A demonstrates spatial (map) and statistical (histogram) distributions of AAMR values (county-level averages between 1999-2017). Panel B illustrates the spatial and statistical distribution of model residuals (predicted AAMR – observed AAMR) for the random forest model. Counties highlighted in Panel B ($n = 762$) were randomly selected for model validation (75:25 model building vs. validation data split) and are color coded to reflect model prediction residuals; counties that were not selected for model validation are shown in yellow. Panel C includes scatterplots of the model residuals versus latitude and longitude (geographic center of each of the randomly selected validation counties), and of the observed versus predicted AAMR values. The 1:1 line illustrates perfect agreement between observed and predicted values, while the dashed line illustrates the actual relationship (least squares regression) between observed and predicted AAMR.

Figure 4 Barplot of variable importance values for the top 20 predictors of AAMR. Variable importance is measured as the percent increase in mean squared error when a given variable is removed from the random forest model. Variable themes are shown in parentheses: S – socioeconomic; C – climate; H – hydrology; P – pollution; M – morphology; L – land use; G - geology. Complete variable importance results for all 76 covariates are provided in Table S1.

Figure 5 Summary of random forest modeling results when the 12 socioeconomic covariates were excluded, leaving the 64 environmental covariates. Panel A shows the distribution of model prediction residuals (predicted – observed AAMR) for independent validation data (75:25 model building vs. validation data split). Panel B shows variable importance values for the top 20 predictors of AAMR, measured as the percent increase in mean squared error. Superscript stars (*) indicate variables that were also ranked among the top 20 covariates in the complete (with socioeconomic covariates) model. Variable themes are shown in parentheses: C – climate; H – hydrology; P – pollution; M – morphology; L – land use; G - geology.

Chapter 2

Fig. 1. Metamodel of hypothesized relationships among geologic, climatic, hydrologic, land cover and socioeconomic variables. Strongest predicted effects are indicated by solid arrows. Weaker or secondary effects are indicated by dashed arrows. Ecosystem health is predicted to be a function of geologic, climatic, hydrologic, and land cover variables. Human health is predicted to be a function of all variable classes.

Fig. 2. Plots showing spatial autocorrelation in stream ecosystem health (BMMI) and human health (AAMR) as functions of neighbor distance, or spatial lag, for U.S. counties. For each of the two variables, Moran's I is plotted against a series of 50 km spatial lags. Lags shown as open circles (100-149, 150-199 and 200-249 km) were used as spatial filters ($b = 100, 150$ and 200 km) to mitigate spatial autocorrelation in the modeling data, prior to building path analysis models. Inset histogram shows the distribution of nearest neighbor distances, measured as the centroid-to-centroid distance between adjacent counties (log₁₀ scale), for all counties in the conterminous U.S.

Fig. 3. Path model diagram of age-adjusted mortality rate (AAMR) in U.S. counties. The five classes of predictor variables are indicated by color: brown = geologic; maroon = climatic; blue = hydrologic; orange = land cover; pink = socioeconomic. Variables shown in white are indices of stream ecosystem health (BMMI) and human health (AAMR). Exogenous variables are shown with an outer ring; all remaining variables are endogenous. All paths shown in the diagram are significant at $p \leq 0.05$, when a spatial filtering distance of 100 km is used. Complete path analysis results (standardized path coefficients, model fit statistics, and coefficients of determination for BMMI and AAMR) are listed for all three filtering distances ($b = 100, 150$ and 200 km) in Tables 1 and 2.

Chapter 3

Fig. 1. Ecoregion and age-adjusted mortality rate (AAMR) maps. Panel A shows the spatial boundaries of the nine aggregate ecoregions of Herlihy et al. (2008). Ecoregion labels are: WMT – Western Mountains (Northwest); XER – Xeric; NPL – Northern Plains; SPL – Southern Plains; TPL – Temperate Plains; UMW – Upper Midwest; CPL – Coastal Plain (Southeast); SAP – Southern Appalachians; NAP – Northern Appalachians (Northeast). Panel B shows the county-level distribution of total AAMR (all demographic groups combined) throughout the conterminous U.S. Panel C shows AAMR for White residents of subset counties in the Northeast, Southeast, and Southern Plains subregions. Panel D shows AAMR for Black residents in each of the three subregions. All maps are shown at a common scale. The color scheme used to illustrate AAMR data is identical in panels B-D.

Fig. 2. Metamodel of hypothesized links between five classes of covariates (geology/ morphology, climate, hydrology, land cover/land use, socioeconomic) and human health.

Fig. 3. Heatmap of covariate importance values, measured as the percent increase in mean squared error (%IncMSE) from random forest models. Each of the 85 covariates is ranked by %IncMSE (rank 1 = most important, or highest %IncMSE), with rankings shown independently (i.e., one set per column) for the six random forest models (3 subregions \times 2 demographic groups) tested in this

study: Black residents in Northeast counties (NE_Black) and White residents in Northeast counties (NE_White); Black and White residents in Southeast counties (SE_Black and SE_White, respectively); and Black and White residents in Southern Plains counties (SP_Black and SP_White, respectively). Numbers shown in each cell are the %IncMSE ranks for a given model, with darker colors indicating more important covariates.

Fig. 4. Path model diagrams of age-adjusted mortality rates (AAMR) for (A) White and (B) Black residents in Northeast counties. The four classes of predictor variables are indicated by color: yellow = geologic; blue = climatic; green = land cover; pink = socioeconomic. Variables shown in grey are indices of human health (AAMR). Exogenous variables are shown with an outer ring; all remaining variables are endogenous. Standardized path coefficients are listed along each path and all paths shown in the diagram are significant at $p \leq 0.05$. Complete path analysis results (model fit statistics and coefficients of determination) are listed in Table 2.

Fig. 5. Path model diagrams of age-adjusted mortality rates (AAMR) for (A) White and (B) Black residents in Southeast counties. The four classes of predictor variables are indicated by color: yellow = geologic; blue = climatic; green = land cover; pink = socioeconomic. Variables shown in grey are indices of human health (AAMR). Exogenous variables are shown with an outer ring; all remaining variables are endogenous. Standardized path coefficients are listed along each path and all paths shown in the diagram are significant at $p \leq 0.05$. Complete path analysis results (model fit statistics and coefficients of determination) are listed in Table 2.

Fig. 6. Path model diagrams of age-adjusted mortality rates (AAMR) for (A) White and (B) Black residents in Southern Plains counties. The four classes of predictor variables are indicated by color: yellow = geologic; blue = climatic; green = land cover; pink = socioeconomic. Variables shown in grey are indices of human health (AAMR). Exogenous variables are shown with an outer ring; all remaining variables are endogenous. Standardized path coefficients are listed along each path and all paths shown in the diagram are significant at $p \leq 0.05$. Complete path analysis results (model fit statistics and coefficients of determination) are listed in Table 2.

Building a macrosystems ecology framework to identify links between environmental and human health

By Felisha Nicole Marie Walls (PhD, MS)

A Dissertation submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy at Virginia Commonwealth University, 2023.

Major Director: Dr. Daniel J. McGarvey, Center for Environmental Studies

Abstract

The structure and functions of natural systems continue to be degraded by human activities such as land-use change. One potential consequence that has received relatively little attention is a corresponding decline in human health. Altered ecosystems can present a host of risks to human health. Yet the consequences of environmental degradation are rarely considered in regional to national-scale research on human well-being. The guiding objective of this work is to compare an array of stream health and socioeconomic metrics with spatially congruent human health metrics throughout the conterminous United States, then to identify potential links between environmental health, socioeconomic factors and human health. Environmental health metrics have been obtained from the Environmental Protection Agency's Stream-Catchment dataset; these data include natural and anthropogenic watershed characteristics for 2.65 million stream and river segments within the U.S. Human health metrics have been obtained from the Center for Disease Control and Prevention Wide-ranging Online Data for Epidemiologic Research database. Socioeconomic data were downloaded from four sources and represents key

behavioral, social and economic determinants of human health outcomes. In Chapter 1, random forest modeling was used to identify key predictors of human health, expressed as county-level, age-adjusted mortality rates, then to predict county-level mortality rate as a function of these covariates. In Chapter 2, path analysis models of human mortality were built to explore direct and indirect pathways between the environment, ecosystem health and human health to reveal potential cause-and-effect pathways at macrosystems scales. Finally, in Chapter 3, the random forest and path analysis modeling techniques used in Chapters 1 and 2 were repurposed to model racial and regional differences in the multi-step pathways that link environmental and socioeconomic factors to AAMR.

Introduction

Previous studies have shown that human activities, such as land use change, negatively impact natural environments (Vitousek et al., 2008). However, little is known about how these environmental disturbances may, in turn, have important consequences for human health. This is because the effects of environmental degradation are rarely considered in research on human health and well-being. When they are, studies are often focused on a specific pollutant or conducted at a specific site.

For instance, significant associations have been reported between particulate matter air pollution and increased risk of stroke, heart disease, and lung disease (Krall et al., 2013; Ostro et al., 2007; Zanobetti & Schwartz, 2009). Higher rates of human mortality and morbidity have been attributed to permitted industrial water pollution discharge (Hendryx et al., 2012; Jian et al., 2017; Ren et al., 2014). Additionally, trends in health disparities have been observed across sociodemographic conditions, such as education, race and income (Cullen et al., 2012; Davids et al., 2014; Jian et al., 2017). Although, these focused based studies have improved our understanding of how human health may respond to environmental disturbance, their findings cannot always be generalized to address concerns at larger scales. This is alarming because the ongoing changes in the environment are expected to have widespread effects on human health.

One way to improve our understanding of the dynamic connections between human and ecosystem health may be to integrate human health data in macrosystems ecology research. Defined as “the integrative study of biological, geophysical and social systems at regional to national scales (Heffernan et al., 2014). Macrosystems research combines traditional tools used in ecology, with new technological advances in the collection and analysis of large data, such as

machine learning (Dodds et al., 2021). It can provide opportunities to investigate the effects of environmental change and socioeconomic factors on human health and well-being (Tromboni et al., 2021).

To understand how direct and indirect exposure to a collection of environmental factors may influence human health outcomes, I examined the impacts of environmental health and socioeconomic factors on human mortality rates throughout the United States. Freshwater ecosystems were used to highlight the complex interactions between health and the environment for two reasons (Dunham et al., 2018). First, humans rely on services provided by freshwater systems, such as drinking water and climate regulation (R. De Groot et al., 2002; Millennium Ecosystem Assessment, 2005; D. Thornbrugh et al., 2018). Second, data characterizing human-induced stressors to stream health are well documented (Carlisle et al., 2011; Esselman et al., 2011). I also recognize that socioeconomic factors often act as intermediaries of the indirect links between humans and the environment (Evans & Kantrowitz, 2002).

In this dissertation research I created a macrosystems ecology framework to identify direct and indirect links between the environment, socioeconomic factors and human health within the conterminous United States (CONUS). As an indicator of human health, the study focused primarily on age-adjusted mortality rates (AAMR) for all counties in the CONUS (<https://wonder.cdc.gov/>). Environmental data represented a range of anthropogenic, climatic, geologic, hydrologic, and land cover influences on local streams and rivers (Hill et al., 2016). Socioeconomic data contained key behavioral, social and economic predictors of human health and well-being (Sanchez et al., 2014; Wheeler et al., 2019).

Chapter 1 describes the random forest modeling process used to: (i) identify a subset of non-redundant environmental and socioeconomic (social and economic) covariates that are

strongly associated with county-level AAMR; and (ii) predict AAMR as a function of these covariates. Results determined that socioeconomic factors were the most influential determinates of human mortality in the U.S (Figure i). In particular, smoking, food insecurity, and lack of physical activity were most important. Furthermore, air temperature and precipitation, were most important among environmental variables.

Chapter 2 applied the environmental, socioeconomic and human health data from Chapter 1 and an independent measure of ecosystem health within a path analysis (PA) model. In Chapter 2, the primary goal was not to make predictions, but to build a systems-level path diagram of direct and indirect effects on human and ecosystem health. PA models revealed: (i) geology and climate plays a foundational role in regulating many of the indirect effects on human and ecosystem health; (ii) hydrology and land cover have distinct effects on human and ecosystem health; and (iii) forest cover is a key link between the environment and the socioeconomic variables that directly influence human health. In Chapter 3, we used the random forest and path analysis modeling techniques outlined in Chapters 1 and 2 to assess differences in environmental and socioeconomic influences on human health outcomes among Black and White populations within three subregions (Northeast, Southeast, and Southern Plains) of the CONUS. I found differences among regions and races, including: (1) socioeconomic variables, smoking and lack of physical activity, consistently had the strongest influences on AAMR; (2) land cover variables were key determinants of education and income; (3) air temperature and precipitation played an key role in regulating many of the direct and indirect effect on all environmental, socioeconomic and human health variables.

I encourage future studies to pursue integrative research on the diverse effects that socioeconomic and environmental influences have on human health. By combining large, mixed

datasets with modern machine learning tools, the research community can begin to anticipate human responses to changing conditions and, hopefully, take proactive steps to mitigate the most damaging outcomes.

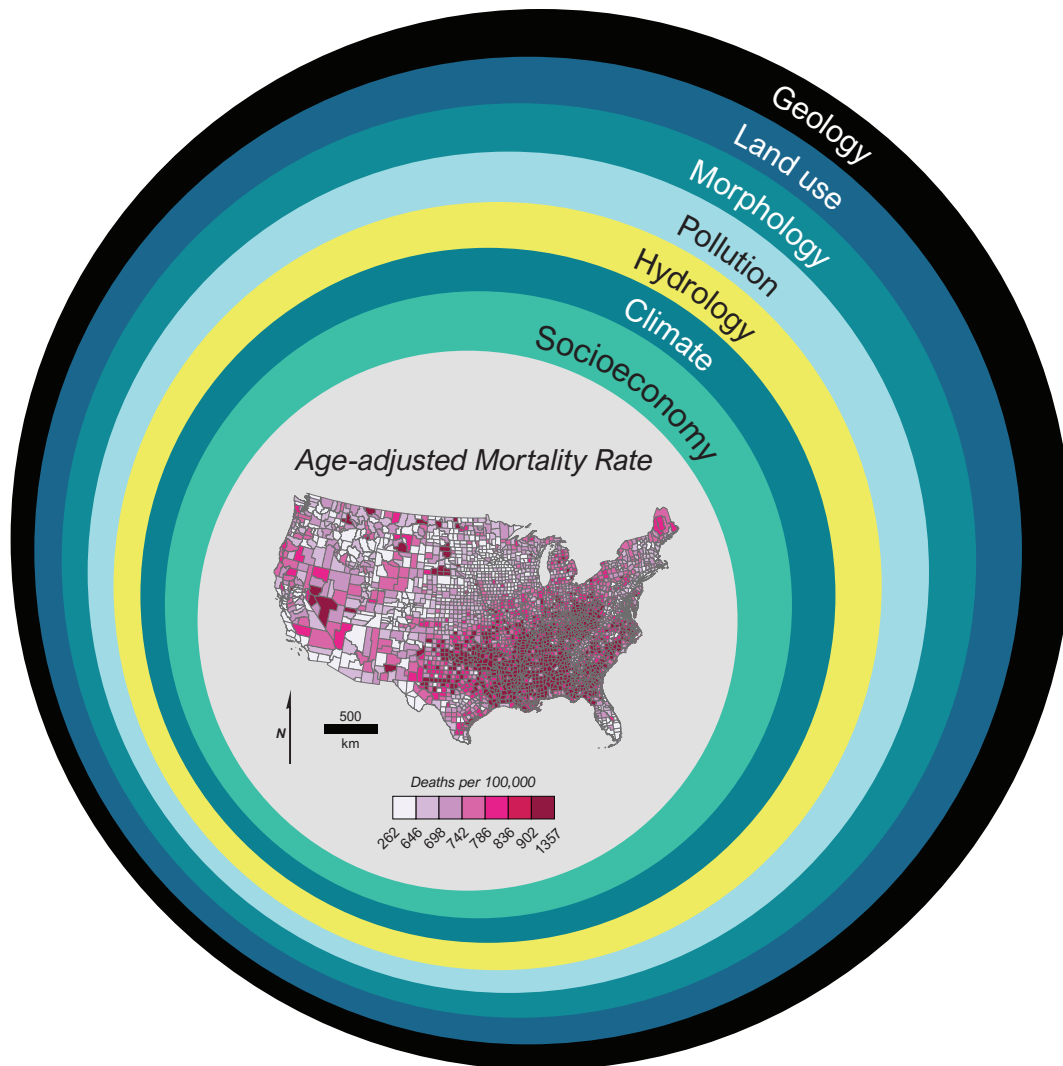


Figure i. Modeling age-adjusted mortality rate as a function of socioeconomic and environmental predictors. A map of mortality rate in U.S. counties is shown at center and seven classes of predictor variables are shown as concentric rings, with the strongest predictors closest to the center.

References

- Carlisle, D., Wolock, D., & Meador, M. (2011). Alteration of streamflow magnitudes and potential ecological consequences: A multiregional assessment. *Frontiers in Ecology and the Environment*, 9(5), 264–270.
- Cullen, M. R., Cummins, C., & Fuchs, V. R. (2012). Geographic and Racial Variation in Premature Mortality in the U . S .: Analyzing the Disparities. *PLoS ONE*, 7(4).
- Dauids, B.-O., Hutchins, S. S., Jones, C. P., & Hood, J. R. (2014). Disparities in life expectancy across US counties linked to county social factors, 2009 Community Health Status Indicators (CHSI). *Journal of Racial and Ethnic Health Disparities*, 1(1), 2–11.
- De Groot, R., Wilson, M., & Boumans, R. (2002). A typology for the classification, description and valuation of ecosystem functions, good and services. *Ecological Economics*, 41, 393–408.
- Dodds, W. K., Rose, K. C., Fei, S., & Chandra, S. (2021). Macrosystems revisited: challenges and successes in a new subdiscipline of ecology. *Frontiers in Ecology and the Environment*, 19(1), 4–10.
- Dunham, J., Angermeier, P., Crausbay, S., Cravens, A., Gosnell, H., Mcevoy, J., Moritz, M., Raheem, N., & Sanford, T. (2018). Rivers are social – ecological systems : Time to integrate human dimensions into riverscape ecology and management. *WIREs Water*, 1–10.
- Esselman, P., Infante, D., Wang, L., Wu, D., Cooper, A., & Taylor, W. (2011). An index of cumulative disturbance to river fish habitats of the Conterminous United States from landscape anthropogenic activities. *Ecological Restoration*, 29(1–2), 133–151.
- Evans, G. W., & Kantrowitz, E. (2002). Socioeconomic status and health: the potential role of environmental risk exposure. *Annual Review of Public Health*, 23(1), 303–331.

- Heffernan, J. B., Soranno, P. A., Angilletta, M. J., Buckley, L. B., Gruner, D. S., Keitt, T. H., Kellner, J. R., Kominoski, J. S., Rocha, A. V., Xiao, J., Harms, T. K., Goring, S. J., Koenig, L. E., McDowell, W. H., Powell, H., Richardson, A. D., Stow, C. A., Vargas, R., & Weathers, K. C. (2014). Macrosystems ecology: Understanding ecological patterns and processes at continental scales. *Frontiers in Ecology and the Environment*, *12*(1), 5–14.
- Hendryx, M., Conley, J., Fedorko, E., Luo, J., & Armistead, M. (2012). Permitted water pollution discharges and population cancer and non-cancer mortality : toxicity weights and upstream discharge effects in US rural-urban areas. *International Journal of Health Geographics*, *11*(9), 1–15.
- Hill, R., Weber, M., Leibowitz, S., Olsen, A., & Thornbrugh, D. (2016). The Stream-Catchment (StreamCat) Dataset: A Database of watershed metrics for the Conterminous United States. *Journal of the American Water Resources Association*, *52*(1).
- Jian, Y., Messer, L., Jagai, J., Rappazzo, K., Gray, C., Grabich, S., & Lobdell, D. (2017). Associations between environmental quality and mortality in the contiguous United States, 2000–2005. *Environmental Health*, *125*(3), 2000–2005.
- Krall, J. R., Brooke Anderson, G., Dominici, F., Bell, M. L., & Peng, R. D. (2013). Short-term exposure to particulate matter constituents and mortality in a national study of U.S. urban communities. *Environmental Health Perspectives*, *121*(10), 1148–1153.
- Millennium Ecosystem Assessment. (2005). Ecosystems and human well-being: Health synthesis. In *World Health Organization* (Vol. 18, Issue 3).
- Ostro, B., Feng, W. Y., Broadwin, R., Green, S., & Lipsett, M. (2007). The effects of components of fine particulate air pollution on mortality in California: Results from CALFINE. *Environmental Health Perspectives*, *115*(1), 13–19.

- Ren, H., Wan, X., Yang, F., Shi, X., Xu, J., Zhuang, D., & Yang, G. (2014). Association between changing mortality of digestive tract cancers and water pollution: A case study in the Huai River Basin, China. *International Journal of Environmental Research and Public Health*, *12*(1), 214–226.
- Sanchez, G. M., Nejadhashemi, A. P., Zhang, Z., Woznicki, S. A., Habron, G., Marquart-Pyatt, S., & Shortridge, A. (2014). Development of a socio-ecological environmental justice model for watershed-based management. *Journal of Hydrology*, *518*, 162–177.
- Thornbrugh, D., Hill, R., Leibowitz, S., Hill, R., Weber, M., Johnson, Z., Olsen, A., Flotemersch, J., Stoddard, J., & Peck, D. (2018). Mapping watershed integrity for the conterminous United States. *Ecological Indicators*, *85*(January 2018), 1133–1148.
- Tromboni, F., Liu, J., Ziaco, E., Breshears, D. D., Thompson, K. L., Dodds, W. K., Dahlin, K. M., LaRue, E. A., Thorp, J. H., Viña, A., Laguë, M. M., Maasri, A., Yang, H., Chandra, S., & Fei, S. (2021). Macrosystems as metacoupled human and natural systems. *Frontiers in Ecology and the Environment*, *19*(1), 20–29.
- Vitousek, P. M., Mooney, H. A., Lubchenco, J., & Melillo, J. M. (2008). Human domination of Earth's ecosystems. *Urban Ecology: An International Perspective on the Interaction Between Humans and Nature*, *277*(July), 3–13.
- Wheeler, D. C., Jones, R. M., Schootman, M., & Nelson, E. J. (2019). Explaining variation in elevated blood lead levels among children in Minnesota using neighborhood socioeconomic variables. *Science of the Total Environment*, *650*, 970–977.
- Zanobetti, A., & Schwartz, J. (2009). The effect of fine and coarse particulate air pollution on mortality: A national analysis. *Environmental Health Perspectives*, *117*(6), 898–903.

Chapter 1 Prelude

Anthropogenic disturbances to natural ecosystems can also impact human health. However, systematic research on the links between human and environmental health is at relatively early stage. In Chapter 1, a large dataset of environmental and socioeconomic variables is assembled. Random forest modeling is then used to predict human mortality rate as a function of these variables and to determine which of the environmental and socioeconomic influences may have the strongest influences on mortality rate. The random forest model identifies smoking, food insecurity, and lack of physical activity as the most influential socioeconomic predictors of mortality rate. However, environmental variables, particularly air temperature and precipitation, were also found to be important predictors of human health. This research was published in 2023, in the journal *People and Nature* (Walls, F., and D. McGarvey. 2023. Building a macrosystems ecology framework to identify links between environmental and human health: A random forest modelling approach. *People and Nature*, 5:183-197. DOI: 10.1002/pan3.10427).

RESEARCH ARTICLE

Building a macrosystems ecology framework to identify links between environmental and human health: A random forest modelling approach

Felisha N. Walls  | Daniel J. McGarvey

Virginia Commonwealth University,
Richmond, Virginia, USA

Correspondence

Felisha N. Walls
Email: wallsfn@vcu.edu

Funding information

National Science Foundation, Grant/
Award Number: DEB 1553111

Handling Editor: Laura Graham

Abstract

1. Anthropogenic activities that degrade natural ecosystems may also impact human health. However, research on the links between human and environmental health has most often been conducted at small scales (e.g. individual cities) and cannot easily be extrapolated to larger scales.
2. We created a macrosystems ecology framework to identify associations between human and environmental health by combining human mortality and socioeconomic data for the conterminous United States with spatially aligned data on the physicochemical characteristics of river basins.
3. Principal component analysis was first used to reduce a list of 596 environmental variables to a subset of 64 environmental covariates, representing six main environmental themes (climate, geology, hydrology, land use, river basin morphology and pollution). Independent, spatially aligned information was then obtained for 12 socioeconomic covariates.
4. Random forest modelling was used to predict age-adjusted mortality rate as a function of the environmental and socioeconomic covariates. An independent data subset (random 75:25 model building vs. testing split) was also used for model validation. The coefficient of determination between predicted and observed mortality rates was 0.76 for the validation data. Furthermore, model residuals (predicted – observed mortality) were centered near zero and normally distributed (1 SD = 62.26), suggesting high model accuracy and precision.
5. Socioeconomic covariates were consistently the most influential predictors of mortality rate. Smoking, food insecurity, and lack of physical activity were particularly important. However, environmental covariates accounted for 5 of the 10 strongest predictors overall, with air temperature and precipitation being most influential among environmental variables.
6. This proof-of-concept study demonstrates the utility of a modelling framework that combines environmental and human health data at macroscales. We suggest

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *People and Nature* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

that further application of macrosystems ecology tools will improve the capacity to anticipate human health responses to ongoing environmental change.

KEYWORDS

age-adjusted mortality rate, air temperature, centers for disease control WONDER database, environmental effects, precipitation, principal component analysis, socioeconomic effects, stream catchment database

1 | INTRODUCTION

Human activities have dramatically impacted natural ecosystems across the globe (Vitousek et al., 1997). These ecosystem disturbances may, in turn, have important consequences for human health. For instance, significant associations between particulate matter air pollution and increased risk of stroke, heart disease and lung disease have been documented (Krall et al., 2013; Ostro et al., 2007; Zanobetti & Schwartz, 2009). Similarly, permitted industrial water pollution has been linked to higher rates of human mortality and morbidity (Hendryx et al., 2012; Jian et al., 2017; Ren et al., 2014). These types of direct, context-specific epidemiological connections are critical for understanding how human health may respond to environmental disturbance.

However, insights gained through research on direct or secondary exposure to discrete pollutants such as airborne particulates or industrial wastewater cannot always be generalized to address health concerns at larger scales. This is concerning because gradual changes in the environment may also have widespread effects on human health. Increasing numbers of hospital admissions due to cardiovascular or respiratory complications are a case in point: they may be aggravated, but not directly caused by increasing air temperatures (Guirguis et al., 2014; Knowlton et al., 2009; Meehl & Tebaldi, 2004).

One strategy to quantify and explain dynamic connections between human and ecosystem health may be to integrate human health data in macrosystems ecology research. Defined as 'the study of diverse ecological phenomena at the scale of regions to continents and their interactions with phenomena at other scales' (Heffernan et al., 2014), macrosystems ecology can provide a nexus to integrate large-scale data on ecosystem structure and function with data on human behaviour and well-being. Indeed, several authors have recently advocated for the inclusion of social factors, such as human health, in macroscale research (Burnside et al., 2012; Fleishman & Brown, 2019; Leitão et al., 2019).

Freshwater ecosystems provide an ideal context for studying macrosystems connections between human health and the environment for at least two reasons (Corley et al., 2018; Dunham et al., 2018). First, freshwater ecosystems provide humans with many ecosystem services. These include resource provisioning services (e.g. food and drinking water), regulating services (e.g. climate regulations and flood mitigation), supporting services (e.g. riparian wildlife habitat), and cultural services (e.g. recreation and aesthetics) (De Groot et al., 2002; MEA, 2005; Thornbrugh

et al., 2018). However, human activities are rapidly degrading the integrity of freshwater ecosystems in the United States (Thornbrugh et al., 2018). Due largely to the prevalence of industrial agriculture (e.g. crop production, animal feeding operations), greenhouse gas emissions (e.g. transportation, electricity production) and hydrologic modifications (e.g. dams, channelization), 46% of all streams and rivers within the United States are estimated to be in 'poor' biological condition (United States Environmental Protection Agency, 2017).

One potent example of the negative impact that industrial activities may have on freshwater ecosystems, and by extension, human health is coal mining in the Appalachian region of the United States. Valley-fill disposal of mountaintop mining waste buries and eliminates headwater streams that play a critical role in ecological processes such as nutrient cycling, stream biodiversity and water quality (Allan, 2004; Meyer et al., 2007; Palmer et al., 2010). Loss of these ecological functions may then exacerbate human health concerns in Appalachian coal mining communities, where rates of heart disease, kidney disease and mortality are among the highest in the nation (Hendryx, 2009; Hendryx & Ahern, 2009). Widespread degradation of Appalachian streams and landscapes is also linked to depression in local communities that have deep cultural ties to the mountains (Hendryx & Innes-Wimsatt, 2013).

Industrial agriculture is another example of a macrosystems link between human activity, freshwater systems, and human health (Corley et al., 2018). Agricultural runoff delivers significant quantities of nutrients, pesticides, sediment, and bacteria to US streams and rivers. Widespread pollutants are, in turn, associated with impaired freshwater communities (Pimentel et al., 1997; Roth et al., 1996) as well as elevated cancer risk in humans (Horrihan et al., 2002). These types of pervasive freshwater disturbances create a logical precedent to search for statistical correlations between indicators of human and freshwater ecosystem health.

A second reason to focus on freshwater ecosystems is the large research investment that has already been made to document and understand the health of these systems at regional or macro-scales. For instance, Carlisle et al. (2011) developed RF models to quantify stream-flow alteration and its potential ecological impacts at 2888 monitoring sites throughout the conterminous United States. They found that anthropogenic stream alterations were a major threat to the structure and function of stream and river ecosystems, particularly in arid climates where exceptionally high rates of water consumption are driven by agricultural irrigation (Carlisle et al., 2011; Poff & Zimmerman, 2010; Resh et al., 1988). Alternatively, Esselman et al. (2011) used a partial constrained ordination to develop an index of cumulative anthropogenic

disturbance to fish habitats while leveraging biological integrity data for approximately 2.23 million river reaches in the US. Their work generated a gradient of disturbance effects on fish habitat, beginning with urban development (most severe effect) and progressing through point-source pollution, pasture lands and dams (least severe effects) (Esselman et al., 2011; Roth et al., 1996).

This study builds a macrosystems framework to study connections between the health of freshwater ecosystems and human populations within the United States. As an indicator of human health, the study focuses on human mortality rate. Average life expectancy in the United States increased through much of the 20th century but stagnated in 2010, then entered a declining phase by 2014 (Woolf et al., 2018; Woolf & Aron, 2018). Now, despite \$3 trillion in annual health care expenditures, Americans are living shorter, less healthy lives (Muennig & Glied, 2010; The US Burden of Disease Collaborators, 2018). This recent inversion in the life expectancy trend underscores the need to identify environmental factors that covary with mortality rate over similar time periods.

Our primary objective is to model mortality rate as a function of the physicochemical characteristics of rivers and their surrounding basins. In essence, we address the question: 'can a large suite of environmental variables predict human mortality with a meaningful level of accuracy and precision?' However, we recognize that socioeconomic factors often act as intermediaries between the environment and human health (Evans & Kantrowitz, 2002). For instance, increased access to greenspace can positively influence physical and mental health, but greenspace is itself a function of forest cover and urban development (Engemann et al., 2019; Frumkin et al., 2017). We, therefore, use a list of socioeconomic variables, each of which is known to impact human health, in combination with the environmental variables. We then compare a human mortality model that includes environmental, as well as socioeconomic predictors with a model that includes only environmental predictors. By doing so, we provide new insight to the complex interrelationships that link environmental and social factors to human health.

Specific study goals are to (1) combine national-scale data on the environmental characteristics of freshwater ecosystems with national-scale data on socioeconomic conditions and human mortality; (2) select a relatively small subset of environmental variables from a much larger database, seeking to minimize collinearity while retaining seven major themes of information (climate, geology, hydrology, land use, river basin morphology, pollution and socioeconomic) and (3) use a RF algorithm to model human mortality as a function of environmental and socioeconomic variables.

2 | MATERIALS AND METHODS

2.1 | Data collection and management

Mortality data were obtained from the Underlying Cause of Death: Detailed Mortality database, accessible through the US Centers for Disease Control and Prevention, Wide-ranging Online Data for

Epidemiologic Research website (<https://wonder.cdc.gov/>; downloaded February 2020). We focused exclusively on age-adjusted mortality rates (AAMRs) for all counties in the United States, between 1999 and 2017. This time interval brackets the inverted life expectancy trend noted in the Introduction (Woolf et al., 2018; Woolf & Aron, 2018). AAMR is the weighted average of age-specific death rates within a given county. Weighting by age accounts for the uneven age distribution among counties. In counties with fewer than 10 annual mortality events, the Centers for Disease Control withholds AAMR data to ensure confidentiality. Our analyses were, therefore, limited to counties with ≥ 10 annual deaths (Jian et al., 2017).

Environmental data were obtained from two online sources. First, the 1:100,000 scale National Hydrography Dataset Plus (NHD), Version 2 digital stream network (Mckay et al., 2012; downloaded January 2020) was used to build a spatial model of all stream and river segments within the United States. NHD attribute tables included six physical variables (e.g. stream order and channel slope) that were used as potential environmental covariates in AAMR models. Second, environmental variables were queried from the US Environmental Protection Agency Stream-Catchment (StreamCat) database (Hill et al., 2016; downloaded January 2020). StreamCat is a large compilation of anthropogenic, geologic, land use, and climatic variables that were 'mapped' to the NHD at two distinct spatial scales. The watershed scale is a regional representation of the entire, cumulative landscape that lies upstream of a given stream/river segment (Figure 1a). The catchment scale is a local representation of the landscape immediately adjacent to a given stream/river segment.

Watershed and catchment scale variables were both included as potential predictors of AAMR to obtain a holistic view of the effects that natural and anthropogenic features may have on freshwater ecosystems and human health (Allan, 2004; Hill et al., 2017; Thornbrugh et al., 2018). This was prudent because the drivers of ecosystem and human health are dynamic and can operate across a range of spatial scales (Allan, 2004; Flotemersch et al., 2016; Hill et al., 2016; Roth et al., 1996; Tsang et al., 2014). For example, Lewis et al. (2019) showed the negative effects of faecal-derived microbial pollutants within the Olema Creek Watershed in California are cumulative and most effectively predicted at the watershed scale. Conversely, Bale et al. (2017) found that water pollutant loads from low-density residential neighbourhoods in California vary from site-to-site and that catchment scale estimates are more useful.

StreamCat variables were downloaded from the master StreamCat database (Hill et al., 2016) for all digital stream and river segments within the United States, using a custom script (provided in Supporting Information) written in the R programming language (R Development Core Team 4.3.0, 2022). StreamCat data were then joined to the NHD data using unique 'COMID' index codes for every stream/river segment in the NHD database. In total, 109 unique predictor variables were queried for each NHD stream/river segment. All variables are defined in McKay et al. (2012) and Hill et al. (2016).

Next, the NHD river network data were superimposed on a shapefile of US counties, within a geographic information system.

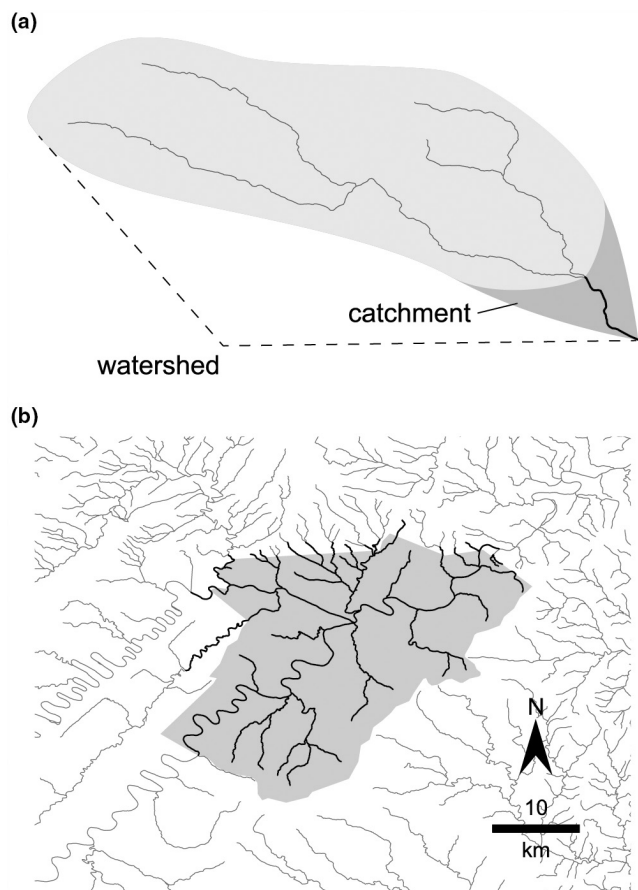


FIGURE 1 Spatial dimensions of the digital stream network data. Panel (a) illustrates the spatial extents of the StreamCat ‘watershed’ and ‘catchment’ scales, relative to a discrete stream segment (indicated by the heavy black line at lower right). Panel (b) illustrates the spatial querying of digital stream and river segments within a given county. In this example, all stream segments that intersect the focal county (grey polygon) were selected in the spatial query (heavy black lines) and included in county-level summary statistics for the focal county.

Spatial queries of all stream and river segments within a given county (multiple NHD segments occurred within each county; see [Figure 1b](#)) were performed with ArcGIS 10.5 software (Environmental Systems Research Institute, Redlands, California). Three county-level summary statistics (minimum, median, and maximum) were then calculated with a custom R script (see Supporting Information) for each of the environmental variables. This process generated 596 predictor variables for each county (3 summary statistics for each NHD variable; 3 summary statistics \times 2 spatial scales for each StreamCat variable). All environmental data were matched to their respective counties using Federal Information Processing Standard county codes.

Twelve socioeconomic variables were also selected, based upon prior human health research, and downloaded (December 2021) for each county. Prevalence of adult smoking, physical inactivity, food insecurity, violent crime and severe housing shortages, as well as the degree of rurality (Little, 1999) and social association rate,

were obtained from the University of Wisconsin Population Health Institute, County Health Rankings and Roadmaps project (www.countyhealthrankings.org). Each of these is known to have a direct or secondary effect on human health and mortality rates (Luken, 2022; Paffenbarger et al., 1993; Sun et al., 2020; Wang & Preston, 2009; Wilkinson et al., 1998; Yang et al., 2011). Prevalence of a college (bachelor’s) degree, access to health insurance and renter occupied housing were obtained from the US Census Bureau, American Community Survey (www.census.gov/programs-surveys/acs). We used the most recent 5-year estimates (2013–2017). These variables are regularly used in epidemiological research and are highly relevant to disadvantaged populations (Sampson, 1997; Sanchez et al., 2014; Wheeler et al., 2019). The leaf area index (LAI), or surface area of green vegetation relative to total surface area (2011–2015 remote sensing data), was obtained from Russette et al. (2021) and used as a surrogate for local greenspace, which tends to promote physical activity and psychological well-being (Engemann et al., 2019; Frumkin et al., 2017). Finally, the overall rate of participation in organized religion, defined as the number of individuals (per 1000 county residents) of any denomination or faith who regularly congregate for worship at a church or other community center, was obtained from Grammich et al. (2012). Similar to greenspace, religious activity may be positively correlated with human health (Sissamis et al., 2022).

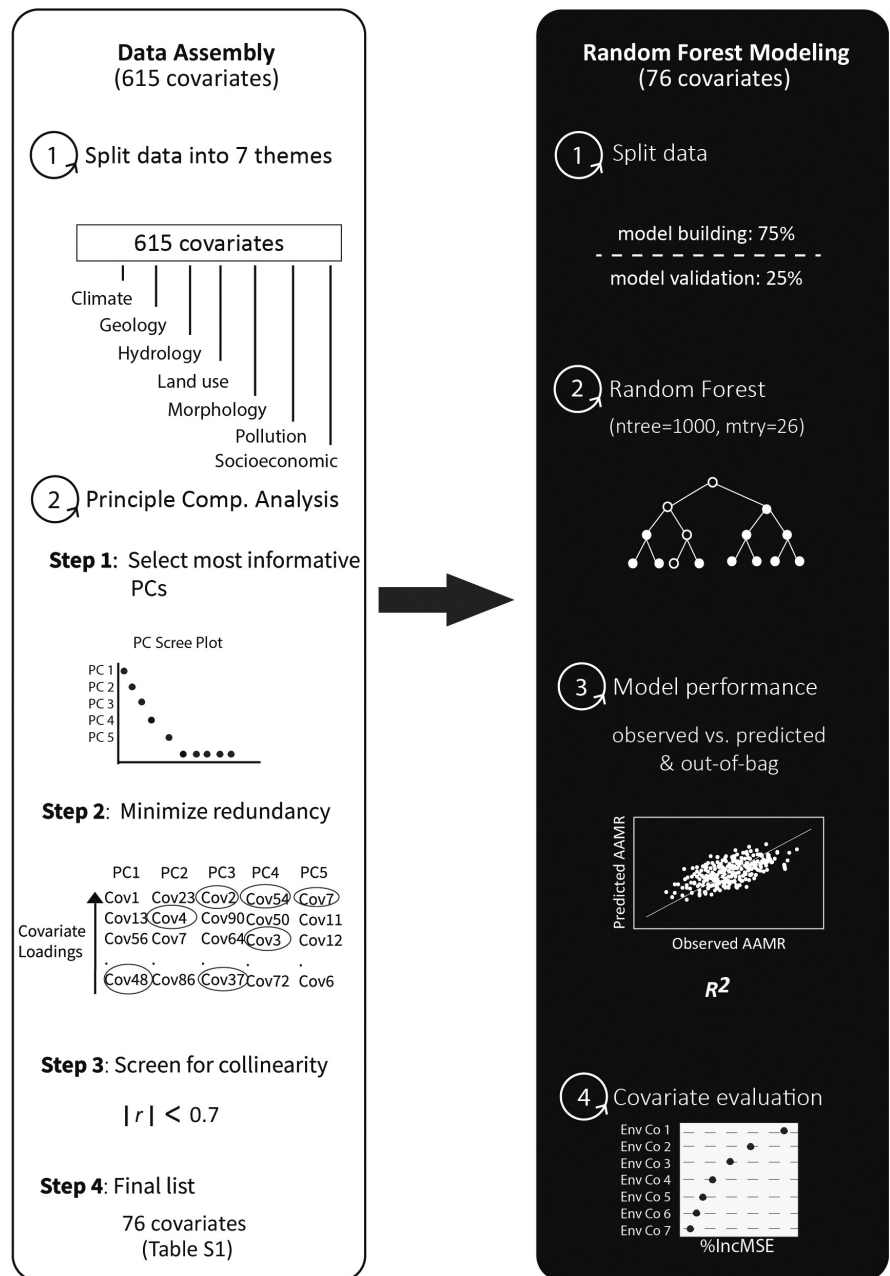
As a final data preparation step, imputation was used to replace missing values within the combined set of predictor variables. Among all county \times variable entries, approximately 0.2% were missing values. Imputed estimates were generated with the `MISSFOREST` package in R (Stekhoven & Bühlmann, 2012). Briefly, `missForest` imputes missing values for a focal variable using all remaining variables within a dataset as predictors in a RF model (see Section 2.2.2 below for additional information on RF modelling). The process is repeated independently for each variable (i.e. each variable with missing values is iteratively treated as the focal variable) until no missing values remain in the dataset. Default settings were used with the `missForest` function. All data and complete R code to repeat our process is provided in Supporting Information.

2.2 | Modelling age-adjusted mortality rate

A 2-stage process was used to (i) reduce complexity and collinearity within the large set of environmental variables, then (ii) model AAMR (see [Figure 2](#)). First, principal component analysis (PCA) was used to identify a subset of environmental variables that represented much of the information encoded in the complete set of 596 environmental variables, while minimizing collinearity. This was necessary because a preliminary examination of the environmental data revealed significant collinearity ($|r| \geq 0.7$) among many of the environmental variables. PCA was not performed for the relatively small set of 12 socioeconomic variables, as significant collinearity was not observed among them.

Second, random forest (RF) modelling was used to predict AAMR throughout the United States. RF is a machine learning algorithm

FIGURE 2 Workflow diagram of the two-stage process used to select a subset of covariates and build the random forest model of age-adjusted mortality rate.



that builds a ‘forest’ of decision trees, each of which independently predicts a categorical or continuous output. An aggregate, forest-level prediction is then calculated as the average of all individual tree predictions. This ensemble approach makes RF robust to outliers, minimizes overfitting, and generally improves the accuracy of model predictions (Breiman, 2001b; Cutler et al., 2007; Fox et al., 2017). The combined PCA and RF modelling process is detailed below.

2.2.1 | Principal component analysis

To reduce the number of environmental variables subsequently used in RF models, we adapted the data reduction process of Olden and Poff (2003). Specifically, we identified PCs within the n -dimensional

environmental data and used the PC loadings to select individual variables that were effective surrogates for the PCs. We began by partitioning the 596 environmental variables among six themes: climate ($n = 24$), geology/lithology ($n = 218$), hydrology ($n = 42$), land use ($n = 230$), river basin morphology ($n = 32$) and pollution ($n = 50$). A separate PCA was then performed for each of the six themes, using the ‘prcomp’ function in R.

For each of the six environmental themes, we used a PC scree plot (percent variance explained by each variable, in rank order) to select the most informative PCs. The number of select PCs was determined by visually identifying the start of asymptotic behaviour in the scree plot; PCs to the left of the asymptote were retained. Next, PC loading scores were used to identify environmental variables that were most strongly associated with each of the retained PCs. These

variables were included in the final RF dataset. At each step, Pearson correlation values were used to ensure that significantly collinear variables were not selected. Several additional variables that were predicted to have a direct or indirect effect on AAMR but were not collinear with previously selected variables, were then manually selected from each theme. Finally, the selected variables from each theme were combined and used to build a master correlation matrix, ensuring that none of the retained variables exhibited significant collinearity. This process resulted in a final list of 76 covariates (64 environmental + 12 socioeconomic; see Table S1). R code to repeat the PC analyses and extract the PC loadings is provided in Supporting Information.

2.2.2 | Random forest modelling

An RF model of AAMR was built with the refined list of 76 covariates. Prior to modelling, we used a random 75:25 data split to create independent model building (75% of the data) and model validation (25% of the data) datasets. Noting that RF model performance increases without penalty as the number of trees in a forest increases (Oshiro et al., 2012), we included an arbitrarily large number of trees, $n_{tree} = 1000$, in the RF model. The number of covariates tested at each tree node was set as $m_{try} = 26$, consistent with the standard regression tree setting, $m_{try} \approx p/3$ (Liaw & Wiener, 2002; Wheeler et al., 2015). Modelling was performed with the 'RANDOMFOREST' package in R (Liaw & Wiener, 2002).

Model performance was assessed by comparing the model predicted AAMR value with the observed AAMR for each of the randomly selected validation counties ($n = 762$). A strong linear relationship between predicted and observed AAMR was interpreted as evidence of an effective model. We also examined the out-of-bag results from the RF model building process. As each tree within an RF is grown, the RF algorithm uses a bootstrap (i.e. sampling with replacement) procedure to select a random subset of the available data for model fitting. This minimizes the risk of overfitting and provides an independent subset of data—the out-of-bag replicates that were not included in the bootstrap sample—that can be used for model validation (Breiman, 2001a). Model-predicted AAMR values were compared with the observed AAMR values for each of the out-of-bag samples. We also recorded the percent increase in mean squared error (%IncMSE) statistic for each of the 76 covariates, then used the %IncMSE values to identify covariates that had the strongest influences on predicted AAMR.

Finally, we removed the 12 socioeconomic variables from the complete list of 76 covariates and repeated the RF modelling process, using only the 64 environmental variables ($n_{tree} = 1000$, $m_{try} = 22$). By comparing RF models with and without the socioeconomic variables, we sought to distinguish environmental variables that may be conflated with one or more of the socioeconomic variables from environmental variables that may represent independent influences on AAMR. R code to repeat all RF procedures is provided in Supporting Information.

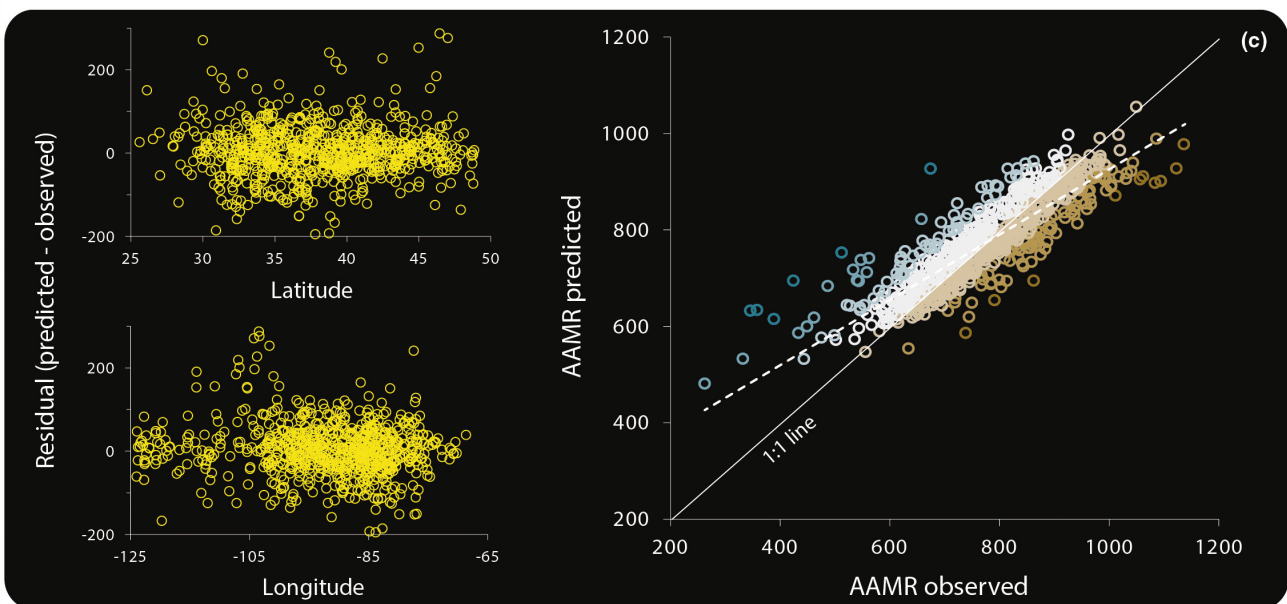
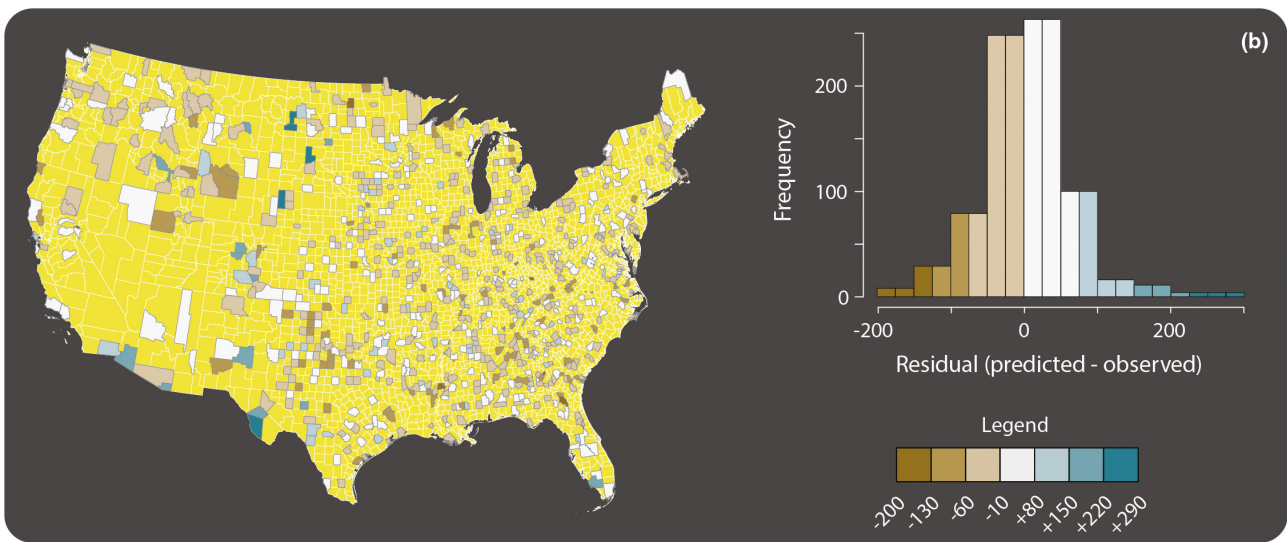
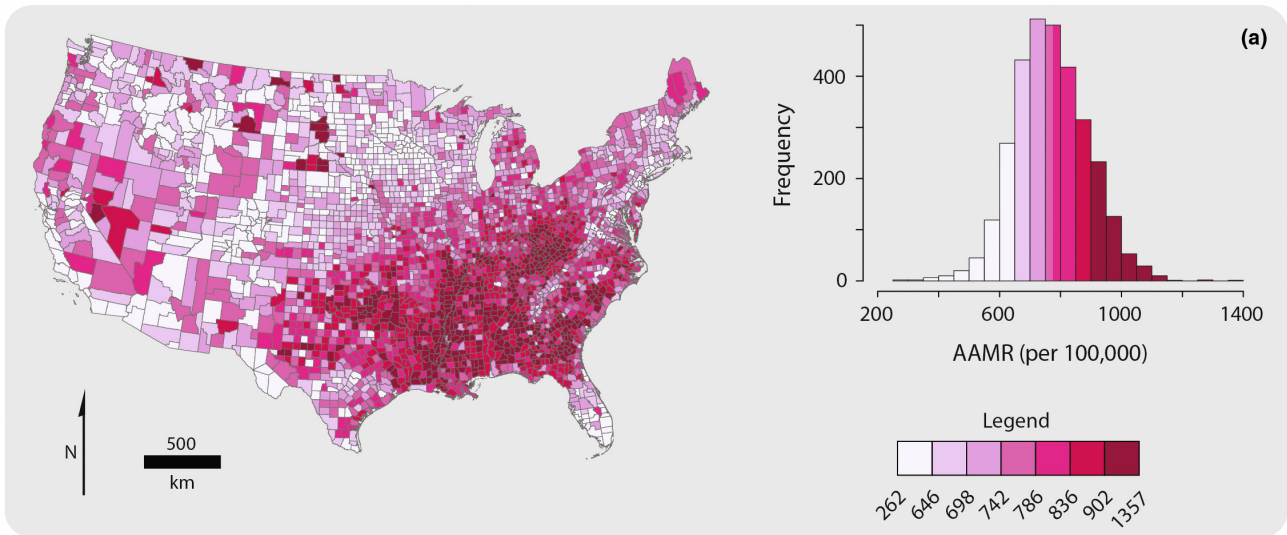
3 | RESULTS

Across the United States, AAMR was normally distributed with a mean value of 769.7 annual deaths per 100,000 county residents (Figure 3a). Spatial clustering of AAMR was evident, with highest rates concentrated in the Southeast and parts of the Ohio River Basin, as noted by Yang et al. (2011). Lowest rates were generally distributed throughout the Upper Midwest and Interior West.

Model prediction residuals (predicted – observed AAMR) for the validation data, when using all 76 covariates (environmental + socioeconomic) were normally distributed and centered near zero (mean residual = 3.7; see Figure 3b). Spread of the prediction residuals (1 SD = 62.3) was also modest in comparison to the spread of observed AAMR values for the US (1 SD = 121.9), indicating that model accuracy and precision were relatively high. Spatial clustering was also mitigated by the RF modelling process. A map of model validation residuals revealed no conspicuous spatial clustering (Figure 3b) and residual scatterplots did not reveal strong associations with latitude or longitude (Figure 3c). Moreover, coefficients of determination (observed vs. predicted AAMR) for the out-of-bag results (mean $r^2 = 0.748$ among 1000 RF trees) were nearly identical to the r^2 value (0.755) for independent validation data, suggesting the validation results were robust to the random 75:25 data split. A scatterplot of observed versus predicted AAMR values did, however, reveal bias in the model predictions. The RF model tended to underestimate AAMR in counties with the highest observed AAMR values and to overestimate AAMR in counties with the lowest values (Figure 3c).

Ranked %IncMSE values showed that smoking, food insecurity and lack of physical activity were the three strongest predictors of AAMR (Figure 4). Air temperature, annual precipitation and the percent of individuals with a college degree were also important. Effects of the remaining covariates decreased rapidly; for most covariates, the %IncMSE value was at least one order of magnitude

FIGURE 3 Summary plots of observed and model-predicted age-adjusted mortality rate (AAMR). Panel (a) demonstrates spatial (map) and statistical (histogram) distributions of AAMR values (county-level averages between 1999 and 2017). Panel (b) illustrates the spatial and statistical distribution of model residuals (predicted AAMR – observed AAMR) for the random forest model. Counties highlighted in panel (b) ($n = 762$) were randomly selected for model validation (75:25 model building vs. validation data split) and are colour coded to reflect model prediction residuals; counties that were not selected for model validation are shown in yellow. Panel (c) includes scatterplots of the model residuals versus latitude and longitude (geographic center of each of the randomly selected validation counties), and of the observed versus predicted AAMR values. The 1:1 line illustrates perfect agreement between observed and predicted values, while the dashed line illustrates the actual relationship (least squares regression) between observed and predicted AAMR.



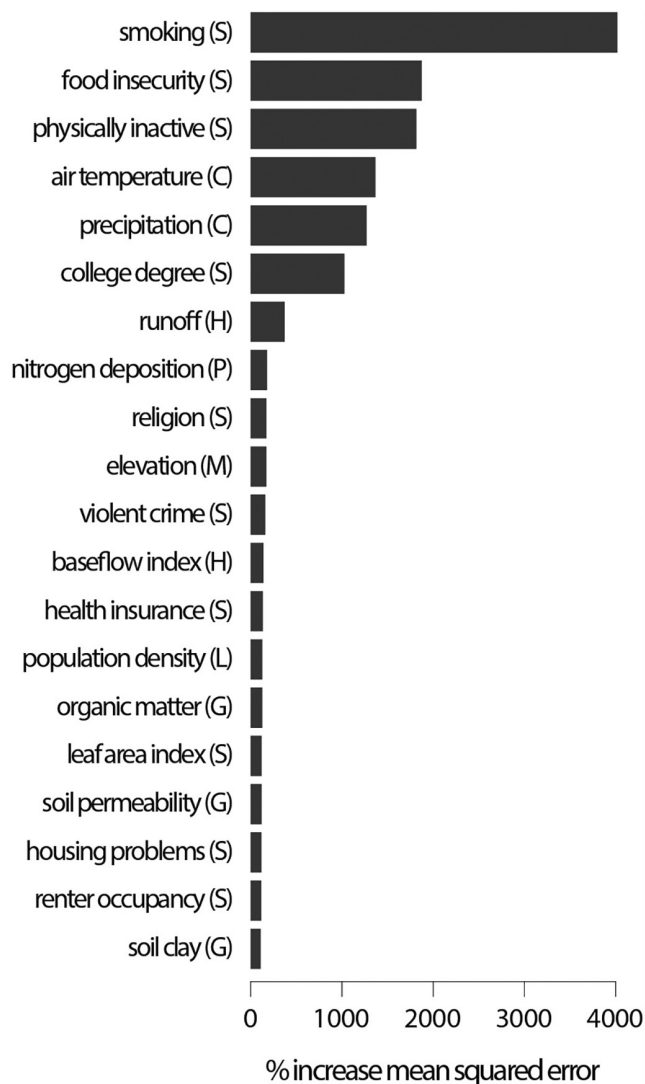


FIGURE 4 Barplot of variable importance values for the top 20 predictors of age-adjusted mortality rate (AAMR). Variable importance is measured as the percent increase in mean squared error when a given variable is removed from the random forest model. Variable themes are shown in parentheses: C, climate; G, geology; H, hydrology; L, land use; M, morphology; P, pollution; S, socioeconomic. Complete variable importance results for all 76 covariates are provided in Table S1.

smaller than the top six %IncMSE values (Table S1). Of the seven predictor themes, socioeconomic variables were clearly the best predictors of AAMR, followed by climate variables. Hydrology, geology, river morphology, land use and pollution variables were less important, although each variable theme was represented by at least one covariate in the top 20 ranked predictors (Figure 4).

Notably, model performance remained high when the 12 socioeconomic covariates were excluded and the RF model was re-run with only the 64 environmental covariates. Residuals for the independent validation data (predicted–observed AAMR) remained normally distributed and centered near zero (mean = -1.871 ; Figure 5a), although a moderate increase in the spread of the residuals (1 SD = 78.541) was observed and r^2 decreased to 0.585. Thus,

removing the socioeconomic variables lowered model precision, but did not have a strong effect on model accuracy. Ranked %IncMSE values also revealed consistency in the relative importance of specific environmental variables when comparing models with and without socioeconomic variables. For example, each of the 10 environmental variables (air temperature, precipitation, runoff, nitrogen deposition, elevation, baseflow index, population density, organic matter, soil permeability and soil clay content) that was ranked among the top 20 predictors when using the complete list of 76 covariates (environmental + socioeconomic covariates; see Figure 4) was also ranked in the top 20 when using the reduced covariate list (Figure 5b).

4 | DISCUSSION

Our study was motivated by the question: can a suite of environmental covariates that represent the physicochemical properties of streams and rivers throughout the United States (Hill et al., 2016), in combination with a shorter list of socioeconomic variables, predict human mortality with a meaningful level of accuracy and precision? In the following discussion, we first seek to answer this question by comparing our RF results with modelling results from other national-scale studies of human health. Next, we reflect on the roles individual model covariates may play in regulating human health. Finally, we conclude by calling for additional efforts to integrate environmental information in human health research.

4.1 | A model performance baseline

Direct comparisons with other model-based studies of human mortality are challenging, due to differences in methods and data, spatial scale and reported summary statistics. However, we believe three previous studies can provide a useful baseline to gauge the effectiveness of our RF model. In a national study of social capital and its influence on human mortality in the United States, Yang et al. (2011) used a combination of ordinary least squares and spatial autoregressive methods to model county-level AAMR as a function of socioeconomic status, local population size and proximity to a major metropolitan area. Their models accounted for 49%–65% of the county-level variance in AAMR and showed that social cohesion (i.e. interpersonal connections maintained through regular social activity) is essential for understanding mortality differences among urban and rural counties. Eichstaedt et al. (2015) also used ordinary least squares regression to model AAMR resulting directly from atherosclerotic heart disease. Predictors included a suite of traditional socioeconomic variables (e.g. income, education, smoking) as well as novel indicators of social media activity. Their county-level models accounted for 13%–18% of the variance in AAMR and showed that regular exposure to negative language on social media may significantly increase the risk of fatal heart disease. Finally, Dwyer-Lindgren et al. (2017) used a large sample of county-level data on socioeconomic and demographic status (e.g. median household income

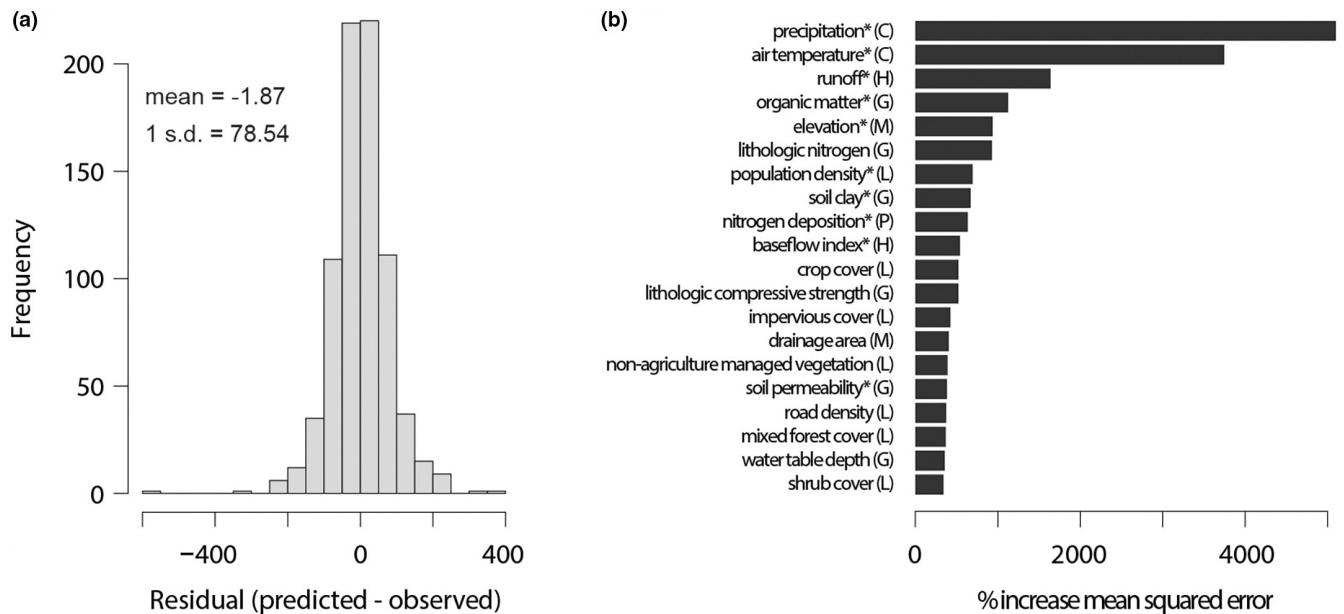


FIGURE 5 Summary of random forest modelling results when the 12 socioeconomic covariates were excluded, leaving the 64 environmental covariates. Panel (a) shows the distribution of model prediction residuals (predicted – observed AAMR) for independent validation data (75:25 model building vs. validation data split). Panel (b) shows variable importance values for the top 20 predictors of AAMR, measured as the percent increase in mean-squared error. Superscript stars (*) indicate variables that were also ranked among the top 20 covariates in the complete (with socioeconomic covariates) model. Variable themes are shown in parentheses: C, climate; G, geology; H, hydrology; L, land use; M, morphology; P, pollution.

and proportions of black and Hispanic residents), behavioural and metabolic risk factors, (e.g. prevalence of smoking and diabetes), and health care access (e.g. percent of residents insured and the number of physicians per capita) to predict life expectancy at birth. Their ordinary least squares regression models accounted for 27%–74% of the variance in life expectancy.

Our model performed as well or better than models from the three baseline studies. When all 76 covariates were included, the RF model accounted for 76% of the variance in observed AAMR, with prediction residuals that were normally distributed and centered near zero (Figure 3b). When the 12 socioeconomic variables were removed, the explained variance in AAMR decreased to 59% and model precision was reduced, but accuracy remained high (prediction residuals centered near zero; Figure 5a). We, therefore, conclude the RF model, which combines socioeconomic information with a suite of environmental variables, is an important advance in national-scale, human health modelling; each of the three baseline studies was limited to socioeconomic predictors of human health.

Further improvement in the accuracy and precision of our model may also be possible with additional tuning of the RF algorithm. We observed a consistent bias in the model predictions, where AAMR in counties with the highest observed values were underestimated (brown points in Figure 3c) and AAMR in counties with the lowest observed values were overestimated (blue-green points in Figure 3c). This particular bias is often observed in RF models (Breiman, 1996; Zhang & Lu, 2012) and can potentially be corrected with iterative bagging (Breiman, 2001b) or residual rotation (Song, 2015). We did not use these tuning methods because our goal was to demonstrate

a macrosystems ecology proof-of-concept, linking the environmental characteristics of freshwater systems with human health. In this initial study, we were not concerned with maximizing model prediction accuracy per se. Nevertheless, these tuning methods could be used in future applications to increase r^2 among observed and predicted AAMR, further underscoring the predictive potential of our macrosystems framework.

4.2 | Individual influences of socioeconomic and environmental variables on human health

Using %IncMSE values to rank model covariates in order of influence, we found that many of the best individual predictors of AAMR were socioeconomic variables (Figure 4), an outcome broadly consistent with traditional epidemiological research (Dwyer-Lindgren et al., 2017). For instance, direct effects on human health are well-documented for the three strongest predictors of AAMR in our model: smoking (Burns, 2003; Doll et al., 2004; Lakier, 1992), food insecurity (Gundersen et al., 2018; Sun et al., 2020), and physical activity (Paffenbarger et al., 1993; Pate et al., 1995). Alternatively, the remaining socioeconomic variables are likely to reflect indirect or mixed effects on human health, ranging from simple, intuitive connections to complex, multi-step pathways (see Evans & Kantrowitz, 2002).

One example of a simple, indirect effect is the connection between college education and AAMR (Figure 4). We hypothesize this is a two-step link that reflects the positive relationship between college

education and income (Pencavel, 1991; Perna, 2003), followed by the association that income shares with direct health influences, such as smoking (Casetta et al., 2017; Evans & Kantrowitz, 2002). Similarly, the prevalence of renter occupied housing is a strong predictor of local income and several forms of social capital that may influence human health (e.g. educational quality and access to greenspace; Rohe et al., 2002).

A more complex example is the association between LAI, our surrogate measure of greenspace, and AAMR (Figure 4). Regular exposure to greenspace may confer numerous health benefits, including enhanced eyesight, improved immune function and lower blood pressure (Frumkin et al., 2017). Mechanisms that underlie these benefits are an active area of research and are likely to include a mix of ecological and psychological factors. For instance, human immune function may benefit from exposure to novel microbes and antigens that are encountered in greenspaces (the 'hygiene' hypothesis; Rook, 2013), while natural killer cell activity may be enhanced by high concentrations of airborne phytoncides in forested habitats (Li et al., 2008). Alternatively, if greenspace is generally an indicator of intact, healthy landscapes, it may promote mental health by mitigating solastalgia – the distress caused by degradation of culturally important environments and natural resources (Galway et al., 2019).

But regardless of the mechanisms, these health benefits will only be realized when residents regularly visit greenspaces to exercise and build social capital (Hunter et al., 2015).

LAI is particularly interesting because it also demonstrates the fundamental role the environment can play in regulating socioeconomic conditions and, by extension, human health. We classified LAI as a socioeconomic variable because it was used as a direct measure of greenspace (Russette et al., 2021) and greenspace is an inherently socioeconomic concept. The principal determinant of the LAI is, however, the environment. By definition, LAI values will be highest in contiguous patches of broadleaf forest (Fang et al., 2019). Human activities such as agriculture and urban development may alter the LAI at a given location, but their influence is secondary to the primary effect of natural vegetation and landcover type.

Another strong environmental influence on human health is runoff. After air temperature and precipitation (discussed separately below), runoff was the most important environmental predictor in the RF model (Figure 4). This is intuitive because runoff is widely recognized as a key source of pollution. Runoff from roads and parking lots, agricultural fields, construction sites, and industrial sources can magnify instream concentrations of organic chemicals, inorganic metals, nutrients, and bacteria (Cole et al., 1984; Müller et al., 2020; Willis & McDowell, 1982). Exposure to these pollutants may then result in a variety of ailments, ranging from skin rashes to sinus infections and gastrointestinal illness (Gaffield et al., 2003; Prüss, 1998; Turbow et al., 2008). Gastrointestinal morbidity is a key health concern, as it results in substantial medical expenses and elevated mortality rates (Chan et al., 2019; Peery et al., 2012).

Finally, we hypothesize that several of the remaining environmental variables from the list of ranked covariates (Figure 4) are linked to groundwater. The baseflow index is a direct measure of

groundwater contribution to local streamflow (Hall, 1968), while soil organic matter, soil permeability and soil clay content are key determinants of groundwater infiltration, movement, and accessibility (O'Geen, 2013). Groundwater may, in turn, be contaminated by excess nutrients (Abdelwaheb et al., 2019; Buss et al., 2004; Schroth & Sinclair, 2003), pesticides (Arias-Estévez et al., 2008), and bacteria (Crane & Moore, 1984) resulting in significant illness when used as a drinking water source (Craun, 1985; Macler & Merkle, 2000). Groundwater irrigation is also a critically limiting resource in many agricultural settings (Siebert et al., 2010). Thus, environmental factors that regulate groundwater availability can have strong influences on the locations of agricultural versus urban communities, which themselves exhibit differing mortality rates (James, 2014).

Additional work will be needed to complete and confirm our interpretations of the individual environmental covariates. For instance, nitrogen deposition, which was the eighth most influential covariate in the RF model (Figure 4) and has known effects on human health (de Vries, 2021; Wolfe & Patz, 2002), may warrant closer inspection. For the moment we submit that the top-ranked environmental covariates reflect logical, material effects on AAMR and these influences are distinct from the 12 socioeconomic variables. Our collinearity screening process ensured the environmental variables were not spuriously correlated with each other or the socioeconomic variables. Furthermore, each of the top-ranked environmental variables from the complete model remained in the list of 20 top-ranked covariates when socioeconomic variables were removed (Figure 5b). This continuity suggests the top-ranked environmental variables were independent of the socioeconomic variables.

4.3 | Integrating environmental information in human health research

Although the individual-level effects of socioeconomic covariates on AAMR were generally strongest, it would be imprudent to discount the importance of environmental influences. Environmental covariates were effective in predicting AAMR when socioeconomic information was excluded (Figure 5) and continued to provide useful, independent information when socioeconomic variables were included (Figures 3 and 4). This reinforces the growing recognition that human health cannot be fully understood and maintained in the absence of environmental information (MEA, 2005).

The foundational importance of the environment is exemplified by air temperature and precipitation. These two covariates were ranked fourth and fifth when all variables were included (Figure 4), then became the first and second most influential covariates when socioeconomic variables were removed (Figure 5b). This was not surprising as both variables can, in extreme cases, be the proximal cause of death. For example, heat wave events may cause heat stroke and death (Hoshiko et al., 2010) while flooding from extreme precipitation events may lead to drowning (Sindall et al., 2022). These types of acute threats to human health are recognized as 'primary' effects of climate change (McMichael, 2013) and are now significant

public health concerns, at national and global scales (Ebi et al., 2018; Romanello et al., 2021). However, the greater significance of temperature and precipitation may stem from their indirect, systems-level effects on human health.

Air temperature and precipitation are fundamental determinants of climate and climate change (Werndl, 2016). As such, they are germane to many of the environmental and socioeconomic factors that affect human health (Portier et al., 2013). Food insecurity will increase in some regions if aridity constrains agricultural production (Dasgupta & Robinson, 2022; Hertel & Rosch, 2010), with predictable health consequences (Gundersen et al., 2018; Sun et al., 2020). Risk of gastrointestinal illness will increase if extreme precipitation events overwhelm water treatment facilities, causing emergency discharge of untreated waste (Jagai et al., 2015; Wade et al., 2014). Endemic diseases, such as influenza, may become more prevalent as temperature and humidity increase (Barreca & Shimshack, 2012) while novel arboviruses, such as dengue and zika, are predicted to expand their ranges (Guernier et al., 2004; Iwamura et al., 2020). Suicide rates may increase as exposure to chronic heat degrades mental health (Burke et al., 2018). And the diverse physical and mental health benefits of greenspace (Coutts & Hahn, 2015; Tzoulas et al., 2007) will be diminished if climate change degrades greenspace resources or limits access to them.

Recognizing the central importance of air temperature and precipitation is also critical from an environmental justice perspective. Wealthy individuals may enhance their immediate socioeconomic reality through small-scale decisions, but climate change is a cosmopolitan problem that can only be mitigated through large-scale, collective action (Adger, 2003). Human populations facing the greatest climate change risks are too often the ones that emit the fewest greenhouse gases and have the least agency to protect themselves (Romanello et al., 2021). It is, therefore, impossible to achieve broad health equity in the absence of an environmental context.

Moving forward, we suggest a productive strategy to advance the science of linked human and environmental health would be to focus on complete networks of connections between environmental variables, socioeconomic variables and human health. In this systems-level approach, we predict that air temperature and precipitation will interact with geologic influences to create baseline conditions for most environmental variables, such as land-cover type and hydrology. The environment is, in turn, predicted to regulate many of the socioeconomic factors that influence human health. By continuing to develop a macrosystems understanding of these interrelationships, scientists will position themselves to play a key role in climate change planning. They may also discover opportunities to create more equitable health outcomes for populations facing different environmental and socioeconomic challenges (Fleishman & Brown, 2019; Heffernan et al., 2014).

AUTHOR CONTRIBUTIONS

Felisha N. Walls conceived the study, led the investigation and writing. Daniel J. McGarvey funded the study, supervised all data analyses and edited the writing. Felisha N. Walls and Daniel J. McGarvey

located the data, developed the analytical methods and created data visualizations.

ACKNOWLEDGEMENTS

We thank Ryan Hill and David Wheeler for providing critical feedback on the data preparation and modelling procedures used in this study. F.N.W. and D.J.M. were funded through a US National Science Foundation grant (DEB 1553111).

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

DATA AVAILABILITY STATEMENT

All data and code used in this paper are freely available in Supporting Information, posted on FigShare <https://doi.org/10.6084/m9.figshare.17430227.v1>.

ORCID

Felisha N. Walls  <https://orcid.org/0000-0001-6000-4853>

REFERENCES

- Abdelwaheb, M., Jebali, K., Dhaouadi, H., & Dridi-Dhaouadi, S. (2019). Adsorption of nitrate, phosphate, nickel and lead on soils: Risk of groundwater contamination. *Ecotoxicology and Environmental Safety*, 179, 182–187. <https://doi.org/10.1016/j.ecoenv.2019.04.040>
- Adger, W. N. (2003). Social capital, collective action, and adaptation to climate change. *Economic Geography*, 79(4), 387–404.
- Allan, J. (2004). Landscapes and riverscapes: The influence of land use on stream ecosystems. *Annual Review of Ecology, Evolution, and Systematics*, 35(2002), 257–284. <https://doi.org/10.1146/annurev.ecolsys.35.120202.110122>
- Arias-Estévez, M., López-Periágo, E., Martínez-Carballo, E., Simal-Gándara, J., Mejuto, J. C., & García-Río, L. (2008). The mobility and degradation of pesticides in soils and the pollution of groundwater resources. *Agriculture, Ecosystems and Environment*, 123(4), 247–260. <https://doi.org/10.1016/j.agee.2007.07.011>
- Bale, A., Greco, S., Pitton, B., Haver, D., & Oki, L. (2017). Pollutant loading from low-density residential neighborhoods in California. *Environmental Monitoring and Assessment*, 189, 386. <https://doi.org/10.1007/s10661-017-6104-2>
- Barreca, A., & Shimshack, J. (2012). Absolute humidity, temperature, and influenza mortality: 30 years of county-level evidence from the United States. *American Journal of Epidemiology*, 176(7), 114–122. <https://doi.org/10.1093/aje/kws259>
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24, 123–140.
- Breiman, L. (2001a). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Breiman, L. (2001b). Using iterated bagging to debias regressions. *Machine Learning*, 45, 261–277.
- Burke, M., González, F., Baylis, P., Heft-Neal, S., Baysan, C., Basu, S., & Hsiang, S. (2018). Higher temperatures increase suicide rates in the United States and Mexico. *Nature Climate Change*, 8(8), 723–729. <https://doi.org/10.1038/s41558-018-0222-x>
- Burns, D. M. (2003). Epidemiology of smoking-induced cardiovascular disease. *Progress in Cardiovascular Diseases*, 46(1), 11–29. [https://doi.org/10.1016/S0033-0620\(03\)00079-3](https://doi.org/10.1016/S0033-0620(03)00079-3)
- Burnside, W., Brown, J., Burger, O., Hamilton, M., Moses, M., & Bettencourt, L. (2012). Human macroecology: Linking pattern and process in big-picture human ecology. *Biological Reviews*, 87(2012), 194–208. <https://doi.org/10.1111/j.1469-185X.2011.00192.x>

- Buss, S. R., Herbert, A. W., Morgan, P., Thornton, S. F., & Smith, J. W. N. (2004). A review of ammonium attenuation in soil and groundwater. *Quarterly Journal of Engineering Geology and Hydrogeology*, 37, 347–359. <https://doi.org/10.1144/1470-9236/04-005>
- Carlisle, D., Wolock, D., & Meador, M. (2011). Alteration of streamflow magnitudes and potential ecological consequences: A multiregional assessment. *Frontiers in Ecology and the Environment*, 9(5), 264–270. <https://doi.org/10.1890/100053>
- Casetta, B., Videla, A. J., Bardach, A., Morello, P., Soto, N., Lee, K., Camacho, P. A., Hermoza Moquillaza, R. V., & Ciapponi, A. (2017). Association between cigarette smoking prevalence and income level: A systematic review and meta-analysis. *Nicotine & Tobacco Research*, 19(12), 1401–1407. <https://doi.org/10.1093/ntr/ntw266>
- Chan, J. S. H., Chao, A. C. W., Cheung, V. C. H., Wong, S. S. K., Tang, W., Wu, J. C. Y., Chan, H. L. Y., Chan, F. K. L., Sung, J. J. Y., & Ng, S. C. (2019). Gastrointestinal disease burden and mortality: A public hospital-based study from 2005 to 2014. *Journal of Gastroenterology and Hepatology*, 34(1), 124–131. <https://doi.org/10.1111/jgh.14377>
- Cole, R. H., Frederick, R. E., Healy, R. P., & Rolan, R. G. (1984). Preliminary findings of the priority pollutant monitoring project of the Nationwide urban runoff program. *Journal of the Water Pollution Control Federation*, 56(7), 898–908.
- Corley, B., Bartelt-Hunt, S., Rogan, E., Coulter, D., Sparks, J., Baccaglini, L., Howell, M., Liaquat, S., Commack, R., & Kolok, A. S. (2018). Using watershed boundaries to map adverse health outcomes: Examples from Nebraska, USA. *Environmental Health Insights*, 12, 1178630217751906. <https://doi.org/10.1177/1178630217751>
- Coutts, C., & Hahn, M. (2015). Green infrastructure, ecosystem services, and human health. *International Journal of Environmental Research and Public Health*, 12(8), 9768–9798. <https://doi.org/10.3390/ijerph120809768>
- Crane, S. R., & Moore, J. A. (1984). Bacterial pollution of groundwater: A review. *Water, Air, and Soil Pollution*, 22, 67–83.
- Craun, G. F. (1985). A summary of waterborne illness transmitted through contaminated groundwater. *Journal of Environmental Health*, 48(3), 122–127.
- Cutler, D., Edwards, T., Beard, K., Cutler, A., Hess, K., Gibson, J., & Lawler, J. (2007). Random forests for classification in ecology. *Ecology*, 88(11), 2783–2792. <https://doi.org/10.1890/07-0539.1>
- Dasgupta, S., & Robinson, E. J. Z. (2022). Attributing changes in food insecurity to a changing climate. *Scientific Reports*, 12, 4709. <https://doi.org/10.1038/s41598-022-08696-x>
- De Groot, R., Wilson, M., & Boumans, R. (2002). A typology for the classification, description and valuation of ecosystem functions, good and services. *Ecological Economics*, 41, 393–408.
- de Vries, W. (2021). Impacts of nitrogen emissions on ecosystems and human health: A mini review. *Current Opinion in Environmental Science and Health*, 21(100249), 1–8. <https://doi.org/10.1016/j.coesh.2021.100249>
- Doll, R., Peto, R., Boreham, J., & Sutherland, I. (2004). Mortality in relation to smoking: 50 Years' observations on male British doctors. *British Medical Journal*, 328(7455), 1519–1528. <https://doi.org/10.1136/bmj.38142.554479.ae>
- Dunham, J., Angermeier, P., Crausbay, S., Cravens, A., Gosnell, H., Mcevoy, J., Moritz, M. A., Raheem, N., & Sanford, T. (2018). Rivers are social–Ecological systems: Time to integrate human dimensions into riverscape ecology and management. *WIREs Water*, 5, e1291. <https://doi.org/10.1002/wat2.1291>
- Dwyer-Lindgren, L., Bertozzi-Villa, A., Stubbs, R., Morozoff, C., Mackenbach, J., van Lenthe, F., Mokdad, A. H., & Murray, C. (2017). Inequalities in life expectancy among US counties, 1980 to 2014: Temporal trends and key drivers. *Journal of the American Medical Association Internal Medicine*, 177(7), 1003–1011. <https://doi.org/10.1001/jamainternmed.2017.0918>
- Ebi, K., Balbus, J., Luber, G., Bole, A., Crimmins, A., Glass, S., Saha, S., Shimamoto, M. M., Trtanj, J., & White-Newsome, J. (2018). Human health. In D. R. Reidmiller, C. W. Avery, D. R. Easterling, K. E. Kunkel, K. L. M. Lewis, T. K. May-cock, & B. C. Stewart (Eds.), *Impacts, risks, and adaptation in the United States: Fourth national climate assessment* (Vol. II, pp. 572–603). U.S. Global Change Research Program. <https://doi.org/10.7930/NCA4.2018.CH14>
- Eichstaedt, J., Schwartz, H., Kern, M., Park, G., Labarthe, D., Merchant, R., Jha, S., Agrawal, M., Dziurzynski, L. A., Sap, M., Weeg, C., Larson, E. E., Ungar, L. H., & Seligman, M. (2015). Psychological language on twitter predicts county-level heart disease mortality. *Psychological Science*, 26(2), 159–169. <https://doi.org/10.1177/0956797614557867>
- Engemann, K., Pedersen, C. B., Arge, L., Tsirogianis, C., Mortensen, P. B., & Svenning, J. C. (2019). Residential green space in childhood is associated with lower risk of psychiatric disorders from adolescence into adulthood. *Proceedings of the National Academy of Sciences of the United States of America*, 116, 5188–5193. <https://doi.org/10.1073/pnas.1807504116>
- Esselman, P., Infante, D., Wang, L., Wu, D., Cooper, A., & Taylor, W. (2011). An index of cumulative disturbance to river fish habitats of the conterminous United States from landscape anthropogenic activities. *Ecological Restoration*, 29(1–2), 133–151.
- Evans, G. W., & Kantrowitz, E. (2002). Socioeconomic status and health: The potential role of environmental risk exposure. *Annual Review of Public Health*, 23(1), 303–331. <https://doi.org/10.1146/annurev.publhealth.23.112001.112349>
- Fang, H., Baret, F., Plummer, S., & Schaepman-Strub, G. (2019). An overview of global leaf area index (LAI): Methods, products, validation, and applications. *Reviews of Geophysics*, 57(3), 739–799. <https://doi.org/10.1029/2018RG000608>
- Fleishman, E., & Brown, H. (2019). Use of macroecology to integrate social justice and conservation. *Global Ecology and Biogeography*, 00, 1–7. <https://doi.org/10.1111/geb.12965>
- Flotemersch, J., Leibowitz, S., Hill, R., Stoddard, J., Thoms, M., & Tharme, R. (2016). A watershed integrity definition and assessment approach to support strategic management of watersheds. *River Research and Applications*, 32, 1654–1671. <https://doi.org/10.1002/rra.2978>
- Fox, E., Hill, R., Leibowitz, S., Olsen, A., Thornbrugh, D., & Weber, M. (2017). Assessing the accuracy and stability of variable selection methods for random forest modeling in ecology. *Environmental Monitoring and Assessment*, 189, 316. <https://doi.org/10.1007/s10661-017-6025-0>
- Frumkin, H., Bratman, G. N., Breslow, S. J., Cochran, B., Kahn, P. H., Jr., Lawler, J. J., Levin, P. S., Tandon, P. S., Varanasi, U., Wolf, K. L., & Wood, S. A. (2017). Nature contact and human health: A research agenda. *Environmental Health Perspectives*, 125, 075001. <https://doi.org/10.1289/EHP1663>
- Gaffield, S. J., Goo, R. L., Richards, L. A., & Jackson, R. J. (2003). Public health effects of inadequately managed stormwater runoff. *American Journal of Public Health*, 93(9), 1527–1533. <https://doi.org/10.2105/AJPH.93.9.1527>
- Galway, L. P., Beery, T., Jones-Casey, K., & Tasala, K. (2019). Mapping the solastalgia literature: A scoping review study. *International Journal of Environmental Research and Public Health*, 16(15), 2662.
- Grammich, C., Hadaway, K., Houseal, R., Jones, D., Krindatch, A., Stanley, R., & Taylor, R. (2012). 2010 U.S. religion census: *Religious congregations and membership study*. Association of Statisticians of American Religious Bodies.
- Guernier, V., Hochberg, M. E., & Guegan, J.-F. (2004). Ecology drives the worldwide distribution of human diseases. *PLoS Biology*, 2(6), 740–746. <https://doi.org/10.1371/journal.pbio.0020141>
- Guirguis, K., Gershunov, A., Tardy, A., & Basu, R. (2014). The impact of recent heat waves on human health in California. *Journal of Applied Meteorology and Climatology*, 53(1), 3–19. <https://doi.org/10.1175/JAMC-D-13-0130.1>
- Gundersen, C., Tarasuk, V., Cheng, J., De Oliveira, C., & Kurdyak, P. (2018). Food insecurity status and mortality among adults in

- Ontario, Canada. *PLoS ONE*, 13(8), 1–10. <https://doi.org/10.1371/journal.pone.0202642>
- Hall, F. R. (1968). Base-flow recessions—A review. *Water Resources Research*, 4, 973–983.
- Heffernan, J. B., Soranno, P. A., Angilletta, M. J., Buckley, L. B., Gruner, D. S., Keitt, T. H., Kellner, J. R., Kominoski, J. S., Rocha, A. V., Xiao, J., Harms, T. K., Goring, S. J., Koenig, L. E., McDowell, W. H., Powell, H., Richardson, A. D., Stow, C. A., Vargas, R., & Weathers, K. C. (2014). Macrosystems ecology: Understanding ecological patterns and processes at continental scales. *Frontiers in Ecology and the Environment*, 12(1), 5–14. <https://doi.org/10.1890/130017>
- Hendryx, M. (2009). Mortality from heart, respiratory, and kidney disease in coal mining areas of Appalachia. *International Archives of Occupational and Environmental Health*, 82, 243–249. <https://doi.org/10.1007/s00420-008-0328-y>
- Hendryx, M., & Ahern, M. (2009). Mortality in appalachian coal mining regions: The value of statistical life lost. *Public Health Reports*, 124(4), 541–550. <https://doi.org/10.1177/003335490912400411>
- Hendryx, M., Conley, J., Fedorko, E., Luo, J., & Armistead, M. (2012). Permitted water pollution discharges and population cancer and non-cancer mortality: Toxicity weights and upstream discharge effects in US rural-urban areas. *International Journal of Health Geographics*, 11(9), 1–15.
- Hendryx, M., & Innes-Wimsatt, K. A. (2013). Increased risk of depression for people living in coal mining areas of central Appalachia. *Ecopsychology*, 5(3), 179–187.
- Hertel, T. W., & Rosch, S. D. (2010). Climate change, agriculture, and poverty. *Applied Economic Perspectives and Policy*, 32(3), 355–385. <https://doi.org/10.1093/aep/pqp016>
- Hill, R., Fox, E., Leibowitz, S., Olsen, A., Thornbrugh, D., & Weber, M. (2017). Predictive mapping of the biotic condition of conterminous U.S. rivers and streams. *Ecological Applications*, 27(8), 2397–2415. <https://doi.org/10.1002/eap.1617>
- Hill, R., Weber, M., Leibowitz, S., Olsen, A., & Thornbrugh, D. (2016). The stream-catchment (StreamCat) dataset: A database of watershed metrics for the conterminous United States. *Journal of the American Water Resources Association*, 52(1), 120–128. <https://doi.org/10.1111/1752-1688.12372>
- Horrigan, L., Lawrence, R., & Walker, P. (2002). How sustainable agriculture can address the environmental and human health harms of industrial agriculture. *Environmental Health Perspectives*, 110(5), 445–456.
- Hoshiko, S., English, P., Smith, D., & Trent, R. (2010). A simple method for estimating excess mortality due to heat waves, as applied to the 2006 California heat wave. *International Journal of Public Health*, 55(2), 133–137. <https://doi.org/10.1007/s00038-009-0060-8>
- Hunter, R. F., Christian, H., Veitch, J., Astell-Burt, T., Hipp, J. A., & Schipperijn, J. (2015). The impact of interventions to promote physical activity in urban green space: A systematic review and recommendations for future research. *Social Science and Medicine*, 124, 246–256. <https://doi.org/10.1016/j.socscimed.2014.11.051>
- Iwamura, T., Guzman-Holst, A., & Murray, K. A. (2020). Accelerating invasion potential of disease vector *Aedes aegypti* under climate change. *Nature Communications*, 11, 2130. <https://doi.org/10.1038/s41467-020-16010-4>
- Jagai, J., Quanlin, L., Wang, S., Messier, K., Wade, T., & Hilborn, E. (2015). Extreme precipitation and emergency room visits for gastrointestinal illness in areas with and without combined sewer systems: An analysis of Massachusetts data, 2003–2007. *Environmental Health Perspectives*, 123(9), 873–879.
- James, W. (2014). All rural places are not created equal: Revisiting the rural mortality penalty in the United States. *American Journal of Public Health*, 104(11), 2122–2129. <https://doi.org/10.2105/AJPH.2014.301989>
- Jian, Y., Messer, L., Jagai, J., Rappazzo, K., Gray, C., Grabich, S., & Lobdell, D. (2017). Associations between environmental quality and mortality in the contiguous United States, 2000–2005. *Environmental Health*, 125(3), 2000–2005.
- Knowlton, K., Rotkin-Ellman, M., King, G., Margolis, H. G., Smith, D., Solomon, G., Trent, R., & English, P. (2009). The 2006 California heat wave: Impacts on hospitalizations and emergency department visits. *Environmental Health Perspectives*, 117(1), 61–67. <https://doi.org/10.1289/ehp.11594>
- Krall, J., Anderson, B., Dominici, F., Bell, M., & Peng, R. (2013). Short-term exposure to particulate matter constituents and mortality in a national study of U.S. urban communities. *Environmental Health Perspectives*, 121(10), 1148–1153. <https://doi.org/10.1289/ehp.1206185>
- Lakier, J. B. (1992). Smoking and cardiovascular disease. *American Journal of Medicine*, 93, S8–S12.
- Leitão, P., Andrew, C., Engelhardt, E., Graham, C., Martinez-Almoyna, C., Mimet, A., Pinkert, S., Schröder, B., Voskamp, A., Hof, C., & Fritz, S. A. (2019). Macroecology as a hub between research disciplines: Opportunities, challenges and possible ways forward. *Journal of Biogeography*, 47, 13–15. <https://doi.org/10.1111/jbi.13751>
- Lewis, D., Voeller, D., Saitone, T., & Tate, K. (2019). Management scale assessment of practices to mitigate cattle microbial water quality impairments of coastal waters. *Sustainability*, 11(5516), 1–10.
- Li, Q., Morimoto, K. I., Kobayashi, M., Inagaki, H., Katsumata, M., Hirata, Y., Hirata, K., Suzuki, H., Li, Y. J., Wakayama, Y., Kawada, T., Park, B. J., Ohira, T., Matsui, N., Kagawa, T., Miyazaki, Y., & Krensky, A. M. (2008). Visiting a forest, but not a city, increases human natural killer activity and expression of anti-cancer proteins. *International Journal of Immunopathology and Pharmacology*, 21(1), 117–127.
- Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News*, 2(3), 18–22.
- Little, J. (1999). Otherness, representation and the cultural construction of rurality. *Progress in Human Geography*, 23(3), 437–442.
- Luken, S. (2022). Policy recommendations to address housing shortages for people with severe mental illness. *Psychiatric Services*, 73(3), 329–334. <https://doi.org/10.1176/appi.ps.202100158>
- Macler, B. A., & Merkle, J. C. (2000). Current knowledge on groundwater microbial pathogens and their control. *Hydrogeology Journal*, 8, 29–40. <https://doi.org/10.1007/PL00010972>
- Mckay, L., Bondelid, T., Dewald, T., Johnston, C., Moore, R., & Rea, A. (2012). *NHDPlus Version 2: User guide* (data model version 2.1). Horizon Systems.
- McMichael, A. J. (2013). Globalization, climate change, and human health. *New England Journal of Medicine*, 368, 1335–1343. <https://doi.org/10.1056/nejmra1109341>
- MEA (Millennium Ecosystem Assessment). (2005). *Ecosystems and human well-being: Health synthesis* (Vol. 18). World Health Organization. <https://doi.org/10.1177/0964663908100335>
- Meehl, G., & Tebaldi, C. (2004). More intense, more frequent, and longer lasting heat waves in the 21st century. *Science*, 305(5686), 994–997. <https://doi.org/10.1126/science.1098704>
- Meyer, J., Strayer, D., Wallace, J., Eggert, S., Helfman, G., & Leonard, N. (2007). The contribution of headwater streams to biodiversity in river networks. *Journal of the American Water Resources Association*, 43(1), 86–103. <https://doi.org/10.1111/j.1752-1688.2007.00008.x>
- Muennig, P., & Glied, S. (2010). What changes in survival rates tell US about US health care. *Health Affairs*, 29(11), 2105–2113. <https://doi.org/10.1377/hlthaff.2010.0073>
- Müller, A., Österlund, H., Marsalek, J., & Viklander, M. (2020). The pollution conveyed by urban runoff: A review of sources. *Science of the Total Environment*, 709, 136125. <https://doi.org/10.1016/j.scitotenv.2019.136125>
- O'Geen, A. (2013). *Soil water dynamics*. <https://www.nature.com/scitable/knowledge/library/soil-water-dynamics-103089121/>
- Olden, J., & Poff, N. (2003). Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. *River Research and Applications*, 19(2), 101–121. <https://doi.org/10.1002/rra.700>

- Oshiro, T., Perez, P., & Baranauskas, J. (2012). How many trees in a random forest? In P. Perner (Ed.), *Machine Learning and Data Mining in Pattern Recognition. MLDM 2012* (pp. 154–168). Springer. https://doi.org/10.1007/978-3-642-31537-4_13
- Ostro, B., Feng, W., Broadwin, R., Green, S., & Lipsett, M. (2007). The effects of components of fine particulate air pollution on mortality in California: Results from CALFINE. *Environmental Health Perspectives*, 115(1), 13–19. <https://doi.org/10.1289/ehp.9281>
- Paffenbarger, R., Hyde, R., Alvin, W., Lee, I.-M., Jung, D., & Kampert, J. (1993). The association of changes in physical-activity level and other lifestyle characteristics with mortality among men. *New England Journal of Medicine*, 328(8), 538–545. <https://doi.org/10.1056/NEJM199309303291401>
- Palmer, M., Bernhardt, E., Schlesinger, W., Eshleman, K., Fofoula-Georgiou, E., Hendryx, M., Lemly, A. D., Likens, G. E., Loucks, O. L., Power, M. E., White, P. S., & Wilcock, P. (2010). Mountaintop mining consequences. *Science*, 327(5962), 148–149. <https://doi.org/10.1021/es1036018>
- Pate, R. R., Pratt, M., Blair, S. N., Haskell, W. L., Macera, C. A., Bouchard, C., Buchner, D., Ettinger, W., Heath, G. W., & King, A. C. (1995). Public health and prevention and the American College of Sports Medicine. *Journal of the American Medical Association*, 273(5), 402–407.
- Peery, A. F., Dellon, E. S., Lund, J., Crockett, S. D., McGowan, C. E., Bulsiewicz, W. J., Gangarosa, L. M., Thiny, M. T., Stizenberg, K., Morgan, D. R., Ringel, Y., Kim, H. P., DiBonaventura, M., Carroll, C. F., Allen, J. K., Cook, S. F., Sandler, R. S., Kappelman, M. D., & Shaheen, N. J. (2012). Burden of gastrointestinal disease in the United States: 2012 update. *Gastroenterology*, 143(5), 1179–1187. <https://doi.org/10.1053/j.gastro.2012.08.002>
- Pencavel, J. (1991). Higher education, productivity, and earnings: A review. *Journal of Economic Education*, 22(4), 331–359. <https://doi.org/10.1080/00220485.1991.10844727>
- Perna, L. W. (2003). The private benefits of higher education: An examination of the earnings premium. *Research in Higher Education*, 44(4), 451–472. <https://doi.org/10.1023/A:1024237016779>
- Pimentel, D., Houser, J., Preiss, E., White, O., Fang, H., Mesnick, L., Barsky, T., Tariche, S., Schreck, J., & Alpert, S. (1997). Water resources: Agriculture, the environment, and society. An assessment of the status of water resources. *BioScience*, 47(2), 97–106.
- Poff, N., & Zimmerman, J. (2010). Ecological responses to altered flow regimes: A literature review to inform the science and management of environmental flows. *Freshwater Biology*, 55, 194–205. <https://doi.org/10.1111/j.1365-2427.2009.02272.x>
- Portier, C. J., Thigpen, K., Carter, S. R., Dilworth, C. H., Grambsch, A. E., Gohlke, J., Hess, J., Howard, S. N., Luber, G., Lutz, J. T., Maslak, T., Prudent, N., Radtke, M., Rosenthal, J. P., Rowles, T., Sandifer, P. A., Scheraga, J., Schramm, P. J., Strickman, D., ... Whung, P.-Y. (2013). A human health perspective on climate change: A report outlining the research needs on the human health effects of climate change. *Journal of Current Issues in Globalization*, 6(4), 621–710.
- Prüss, A. (1998). Review of epidemiological studies on health effects from exposure to recreational water. *International Journal of Epidemiology*, 27(1), 1–9. <https://doi.org/10.1093/ije/27.1.1>
- R Development Core Team 4.3.0. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.r-project.org>
- Ren, H., Wan, X., Yang, F., Shi, X., Xu, J., Zhuang, D., & Yang, G. (2014). Association between changing mortality of digestive tract cancers and water pollution: A case study in the Huai River basin, China. *International Journal of Environmental Research and Public Health*, 12(1), 214–226. <https://doi.org/10.3390/ijerph120100214>
- Resh, V., Brown, A., Covich, A., Gurtz, M., Li, H., Minshall, G., Reice, S. R., Sheldon, A. L., Wallace, J. B., & Wissmar, R. (1988). The role of disturbance in stream ecology. *Journal of the North American Benthological Society*, 7(4), 433–455.
- Rohe, W. M., Van Zandt, S., & McCarthy, G. (2002). Home ownership and access to opportunity. *Housing Studies*, 17(1), 51–61. <https://doi.org/10.1080/02673030120105884>
- Romanello, M., McGushin, A., Di Napoli, C., Drummond, P., Hughes, N., Jamart, L., Kennard, H., Lampard, P., Rodriguez, B. S., Arnell, N., Ayeb-Karlsson, S., Belesova, K., Cai, W., Campbell-Lendrum, D., Capstick, S., Chambers, J., Chu, L., Ciampi, L., Dalin, C., ... Hamilton, I. (2021). The 2021 report of the lancet countdown on health and climate change: Code red for a healthy future. *The Lancet*, 398(10311), 1619–1662.
- Rook, G. A. (2013). Regulation of the immune system by biodiversity from the natural environment: An ecosystem service essential to health. *Proceedings of the National Academy of Sciences of the United States of America*, 110(46), 18360–18367. <https://doi.org/10.1073/pnas.1313731110>
- Roth, N., Allan, J., & Erickson, D. (1996). Landscape influences on stream biotic integrity assessed at multiple spatial scales. *Landscape Ecology*, 11(3), 141–156.
- Russette, H., Graham, J., Holden, Z., Semmens, E., Williams, E., & Landguth, E. (2021). Greenspace exposure and COVID-19 mortality in the United States: January–July 2020. *Environmental Research*, 198, 111195.
- Sampson, R. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, 277(5328), 918–924.
- Sanchez, G. M., Nejadhashemi, A. P., Zhang, Z., Woznicki, S. A., Habron, G., Marquart-Pyatt, S., & Shortridge, A. (2014). Development of a socio-ecological environmental justice model for watershed-based management. *Journal of Hydrology*, 518, 162–177. <https://doi.org/10.1016/j.jhydrol.2013.08.014>
- Schroth, G., & Sinclair, F. L. (Eds.). (2003). *Trees, crops and soil fertility: Concepts and research methods*. Cabi.
- Siebert, S., Burke, J., Faures, J. M., Frenken, K., Hoogeveen, J., Döll, P., & Portmann, F. T. (2010). Groundwater use for irrigation—A global inventory. *Hydrology and Earth System Sciences*, 14, 1863–1880. <https://doi.org/10.5194/hess-14-1863-2010>
- Sindall, R., Mecrow, T., Queiroga, A. C., Boyer, C., Koon, W., & Peden, A. E. (2022). Drowning risk and climate change: A state-of-the-art review. *Injury Prevention*, 28, 185–191. <https://doi.org/10.1136/injuryprev-2021-044486>
- Sissamis, F., Villalba, K., Garcia, J., Melus, V., Markentell, E. J., Perez, L. D., & Ramirez, G. (2022). Religion and death in the United States: A meta-regression comparative assessment of between-county mortality heterogeneity in the United States. *International Journal of Environmental Research and Public Health*, 19(2), 1–16. <https://doi.org/10.3390/ijerph19020757>
- Song, J. (2015). Bias corrections for random Forest in regression using residual rotation. *Journal of the Korean Statistical Society*, 44, 321–326.
- Stekhoven, D., & Bühlmann, P. (2012). Missforest-non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1), 112–118. <https://doi.org/10.1093/bioinformatics/btr597>
- Sun, Y., Liu, B., Rong, S., du, Y., Xu, G., Snetselaar, L. G., Wallace, R. B., & Bao, W. (2020). Food insecurity is associated with cardiovascular and all-cause mortality among adults in the United States. *Journal of the American Heart Association*, 9(19), e014629. <https://doi.org/10.1161/JAHA.119.014629>
- The US Burden of Disease Collaborators. (2018). The state of US health, 1990–2016: Burden of diseases, injuries, and risk factors among US states. *JAMA: The Journal of the American Medical Association*, 319(14), 1444–1472. <https://doi.org/10.1001/jama.2018.0158>
- Thornbrugh, D., Leibowitz, S. G., Hill, R. A., Weber, M. H., Johnson, Z. C., Olsen, A. R., Flotemersch, J. E., Stoddard, J. L., & Peck, D. (2018). Mapping watershed integrity for the conterminous United States. *Ecological Indicators*, 85, 1133–1148. <https://doi.org/10.1016/j.ecolind.2017.10.070>
- Tsang, Y., Wiefelich, D., Fung, K., Infante, D., & Cooper, A. (2014). An approach for aggregating upstream catchment information to support

- research and management of fluvial systems across large landscapes. *Springerplus*, 3, 1–9. <https://doi.org/10.1186/2193-1801-3-589>
- Turbow, D. J., Kent, E. E., & Jiang, S. C. (2008). Web-based investigation of water associated illness in marine bathers. *Environmental Research*, 106(1), 101–109. <https://doi.org/10.1016/j.envres.2007.06.006>
- Tzoulas, K., Korpela, K., Venn, S., Yli-Pelkonen, V., Kaźmierczak, A., Niemela, J., & James, P. (2007). Promoting ecosystem and human health in urban areas using green infrastructure: A literature review. *Landscape and Urban Planning*, 81(3), 167–178. <https://doi.org/10.1016/j.landurbplan.2007.02.001>
- United States Environmental Protection Agency. (2017). *National Water Quality Inventory: Report to Congress, EPA 841-R-16-011*.
- Vitousek, P., Mooney, H., Lubchenco, J., & Melillo, J. (1997). Human domination of Earth's ecosystems. *Science*, 277(5325), 494–499. https://doi.org/10.1007/978-0-387-73412-5_1
- Wade, T., Lin, C., Jagai, J., & Hilborn, E. (2014). Flooding and emergency room visits for gastrointestinal illness in Massachusetts: A case-crossover study. *PLoS ONE*, 9(10), e110474. <https://doi.org/10.1371/journal.pone.0110474>
- Wang, H., & Preston, S. H. (2009). Forecasting United States mortality using cohort smoking histories. *Proceedings of the National Academy of Sciences of the United States of America*, 106(2), 393–398. <https://doi.org/10.1073/pnas.0811809106>
- Werndl, C. (2016). On defining climate and climate change. *British Journal for the Philosophy of Science*, 67, 337–364. <https://doi.org/10.1093/bjps/axu048>
- Wheeler, D., Nolan, B., Flory, A., DellaValle, C., & Ward, M. (2015). Modeling groundwater nitrate concentrations in private wells in Iowa. *Science of the Total Environment*, 536, 481–488.
- Wheeler, D. C., Jones, R. M., Schootman, M., & Nelson, E. J. (2019). Explaining variation in elevated blood lead levels among children in Minnesota using neighborhood socioeconomic variables. *Science of the Total Environment*, 650, 970–977. <https://doi.org/10.1016/j.scitotenv.2018.09.088>
- Wilkinson, R. G., Kawachi, I., & Kennedy, B. P. (1998). Mortality, the social environment, crime and violence. *Sociology of Health & Illness*, 20(5), 578–597. <https://doi.org/10.1111/1467-9566.00120>
- Willis, H., & McDowell, L. L. (1982). Pesticides in agricultural runoff and their effects on downstream water quality. *Environmental Toxicology and Chemistry*, 1(1), 267–279.
- Wolfe, A. H., & Patz, J. A. (2002). Reactive nitrogen and human health: Acute and long-term implications. *Ambio: A Journal of the Human Environment*, 31(2), 120–125.
- Woolf, S., & Aron, L. (2018). Failing health of the United States. *BMJ*, 360, 1–2. <https://doi.org/10.1136/bmj.k496>
- Woolf, S., Chapman, D., Buchanich, J., Bobby, K., Zimmerman, E., & Blackburn, S. (2018). Changes in midlife death rates across racial and ethnic groups in the United States: Systematic analysis of vital statistics. *BMJ*, 362, k3096. <https://doi.org/10.1136/BMJ.K3096>
- Yang, G., Bowling, L. C., Cherkauer, K. A., & Pijanowski, B. C. (2011). The impact of urban development on hydrologic regime from catchment to basin scales. *Landscape and Urban Planning*, 103(2), 237–247. <https://doi.org/10.1016/j.landurbplan.2011.08.003>
- Zanobetti, A., & Schwartz, J. (2009). The effect of fine and coarse particulate air pollution on mortality: A national analysis. *Environmental Health Perspectives*, 117(6), 898–903. <https://doi.org/10.1289/ehp.0800108>
- Zhang, G., & Lu, Y. (2012). Bias-corrected random forests in regression. *Journal of Applied Statistics*, 39(1), 151–160.

How to cite this article: Walls, F. N., & McGarvey, D. J. (2023). Building a macrosystems ecology framework to identify links between environmental and human health: A random forest modelling approach. *People and Nature*, 5, 183–197. <https://doi.org/10.1002/pan3.10427>

Chapter 2 Prelude

Indirect effects of environmental change can pose the greatest threat to human health. However, few studies have examined these diffuse effects on both ecosystem and human health. In Chapter 2, an independent measure of ecosystem health is added to the large database of environmental and socioeconomic variables from Chapter 1. Path analysis is then used to create a systems-level model of direct and indirect effects on human and ecosystem health. The path analysis model identifies geology and climate variables as key drivers of human and ecosystem health outcomes. However, forest cover is found to be a critical, intermediate link between the environmental and the socioeconomic variables that have the strongest, direct influences on human health. This research was published in 2023, in the journal of *Science of the Total Environment* (Walls, F., and D. McGarvey. 2023. A systems-level model of direct and indirect links between environmental health, socioeconomic factors and human mortality. *Science of the Total Environment*, 874: 162486. DOI: <http://dx.doi.org/10.1016/j.scitotenv.2023.162486>).



A systems-level model of direct and indirect links between environmental health, socioeconomic factors, and human mortality

Felisha N. Walls^a, Daniel J. McGarvey^{b,*}

^a Integrative Life Sciences Doctoral Program, Virginia Commonwealth University, 1000 West Cary Street, Richmond, VA 23284, USA

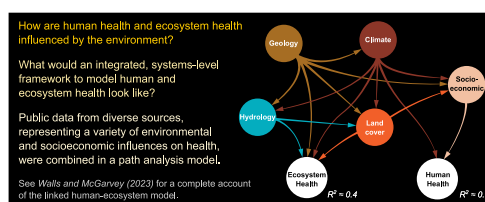
^b Center for Environmental Studies, Virginia Commonwealth University, 1000 West Cary Street, Richmond, VA 23284, USA



HIGHLIGHTS

- Path analysis is used to build a systems-level model of human and ecosystem health.
- Environmental conditions have differing effects on human and ecosystem health.
- Geology, climate, hydrology, and land cover are key drivers of ecosystem health
- Socioeconomic factors and climate explain ~60 % of the variation in human mortality.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Lidia Minguez Alarcon

Keywords:

Path analysis
 Age-adjusted mortality rate
 Climate effects
 Macrosystems ecology

ABSTRACT

Major efforts are being made to better understand how human health and ecosystem health are influenced by climate and other environmental factors. However, studies that simultaneously address human and ecosystem health within a systems-level framework that accounts for both direct and indirect effects are rare. Using path analysis and a large database of environmental and socioeconomic variables, we create a systems-level model of direct and indirect effects on human and ecosystem health in counties throughout the conterminous United States. As indicators of human and ecosystem health, we use age-adjusted mortality rate and an index of biological integrity in streams and rivers, respectively. We show that: (i) geology and climate set boundary conditions for all other variables in the model; (ii) hydrology and land cover have predictable but distinct effects on human and ecosystem health; and (iii) forest cover is a key link between the environment and the socioeconomic variables that directly influence human health.

1. Introduction

Human and environmental health are inextricably linked. This is the core message – and warning – of the Lancet Countdown report (Romanello et al., 2021). As global climate change progresses, natural ecosystems are responding, often with significant consequences for human well-being. In the extreme, climate change may be fatal to humans. Since

the turn of the century, heat waves have directly caused or contributed to thousands of deaths in western Europe (Kovats and Hajat, 2008). Similarly, the frequency and magnitude of flash flood events are predicted to increase in some regions, posing a greater drowning risk to local populations (Haines et al., 2006; Zhang et al., 2021).

A basic typology of climate induced health risks includes three classes (McMichael, 2013). Primary risks are direct, acute health threats, such as heat waves and floods. Secondary risks are indirect, yet readily observable pathways between environmental perturbations and human health. For instance, increasing aridity is expected to diminish agricultural production and exacerbate food insecurity in many regions (Hertel and Rosch, 2010; Mendelsohn, 2008). Alternatively, infectious diseases (e.g., dengue and Zika virus) are predicted to spread as tropical climates expand to temperate latitudes (Guernier et al., 2004; Iwamura et al., 2020). Tertiary risks are the

Abbreviations: AAMR, age-adjusted mortality rate; CONUS, conterminous United States; PA, path analysis; BMMI, benthic multi-metric index; SRMR, Standardized Root Mean Squared Residual; CFI, Comparative Fit Index; TLL, Tucker-Lewis Index.

* Corresponding author.

E-mail addresses: wallsfn@vcu.edu (F.N. Walls), djmcgarvey@vcu.edu (D.J. McGarvey).

most diffuse, as demonstrated by green infrastructure. Green infrastructure (e.g., urban parks) can significantly enhance physical and mental health, but access to green infrastructure is a joint function of ecological and socioeconomic factors that are themselves sensitive to climate change (Coutts and Hahn, 2015; Matthews et al., 2015; Tzoulas et al., 2007). Of the three risk classes, secondary and tertiary risks may pose the greatest cumulative threat because their indirect effects will be the most widespread (Butler and Harley, 2010). This underscores the need for additional research on diffuse connections between climate, the environment and human health (McMichael et al., 2006; Myers and Bernstein, 2011; Portier et al., 2013).

In this study, we demonstrate a macrosystems ecology approach to mapping indirect connections between climate, humans and the environment. Macrosystems ecology is broadly defined as the integrative study of biological, geophysical, and social systems at regional to continental scales (Heffernan et al., 2014). It incorporates many of the tools used in traditional ecology and biogeography, but deliberately seeks to place this information within a social or human context (Leitão et al., 2019; Tromboni et al., 2021). Macrosystems research also capitalizes on new technological developments in large-scale data collection and analysis, such as machine learning and open-source computer code (Dodds et al., 2021).

A recent example of macrosystems research on connections between human and environmental health used random forest modeling to identify key predictors of age-adjusted mortality rate (AAMR) in counties throughout the conterminous United States (CONUS) (Walls and McGarvey, 2023). Potential predictors of human AAMR included a diverse list of socioeconomic, climatic, geologic, and environmental variables. Socioeconomic and behavioral factors, including smoking, physical activity, and food insecurity were found to be the strongest predictors of AAMR. However, climatic and environmental variables, including air temperature, precipitation, and runoff were also strong predictors. Overall, the random forest model was an effective tool for predicting AAMR, with model predictions accounting for 76 % of the variation in observed AAMR values. More importantly, the model demonstrated a workflow to assemble a very large number of potential predictors (total $n > 600$), then rank the predictors by strength of influence on AAMR.

The present study advances the macrosystems science of linked human and environmental health in two ways. First, it uses path analysis (PA) to explore a network of potential cause-and-effect relationships between climate, the physical environment, socioeconomic factors, and human health. Specifically, it orients many of the variables that were previously identified as strong predictors of county-level AAMR (Walls and McGarvey, 2023) within a systems-level path diagram, testing the significance of each path as well as overall model fit. Our PA model was inspired by prior conceptual models of linked human-environment systems (Forget and Lebel, 2001; McMichael, 2013; Myers et al., 2013), but goes further by expressing the direct and indirect links as quantitative, linear equations. The PA model also tracks interrelationships among the predictors of health; this network approach is the key innovation that elevates our model to a systems-level study (Van Dyne, 1966).

Second, our study incorporates independent measures of human health, represented by AAMR, and ecosystem health. By simultaneously modeling human and ecosystem health, we aim to distinguish common environmental influences from effects that are unique to each of the two endpoints. Ecosystem health is represented by a benthic multi-metric index (BMMI) that was originally used to assess biological integrity in streams and rivers throughout the CONUS (Hill et al., 2017; Stoddard et al., 2008). We employ a measure of aquatic ecosystem health because the prevalence and effects of many disturbances, including agriculture (Riseng et al., 2011), forestry (Nislow and Lowe, 2006), mining (Giam et al., 2018), dams (Ligon et al., 1995), fire (Gomez Isaza et al., 2022), urbanization (Violin et al., 2011), and pharmaceutical waste (Rosi-Marshall and Royer, 2012) are thoroughly documented for streams and rivers. Furthermore, rivers and their watersheds provide a uniquely well-integrated context for research on social-ecological systems. Most human populations depend on riverine ecosystem services (Dunham et al., 2018), while the watershed is a natural geospatial unit for modeling environmental exposure and predicting epidemiological outcomes (Corley et al., 2018).

We begin with a metamodel (Grace et al., 2010) of predicted relationships among five broad classes of variables: geologic, climatic, hydrologic, land cover/land use, and socioeconomic. Ecosystem health is predicted to be a function of geology, climate, hydrology, and land use, while human health is predicted to be a function of all five variable classes (Fig. 1). The metamodel, which incorporates only general relationships among variable classes, provides a flexible framework to orient specific variables within the PA model. For instance, at regional to continental scales, elevation (a geologic variable) should have a predictable effect on air temperature and precipitation (climatic variables), while air temperature and precipitation are expected to influence hydrology and land cover.

The generalized nature of the metamodel is also complimentary to the exploratory intent of this study. We are careful not to assert mechanistic understanding because we did not start with explicit, a priori hypotheses, as required in confirmatory PA (Grace and Pugeseck, 1998). Rather, our goal is to build a quantitative, systems-level model of interconnections between the environment, ecosystem health, and human health. Through an iterative model fitting process, we distinguish secondary and tertiary influences on health from more familiar, primary influences. We also distinguish model links that are consistent with general knowledge from links that surprised us and may warrant closer inspection.

2. Material and methods

2.1. Data collection and management

All environmental, socioeconomic, and health data used in this study were previously published and fully detailed (Walls and McGarvey, 2023). Briefly, environmental data were downloaded from the 1:100,000 scale National Hydrography Dataset (Mckay et al., 2012) and the U.S. Environmental Protection Agency's Stream-Catchment database (Hill et al., 2016; downloaded February 2020). These two sources provided >600 environmental variables for the CONUS, representing a range of climatic, geologic, hydrologic, and land cover influences on local stream and river ecosystems. Within the land cover class, a new 'forest cover' variable was created by summing values for deciduous, coniferous, and mixed forest cover. For each of the four classes of environmental variables, principal component analysis was used to select a subset of variables that retained much of the overall information while minimizing collinearity ($|r| < 0.7$)

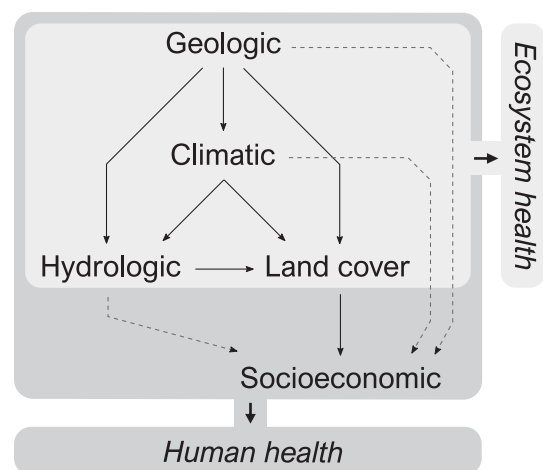


Fig. 1. Metamodel of hypothesized relationships among geologic, climatic, hydrologic, land cover or land use and socioeconomic variables. Strongest predicted effects are indicated by solid arrows. Weaker or secondary effects are indicated by dashed arrows. Ecosystem health is predicted to be a function of geologic, climatic, hydrologic, and land cover variables (all predictors within the inset light gray region). Human health is predicted to be a function of all variable classes (all predictors within the complete darker region).

among variables. This resulted in a final, combined list of 64 environmental variables (see Appendix Table A1).

Socioeconomic data were obtained from four online sources (downloaded December 2021). First, we downloaded nine behavioral and social predictors of human well-being (adult smoking, physical inactivity, food insecurity, violent crime, severe housing shortages, degree of ruralness, and social association rate) from the County Health Rankings and Roadmaps website (www.countyhealthrankings.org). This information was compiled from multiple sources, including the National Center for Health Statistics and the U.S. Bureau of Labor Statistics, by the University of Wisconsin Population Health Institute (Remington et al., 2015). Three additional indicators of education and economic status were then collected from the U.S. Census Bureau, American Community Survey (www.census.gov/programs-surveys/acs/), including county-level prevalence of a bachelor's degree, health insurance, and renter occupied housing. Following Russette et al. (2021), the county-level Leaf Area Index (LAI) was used as a measure of greenspace exposure, which may benefit human health by facilitating physical activity and social connectedness (Frumkin et al., 2017). Finally, the prevalence of religious activity and affiliation (all denominations and groups combined), which is also associated with positive health outcomes (Milstein et al., 2020), was obtained from the 2010 U.S. Religion Census (Grammich et al., 2012). All 12 socioeconomic variables included in this study are listed with units of measure and data sources in Table A1.

We obtained AAMR records (county-level means between 1999 and 2017) from the U.S. Centers for Disease Control and Prevention, Wide-ranging Online Data for Epidemiologic Research website (<https://wonder.cdc.gov/>; downloaded February 2020). Our ecosystem health indicator, BMMI, was distributed with the Stream-Catchment database (Hill et al., 2016; downloaded February 2020). The BMMI was originally built using a large, national-scale survey of benthic invertebrate field samples ($n = 3045$; Stoddard et al., 2008). Following a standardized protocol, separate multi-metric indices were calibrated for each of nine distinct ecoregions (Omernik, 1987), with higher index scores reflecting healthier aquatic communities. Notably, a review of >400 published studies (Ruaro et al., 2020) showed that many of the individual metrics selected by Stoddard et al. (2008) are commonly included in multi-metric indices (e.g., taxa richness, pollution intolerant taxa, mayfly/stonefly/caddisfly abundance), suggesting our results may align well with other aquatic biomonitoring programs. The empirical (site-specific) index scores of Stoddard et al. (2008) were then used to calibrate a model of aquatic community health for every stream and river in the CONUS. Specifically, a separate random forest model was developed for each of the nine ecoregions, predicting the probability that a given stream or river would be in 'good' condition (Hill et al., 2017). This model-predicted probability of being in good condition ranged from 0 to 1 and was used as the BMMI score for each stream and river segment.

County-level environmental, socioeconomic, and health data were cross-referenced with Federal Information Processing Standard codes. Environmental and socioeconomic variables were transformed to improve normality, using the 'Rita' package (Mattei and Ruscio, 2022) with default settings (Shapiro-Wilk normality test, $\alpha = 0.05$) in R (R Core Team, 2022). Resulting transformations are listed in Table A1. All variables were then standardized (z -transformed) to prevent PA model convergence errors that may result from large covariance (>3 orders of magnitude) discrepancies.

2.2. Spatial filtering

Spatial filtering (Boria et al., 2014) was used to mitigate spatial autocorrelation in the county-level BMMI (first-order Moran's $I = 0.20$; $p < 0.01$) and AAMR data (Moran's $I = 0.12$; $p < 0.01$). The filtering process began by locating the centroid, calculated as one-half the distance between the maximum and minimum longitudinal and latitudinal coordinates, of every county within the CONUS. One-hundred iterations of a random walk algorithm were then used to select subsets of counties with centroids that were no closer than a specified filtering distance b . Each iteration

began by randomly selecting a single focal county, then simulating a random walk to a neighboring county with a centroid at least b distance from the centroid of the focal county. The random walk continued, ensuring at each step the centroid of the next county was separated from the centroids of all previously selected counties by a distance $\geq b$, until no selectable counties remained. Selected counties were then queried from the complete list of CONUS counties and the process was repeated until 100 random subsets of filtered counties were obtained. Spatial filtering was conducted with the 'spThin' package (Aiello-lammens et al., 2015) and a custom loop in R.

To identify a suitable b value, we created a series of spatial lags at 50 km intervals (0–49 km, 50–99, 100–149, etc.). Moran's I was then calculated for each lag, using the minimum and maximum extent of the respective lag to define neighboring counties, when measured as centroid-to-centroid distances. For instance, Moran's I for the 50–99 km lag included all county centroids that were within 50–99 km of a focal county centroid. Moran's I values were calculated separately for the two response variables, AAMR and BMMI, then plotted to characterize the relationship between each lag distance and Moran's I (Fig. 2). From these plots, we concluded b values of 100, 150, and 200 km would be suitable for spatial filtering, as Moran's I decreased to a near asymptotic trend between the 100–149 and 150–199 km lags for AAMR. For BMMI, Moran's I decreased rapidly between the 100–149 and 200–249 km lags. Notably, the nearest-neighbor (centroid-to-centroid) distance was <100 km for 98.3 % of all CONUS counties (Fig. 2 inset). Thus, the 100, 150 and 200 km filtering distances substantially reduced autocorrelation in BMMI and AAMR.

2.3. Path analysis

Guided by the metamodel (Fig. 1), we first created a simple PA model of BMMI and AAMR where each of the five variable classes (geologic, climatic, hydrologic, land cover, and socioeconomic) was represented by one or two variables. We then added additional variables in an exploratory, stepwise manner that gradually increased model complexity while preserving the basic metamodel structure. Iterative decisions to add or remove paths and/or variables were informed by path significance levels and PA modification indices. Three fit statistics were used to assess overall model fit (Fan et al., 2016): the Standardized Root Mean Squared Residual (SRMR), Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI).

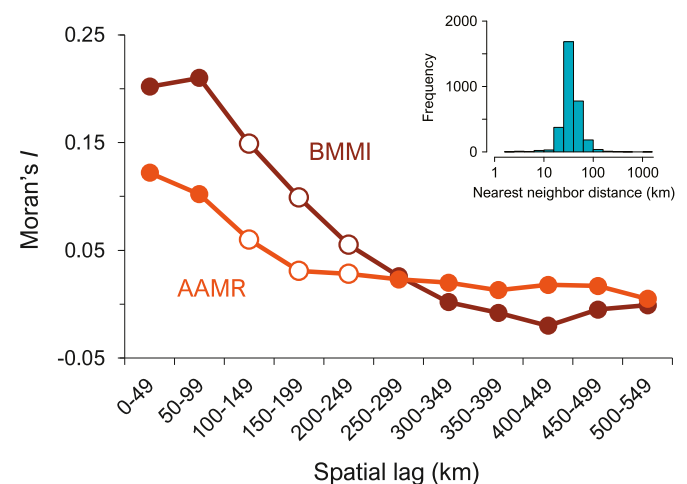


Fig. 2. Plots showing spatial autocorrelation in stream ecosystem health (BMMI) and human health (AAMR) as functions of neighbor distance, or spatial lag, for U.S. counties. For each of the two variables, Moran's I is plotted against a series of 50 km spatial lags. Lags shown as open circles (100–149, 150–199 and 200–249 km) were used as spatial filters ($b = 100, 150$ and 200 km) to mitigate spatial autocorrelation in the modeling data, prior to building path analysis models. Inset histogram shows the distribution of nearest neighbor distances, measured as the centroid-to-centroid distance between adjacent counties (log₁₀ scale), for all counties in the conterminous U.S.

A single, initial subsample of spatially filtered counties was used with the $b = 100$ km filter to build the PA model shown in Fig. 3. Additional subsamples were then drawn at random and used to determine whether the PA modeling results were robust to the spatial filtering process. A custom software loop was written in R to repeat the spatial filtering and PA model fitting process 99 times. At the conclusion of each iteration, we recorded the standardized path coefficient and significance level for every model path, R^2 for AMMR and BMMI, and the three fit indices. Averages and standard deviations were then calculated among the 100 iterations for each of the stored modeling results.

Next, the complete PA modeling process was independently repeated for each of the two remaining spatial filters ($b = 150$ and 200 km). For each of the three b values, we recorded means and standard deviations from 100 filtering iterations, for all PA results. PA model summary statistics were then compared among the three spatial filters. In effect, the iterative results were used as a cross-validation tool to assess model robustness.

PA modeling was performed with the ‘lavaan’ package in R (Rosseel, 2012). R code to repeat all data transformation, standardization, spatial filtering, and path analysis procedures is provided with the raw data in Supporting Information.

3. Results and discussion

We compiled 76 predictor variables (32 geologic, 2 climatic, 4 hydrologic, 26 land cover, and 12 socioeconomic; Table A1), representing 3048 of 3142 counties and county equivalents throughout the CONUS. Four exogenous (geologic) and 17 endogenous (climatic, hydrologic, land cover, and socioeconomic) variables were included in the final PA model (Fig. 3). All paths were significant at $p \leq 0.05$ (mean of 100 iterations) when $b = 100$ km (Table 1). When $b = 150$ km, 59 of 64 paths remained significant and when $b = 200$ km, 53 paths were significant. Overall model fit was good, as indicated by standard fit index thresholds (Fan

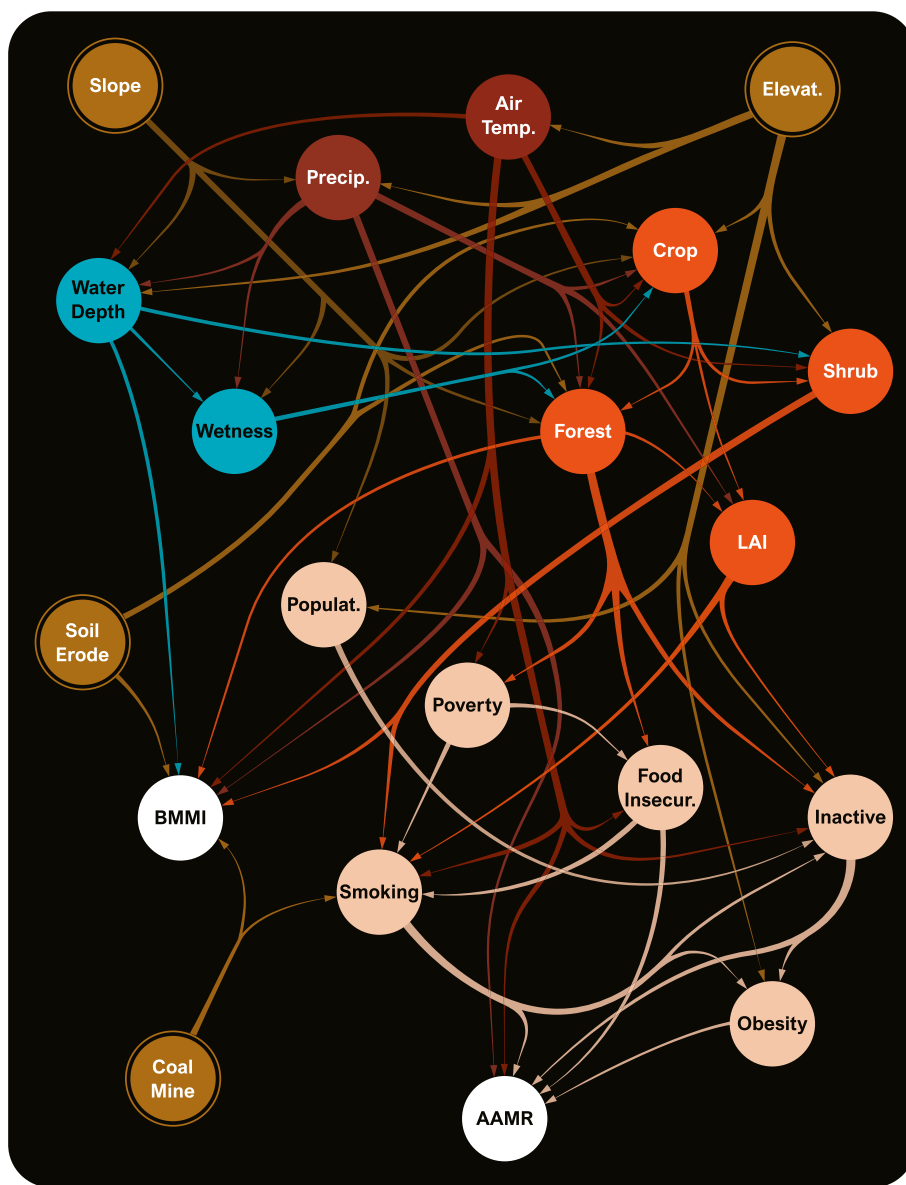


Fig. 3. Path model diagram of age-adjusted mortality rate (AAMR) in U.S. counties. The five classes of predictor variables are indicated by color: brown = geologic; maroon = climatic; blue = hydrologic; orange = land cover; pink = socioeconomic. Variables shown in white are indices of stream ecosystem health (BMMI) and human health (AAMR). Exogenous variables are shown with an outer ring; all remaining variables are endogenous. All paths shown in the diagram are significant at $p \leq 0.05$, when a spatial filtering distance of 100 km is used. Complete path analysis results (standardized path coefficients, model fit statistics, and coefficients of determination for BMMI and AAMR) are listed for all three filtering distances ($b = 100, 150$ and 200 km) in Tables 1 and 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

et al., 2016). Mean SRMR values exceeded the significance threshold (<0.09) for all *b* values (Table 2) and mean TLI values met the threshold (≥ 0.90) for all *b* values except 150 km. CFI values were close to the significance threshold (≥ 0.95); for each *b* value, mean CFI was 0.01–0.02 units below the threshold. Among the three *b* values, mean coefficients of determination (R^2) for BMMI ranged from 0.39 to 0.42 (Table 2). For AAMR, the mean R^2 ranged from 0.60 to 0.63. Small standard deviations for all path coefficients (Table 1), model fit statistics, and R^2 values (Table 2) indicate the modeling results were robust to the random spatial filtering process.

3.1. Socioeconomic drivers of AAMR

The four strongest predictors of AAMR were all socioeconomic variables (Fig. 3). Smoking, obesity, physical inactivity, and food insecurity each had a positive effect on AAMR (Table 1). This was not surprising, as previous epidemiological research has demonstrated the importance of these behavioral and metabolic risk factors. Smoking is a leading cause of cancer, cardiovascular and respiratory disease, and is associated with hypertension, ulcers, and insulin resistance (Burns, 2003; Doll et al., 2004; Skurnik and Shoenfeld, 1998). Obesity is linked to cardiovascular disease, kidney and liver disease, diabetes, hyperlipidemia, hypertension, and cancer (Abdelaal et al., 2017; Flegal, 2005; Solomon and Manson, 1997). Physical inactivity increases the risk of heart disease, hypertension, anxiety, and depression (Cecchini et al., 2010; Paffenbarger et al., 1993; Pate et al., 1995). Similarly, food insecurity is associated with cardiovascular disease, diabetes, hyperlipidemia, and hypertension (Gundersen et al., 2018; Sun et al., 2020). In general, the congruence between epidemiological literature and model-identified drivers of AAMR, combined with the high R^2 values for AAMR (≥ 0.60 for all *b* values; Table 2), suggests our PA results are a good first attempt at a systems-level understanding of environmental effects on human health.

Modeled relationships among the socioeconomic influences on AAMR (see Fig. 3) were also consistent with prior health and behavior research. For example, the positive effect of poverty on food insecurity and smoking (Table 1) is well-documented in the broader literature on income and socioeconomic status (Casetta et al., 2017; Evans and Kantrowitz, 2002; Loopstra and Tarasuk, 2013). Food insecurity had a direct, positive effect on smoking that may reflect psychological distress caused by food shortage, followed by smoking as a coping mechanism (Kim-Mozeleski et al., 2019). The negative effect of population density on physical inactivity (i.e., residents of more populous counties are more likely to be active) has been documented elsewhere and partially attributed to use of walking and cycling as transportation modes in dense urban communities (Forsyth et al., 2007; McCormack and Shiell, 2011). Smoking was also a significant predictor of physical inactivity (Kaczynski et al., 2008) and obesity (Dare et al., 2015). This combined, multi-step effect of smoking on AAMR was particularly notable as it underscored the cumulative, negative influence that smoking has on human health.

3.2. Drivers of BMMI and its broader relevance for environmental health

Stream health was not significantly influenced by socioeconomic variables in the PA model (Fig. 3). Instead, BMMI was linked exclusively to a mix of environmental variables. Forest cover and shrub cover each had a significant, positive effect on BMMI, while soil erodibility and coal mining had negative effects (Table 1). The positive effect of forest cover was expected, as intact forests benefit water quality and aquatic biota in numerous ways (Noteboom et al., 2021; Roth et al., 1996). At the riparian zone scale, allochthonous debris from deciduous and coniferous trees accounts for much of the annual carbon that aquatic consumers depend on, particularly in small to midsize streams (Webster and Meyer, 1997). Riparian forests also protect the stream channel from erosion, filter pollutants from overland runoff, and maintain cooler water temperatures in shaded reaches (Gregory et al., 1991; Triska et al., 1982). Riparian shrubs can provide similar benefits when local soils, climates, or disturbance regimes constrain mature forest stands. For example, riparian shrubs enhance instream

Table 1

Standardized path coefficients for all paths shown in the final path analysis model (Fig. 3). For each path, directionality moves from the ‘Predictor’ variable to the ‘Response’ variable. Coefficients are mean averages from 100 spatial filtering iterations, with 1 s.d. shown in parentheses. Separate results are included for the three spatial filtering distances (100, 150 and 200 km). All paths are significant at $p \leq 0.05$ (mean *p*-value among 100 iterations) unless their respective coefficients are highlighted in italic text; for these paths, the mean $p > 0.05$.

Response	Predictor	Spatial filter (b)		
		100 km	150 km	200 km
Air temp.	Elevat.	-0.51 (0.01)	-0.51 (0.01)	-0.49 (0.02)
	Precip.	-0.80 (0.02)	-0.85 (0.03)	-0.78 (0.05)
Water depth	Slope	0.17 (0.02)	0.21 (0.03)	0.15 (0.06)
	Air temp.	0.32 (0.01)	0.29 (0.02)	0.24 (0.03)
	Elevat.	0.59 (0.03)	0.64 (0.04)	0.56 (0.05)
	Precip.	-0.21 (0.02)	-0.16 (0.04)	-0.23 (0.05)
Wetness	Slope	0.15 (0.02)	0.14 (0.03)	0.14 (0.04)
	Precip.	-0.18 (0.03)	-0.17 (0.04)	-0.15 (0.07)
	Slope	0.34 (0.03)	0.35 (0.04)	0.33 (0.06)
Crop	Water depth	-0.25 (0.03)	-0.26 (0.07)	-0.28 (0.09)
	Air temp.	-0.22 (0.01)	-0.18 (0.02)	-0.18 (0.04)
	Elevat.	-0.23 (0.02)	-0.20 (0.04)	-0.18 (0.06)
	Precip.	-0.15 (0.02)	-0.15 (0.03)	-0.13 (0.04)
Shrub	Slope	-0.26 (0.02)	-0.27 (0.03)	-0.26 (0.05)
	Soil erode	0.36 (0.02)	0.38 (0.03)	0.36 (0.04)
	Wetness	-0.15 (0.02)	-0.12 (0.03)	-0.13 (0.06)
	Air temp.	0.16 (0.02)	0.12 (0.02)	0.12 (0.04)
	Crop	-0.68 (0.02)	-0.69 (0.03)	-0.64 (0.05)
Forest	Elevat.	0.22 (0.02)	0.20 (0.03)	0.17 (0.05)
	Water depth	0.17 (0.02)	0.19 (0.04)	0.26 (0.06)
	Air temp.	-0.34 (0.01)	-0.32 (0.02)	-0.35 (0.03)
	Crop	-0.47 (0.02)	-0.43 (0.03)	-0.44 (0.04)
LAI	Precip.	0.55 (0.01)	0.50 (0.02)	0.46 (0.02)
	Slope	0.15 (0.02)	0.13 (0.03)	0.12 (0.04)
	Soil erode	-0.08 (0.02)	-0.09 (0.03)	-0.05 (0.05)
	Wetness	-0.09 (0.02)	-0.06 (0.03)	-0.04 (0.04)
	Crop	0.75 (0.02)	0.77 (0.03)	0.75 (0.04)
Populat.	Forest	0.83 (0.02)	0.83 (0.04)	0.77 (0.07)
	Precip.	0.15 (0.02)	0.13 (0.03)	0.16 (0.03)
	Elevat.	-0.46 (0.02)	-0.45 (0.04)	-0.51 (0.06)
Poverty	Slope	0.31 (0.03)	0.30 (0.05)	0.31 (0.07)
	Air temp.	0.34 (0.02)	0.36 (0.04)	0.33 (0.05)
Food insecur.	Forest	0.15 (0.03)	0.13 (0.04)	0.13 (0.07)
	Air temp.	0.18 (0.02)	0.12 (0.03)	0.16 (0.04)
	Forest	0.22 (0.02)	0.23 (0.03)	0.27 (0.05)
Inactive	Poverty	0.49 (0.03)	0.48 (0.05)	0.43 (0.05)
	Air temp.	0.22 (0.02)	0.20 (0.03)	0.17 (0.04)
	Elevat.	-0.17 (0.02)	-0.17 (0.03)	-0.19 (0.04)
	Forest	-0.18 (0.03)	-0.14 (0.03)	-0.19 (0.06)
Smoking	LAI	0.20 (0.02)	0.18 (0.04)	0.19 (0.04)
	Populat.	-0.27 (0.02)	-0.26 (0.03)	-0.24 (0.04)
	Smoking	0.48 (0.03)	0.48 (0.04)	0.49 (0.06)
	Air temp.	-0.15 (0.02)	-0.20 (0.03)	-0.17 (0.05)
	Coal mine	0.15 (0.02)	0.15 (0.04)	0.14 (0.05)
	Food insecur.	0.21 (0.03)	0.25 (0.04)	0.30 (0.06)
	LAI	0.16 (0.02)	0.10 (0.03)	0.15 (0.04)
	Poverty	0.54 (0.04)	0.55 (0.06)	0.47 (0.09)
Obesity	Shrub	-0.13 (0.02)	-0.20 (0.03)	-0.19 (0.05)
	Elevat.	-0.12 (0.02)	-0.11 (0.03)	-0.09 (0.03)
	Inactive	0.50 (0.03)	0.47 (0.04)	0.50 (0.06)
	Smoking	0.30 (0.03)	0.33 (0.05)	0.31 (0.07)
BMMI	Air temp.	-0.27 (0.02)	-0.26 (0.03)	-0.24 (0.05)
	Coal mine	-0.08 (0.02)	-0.08 (0.03)	-0.12 (0.06)
	Forest	0.23 (0.03)	0.24 (0.05)	0.21 (0.07)
	Precip.	0.14 (0.03)	0.12 (0.04)	0.12 (0.06)
	Shrub	0.21 (0.03)	0.21 (0.05)	0.21 (0.07)
AAMR	Soil erode	-0.12 (0.02)	-0.11 (0.03)	-0.09 (0.04)
	Water depth	0.28 (0.03)	0.27 (0.04)	0.28 (0.07)
	Air temp.	0.11 (0.02)	0.11 (0.03)	0.09 (0.04)
	Food insecur.	0.19 (0.03)	0.18 (0.04)	0.14 (0.06)
	Inactive	0.14 (0.04)	0.16 (0.06)	0.16 (0.07)
	Obesity	0.17 (0.03)	0.15 (0.05)	0.14 (0.07)
	Precip.	0.09 (0.01)	0.09 (0.02)	0.12 (0.03)
	Smoking	0.32 (0.04)	0.31 (0.06)	0.36 (0.09)

Table 2

Summary statistics for the final path analysis model (Fig. 3). Results include the number of counties (n) selected by the random spatial filtering process, three indices of model fit (SRMR, TLI, and CFI), and coefficients of determination (R^2) for the model response variables BMMI and AAMR. Values shown are mean averages from 100 spatial filtering iterations with 1 s.d. shown in parentheses. Separate results are included for the three spatial filtering distances (100, 150 and 200 km).

Statistic	Spatial filter (b)		
	100 km	150 km	200 km
n	465.87 (4.13)	230.6 (3.24)	133.43 (2.26)
SRMR (<0.09)	0.06 (<0.01)	0.07 (<0.01)	0.07 (0.01)
TLI (≥ 0.90)	0.90 (0.01)	0.89 (0.01)	0.90 (0.02)
CFI (≥ 0.95)	0.94 (<0.01)	0.93 (0.01)	0.94 (0.01)
R^2 BMMI	0.41 (0.02)	0.42 (0.03)	0.39 (0.04)
R^2 AAMR	0.63 (0.02)	0.60 (0.03)	0.60 (0.04)

water quality by removing nitrate from hyporheic and parafluvial water (Hill, 2019; Schade et al., 2001). At larger scales, forest and shrub dominated landscapes are less likely to route pollutants, including nutrients, toxic chemicals, or excess sediments to streams and rivers than urban or agricultural landscapes (Allan, 2004; Sponseller et al., 2001; Strayer et al., 2003). Intact, forested landscapes also sustain natural flow regimes (Poff et al., 1997; Zhang et al., 2017) and facilitate dispersal of flying aquatic insects (maintaining metacommunity structure; (Urban et al., 2006)) more effectively than highly modified or fragmented landscapes.

Negative effects of soil erodibility and coal mining on BMMI were also predictable. Looser, more easily entrained soils accumulate in streams and rivers, where they may have multiple deleterious effects on aquatic biota (Pimentel and Kounang, 1998). These include increased turbidity and diminished light penetration (i.e., decreased primary production), loss of hard substrate suitable for periphyton growth, interference with fish and insect respiratory surfaces, and clogging of interstitial spaces used by benthic invertebrates and gravel-spawning fishes (Henley et al., 2000; Wood and Armitage, 1997). Coal mining effects on aquatic ecosystems include extensive soil erosion and elevated instream turbidity, increased runoff and flashier hydrographs, removal of riparian forest, acute loss of stream channels through surface excavation and valley fills, increased acidity and conductivity, and elevated concentrations of toxic metals (Hopkins et al., 2013; Lindberg et al., 2011; Palmer et al., 2010). Invertebrate and fish assemblage structure are, in turn, sensitive to these water quality and physical habitat disturbances (Giam et al., 2018; Hitt and Chambers, 2014; Pond et al., 2008).

The positive effect of water depth – specifically, mean seasonal water table depth – on BMMI (Table 1) is more difficult to interpret. We anticipated a negative effect because a shallower water table will, on average, provide more stable baseflow to the local stream network (Marani et al., 2001) and enhanced flow stability is generally a benefit to aquatic biota (Hayashi and Rosenberry, 2002). One explanation for the counterintuitive, positive link between water table depth and BMMI is the mitigating effect that deep groundwater can have on instream nitrate. Groundwater is a key nitrate reservoir (Ascott et al., 2017) and can deliver large quantities of nitrate to local streams and rivers when exchange rates with surface water are high (Jordan et al., 1997; Molenat et al., 2008). A deeper water table may therefore improve surface water quality and the health of aquatic biota by sequestering nitrate from the stream network. Clearly, this nitrate-based explanation is incomplete. The dynamics of ground-surface water exchange vary with geology, landscape morphology and land use (Groffman et al., 2002; Jasechko et al., 2016; Sophocleous, 2002), and are influenced by biogeochemical processes in riparian soils (Brunke and Gonser, 1997; Hill, 2019). Nevertheless, we suggest nitrate transfer is a useful starting point to interpret the observed, positive relationship between water table depth and BMMI.

Finally, we recognize that BMMI, as an indicator of aquatic ecosystem health, is an incomplete measure of environmental health and suggest that integrating new information on terrestrial ecosystem health should be a priority for future research. That said, the river does provide a

fundamental link between human activity and the environment. Many anthropogenic activities, such as agriculture, irrigation, mining, and urban development have strong, predictable influences on rivers because pollution moves downgradient to the local stream or river channel. Hence, a river is a direct reflection of the state of its parent watershed. This is a core tenet of riverscape ecology and a compelling reason to prioritize aquatic systems in social-ecological systems research (Corley et al., 2018; Dunham et al., 2018).

The Central Rio Grande Valley is a case in point. Rapid expansion of agriculture, urban development, and invasive species throughout the Valley has dramatically altered the landscape, with significant consequences for the health of the River and surrounding human communities. Instream flows are now diminished or absent through much of the year, water quality is consistently degraded, and zoonotic diseases are more prevalent (Rappport et al., 1998; Esteve-Gassent et al., 2014). These concerns catalyzed the Middle Rio Grande Conservation Action Plan, which seeks to improve ecosystem health throughout the Valley, but aims to do so primarily by protecting and restoring the River (Muldavin et al., 2019).

3.3. History as a bridge between environmental and socioeconomic influences on health

Despite the complexity of the links among environmental and socioeconomic variables, a tractable narrative did emerge from the PA model. As predicted by the metamodel (Fig. 1), geology, climate and hydrology regulated land cover, which in turn accounted for much of the environmental influence on socioeconomic variables (Fig. 3). In this system, the roles of crop cover and forest cover were particularly important. Crop cover was a significant predictor of each of the remaining land cover variables, while forest cover was linked to three of the socioeconomic variables (Table 1). We hypothesize that these land cover to socioeconomic links are, in part, reflections of the history of human settlement and land conversion across the CONUS.

Westward expansion in the 19th century prompted largescale timber harvest and conversion of forest land to agricultural fields (Dakubo, 2010; Li et al., 2022). These changes accelerated through the Industrial Revolution with increased access to steam power, rapid transport by railway, enhanced crop yields made possible by chemical fertilizers, and the rise of major urban centers. Benefits of modern agriculture and urbanization were not, however, uniformly distributed. Smaller populations that remained in forested areas began to experience socioeconomic disparities, as they were isolated from the technological, social, and economic gains experienced in cities (James, 2014). This may explain the model-implied positive effect of forest cover on poverty and food insecurity.

Previous research has demonstrated a strong, positive correlation between poverty and proximity to forested areas (Sunderlin et al., 2007). This correlation is exemplified by economic status in rural and urban parts of Appalachia. A historical legacy of resource-limited industries, particularly logging and coal mining, has left much of Appalachia vulnerable to high rates of unemployment and poverty. This extractive history has also constrained urban development throughout much of the region, placing many residents in small, rural communities where economic, educational, and medical resources are underdeveloped, relative to larger urban centers. Predictably, human health in these impoverished, rural communities has suffered (Borak et al., 2012; Halverson et al., 2004).

Similarly, food insecurity is high in many of the remote, nonmetropolitan communities that occur in or adjacent to forested areas. Food insecurity rates are particularly high in the southern U.S., where a frequent lack of infrastructure, including low densities of grocery stores and sparse public transportation, can create food deserts (Piontak and Schulman, 2014). Food insecurity generally decreases with economic development (Tweeten, 1999), but economic development has historically been constrained in many forested regions of the CONUS (Drummond and Loveland, 2010). Thus, the modeled link between forest cover and food insecurity is both logical and consistent with the broader literature on rurality (Nelson et al., 2021). Indeed, a strategic effort to integrate new theory and

data from rurality research, including urban-to-rural gradients in education (Koricich et al., 2018), healthcare (Caldwell et al., 2016), and environmental justice (Pellow, 2016) may enhance our modeling results and provide additional insight to the links between human health and the environment.

One notable exception to the negative effect of forest cover on human health was the inverse relationship between forest cover and physical inactivity. At the CONUS scale, inactivity tended to decrease (i.e., residents became more active) with increasing forest cover (Table 1). We believe this reflects a link between greenspace access and physical activity. Forested land that remains accessible to the public is frequently used as greenspace (Zhu and Zhang, 2008), facilitating physical recreation and active lifestyles (Huynh et al., 2022; Kim et al., 2020). If our greenspace hypothesis is correct, it highlights the value of a systems-level approach. Rather than summarizing the net effect of forest cover with a single coefficient, as a traditional regression model would, our PA model identified negative as well as positive effects of forest cover on human health.

The PA model also highlights the specific value of forested greenspace, relative to a more general definition of greenspace. While forest cover was negatively associated with physical inactivity, the LAI was positively associated with inactivity (i.e., residents became less active; Table 1). We were surprised by this result because we assumed LAI, as a measure of greenspace, would be inversely correlated with physical inactivity. But on closer inspection, we learned the particular LAI used here (Russette et al., 2021) was tuned to maximize the spectral signatures of agricultural crops (Gitelson, 2004). This accounts for the strong positive effect of crop cover on LAI (Fig. 3, Table 1) and may help explain the counterintuitive, positive link between LAI and physical inactivity. If the LAI is equally representative of agricultural and forested land use, then a direct correlation between outdoor recreation and high LAI values should no longer be assumed. Indeed, this context-dependent interpretation is consistent with the growing recognition that greenspace is not a homogenous concept (Wheeler et al., 2015). Rather, the human health benefits of greenspace are often contingent on the location and specific type of greenspace in question (Akpınar et al., 2016; Jarvis et al., 2020; Nguyen et al., 2021).

3.4. The pervasive influence of climate

The foundational role that climate plays in regulating human and ecosystem health was particularly evident in the PA results. As predicted by the metamodel (Fig. 1), the climate variables air temperature and precipitation had significant influences on each of the hydrology and land cover variables, and on many of the socioeconomic variables (Fig. 3). In this way, climate was confirmed to be an indirect yet fundamental driver of AAMR and BMMI: climate sets the stage for many of the dynamic pathways that regulate human and environmental health. However, the PA model also provided evidence of direct air temperature and precipitation effects on AAMR and BMMI. Specifically, precipitation had a positive effect on AAMR and BMMI, while air temperature had a positive effect on AAMR, but a negative effect on BMMI (Table 1).

Though we did not anticipate direct climate effects on AAMR or BMMI (Fig. 1), they are entirely plausible. For instance, the positive effect of air temperature on AAMR may reflect deaths from heat wave events (Hoshiko et al., 2010). This may be particularly dangerous in humid urban settings, where exposure to high heat index values can be exacerbated by the urban heat island effect and diminished overnight cooling (Basu and Samet, 2002). The positive effect of precipitation may reflect drowning in catastrophic floods (Sindall et al., 2022) or increased exposure to waterborne disease as municipal treatment facilities are overwhelmed (Curriero et al., 2001). The direct, positive link between precipitation and BMMI may reflect functional or life history adaptations of aquatic biota to instream flows (Lytle, 2008; Poff and Allan, 1995), which are tightly associated with precipitation. Alternatively, the negative effect of air temperature on BMMI may reflect warming-induced respiratory stress (Verberk and Bilton, 2013) or the loss of stenothermic, keystone functional groups, such as leaf-shredding insects (Boyero et al., 2012).

The central role of climate in the PA model is also consistent with the growing literature on climate change and its effects. Considerable efforts have been made to identify human populations that are most vulnerable to climate change (Ebi et al., 2018; Romanello et al., 2021) and to predict new threats like the spread of novel infectious diseases (Guernier et al., 2004; Iwamura et al., 2020). In freshwater ecosystems, temperature and precipitation are known to have key influences on the structure of fish and invertebrate communities (Pyne and Poff, 2017), as well as secondary production (Patrick et al., 2019). Our modeling results are well-aligned with this larger body of climate effects research, but they are also an important innovation. By adopting a macrosystems ecology approach, combining biological and socioeconomic data within a novel and integrative context, we have created a fully quantitative, systems-level framework to identify pathways linking humans and wildlife to the geophysical world.

3.5. Next steps

By adding new variables to the dataset, we predict it will be possible to expand the PA model, test additional pathways, and move closer to truly mechanistic understanding of linked human-environmental health systems. For the moment, however, we suggest two priorities for future research. First, sub-models should be developed to test for differences in model structure among regions and demographic sub-groups. Spatial clustering in human life expectancy and mortality rate has previously been documented for U.S. counties (Dwyer-Lindgren et al., 2017) and strong differences in AAMR have been reported among racial groups (Cunningham et al., 2017). Similarly, the environmental influences on BMMI were shown to differ among CONUS sub-regions (Hill et al., 2017). In both cases, regional variation is likely to reflect co-variation in the environmental factors that regulate AAMR and BMMI.

For instance, forest cover had a significant influence on BMMI and three of the socioeconomic variables (Fig. 3). But forested land is most prevalent in the east/southeast and northwest; forest cover is much lower in the interior CONUS (Zhu and Evans, 1994). This regional, coarse-scale variation is amenable to our historical interpretation of forest cover and socioeconomic influences on human health (see above). However, our CONUS-level analysis is unlikely to detect fine-scale patterns in forest cover that may also have important ramifications for human health. The ‘dilution effect’ that links forest fragmentation to the spread of Lyme disease is a telling example. Degradation of continuous forest habitat diminishes native biodiversity while increasing the relative abundance of Lyme’s primary mammalian host (white-footed mouse, *Peromyscus leucopus*), thereby magnifying the probability a tick will consume infected blood (Allan et al., 2003). Human exposure to infected ticks will simultaneously increase as forest fragmentation creates more easily accessed edge habitat, often in close proximity to urban settings. This kind of local-scale dynamic, which would not be detected in our CONUS-scale analysis, demonstrates the potential value of regional sub-models.

Second, path models that explicitly account for spatial autocorrelation in AAMR and BMMI should be built and compared with the current results. Our spatial filtering algorithm was a simple tool to mitigate autocorrelation in the county-level AAMR and BMMI data. However, the iterative thinning process sacrificed information and truncated our sample sizes, particularly at larger b values (Table 2). This resulted in diminished statistical power and failure to detect statistical significance for some model paths (Table 1). An alternative would be to model the spatial autocorrelation itself, then include it as a random variable in a confirmatory or ‘piecewise’ PA model that retains all available data (Lefcheck, 2016). We did not build a piecewise model because the local estimation process requires prior information on the expected relationship (e.g., linear, exponential, or logistic) between each pair of linked variables; with 16 endogenous variables and 64 paths in our model, we were not prepared to specify all components of a piecewise model. Instead, we used a traditional linear model structure in combination with transformed (to increase normality) variables. This simpler approach was prudent, as our initial focus was on global estimation and overall model fit (Grace et al., 2010). But future work would

benefit from a confirmatory PA test of the systems-level model shown in Fig. 3.

We foresee many opportunities to build upon and improve the PA model, particularly if others approach this work with the creative and expansive philosophy that typifies macrosystems ecology research. Climate change and population growth pose significant threats to human and environmental health. But by integrating diverse data resources, it may also be possible to anticipate and mitigate those threats.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2023.162486>.

CRedit authorship contribution statement

Felisha Walls: Conceptualization, methodology, investigation, visualization, writing.

Daniel McGarvey: Project lead and funding, methodology, visualization, supervision, writing, review and editing.

Data availability

All data and code used in this paper are publicly available in the Supporting Information, posted on Figshare (10.6084/m9.figshare.21606411).

Declaration of competing interest

The authors confirm that they have no known competing financial or personal interests that may have influenced the data used or results reported in this paper.

Acknowledgements

We thank Ryan Hill and David Wheeler for providing critical feedback on the data preparation and modeling procedures used in this study. Funding was provided by the U.S. National Science Foundation [grant DEB-1553111].

References

- Abdelaal, M., le Roux, C.W., Docherty, N.G., 2017. Morbidity and mortality associated with obesity. *Ann. Transl. Med.* 5 (7), 1–12. <https://doi.org/10.21037/atm.2017.03.107>.
- Aiello-Iammens, M.E., Boria, R.A., Radosavljevic, A., Vilela, B., Anderson, R.P., 2015. spThin : an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography* 38, 541–545. <https://doi.org/10.1111/ecog.01132>.
- Akpinar, A., Barbosa-Leiker, C., Brooks, K.R., 2016. Does green space matter? Exploring relationships between green space type and health indicators. *Urban For. Urban Green.* 20, 407–418. <https://doi.org/10.1016/j.ufug.2016.10.013>.
- Allan, J., 2004. Landscapes and riverscapes: the influence of land use on stream ecosystems. *Annu. Rev. Ecol. Syst.* 35 (2002), 257–284. <https://doi.org/10.1146/annurev.ecolsys.35.120202.110122>.
- Allan, B.F., Keesing, F., Ostfeld, R.S., 2003. Effect of forest fragmentation on Lyme disease risk. *Conserv. Biol.* 17 (1), 267–272. <https://doi.org/10.1046/j.1523-1739.2003.01260.x>.
- Ascott, M.J., Gooddy, D.C., Wang, L., Stuart, M.E., Lewis, M.A., Ward, R.S., Binley, A.M., 2017. Global patterns of nitrate storage in the vadose zone. *Nat. Commun.* 8 (1), 1416. <https://doi.org/10.1038/s41467-017-01321-w>.
- Basu, R., Samet, J.M., 2002. Relation between elevated ambient temperature and mortality: a review of the epidemiologic evidence. *Epidemiol. Rev.* 24 (2), 190–202. <https://doi.org/10.1093/epirev/mxf007>.
- Borak, J., Salipante-Zaidel, C., Slade, M.D., Fields, C.A., 2012. Mortality disparities in Appalachia: reassessment of major risk factors. *J. Occup. Environ. Med.* 54 (2), 146–156. <https://doi.org/10.1097/JOM.0b013e318246f395>.
- Boria, R.A., Olson, L.E., Goodman, S.M., Anderson, R.P., 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecol. Model.* 275, 73–77. <https://doi.org/10.1016/j.ecolmodel.2013.12.012>.
- Boyer, L., Pearson, R.G., Dudgeon, D., Ferreira, V., Manuel, A., Graça, S., Gessner, M.O., Boulton, A.J., Chauvet, E., Yule, C.M., Lamothe, S., 2012. Global patterns of stream detritivore distribution: implications for biodiversity loss in changing climates. *Global Ecology and Biogeography* 21, 134–141.
- Brunke, M., Gonser, T., 1997. The ecological significance of exchange processes between rivers and groundwater. *Freshw. Biol.* 37, 1–33.
- Burns, D.M., 2003. Epidemiology of smoking-induced cardiovascular disease. *Prog. Cardiovasc. Dis.* 46 (1), 11–29. [https://doi.org/10.1016/S0033-0620\(03\)00079-3](https://doi.org/10.1016/S0033-0620(03)00079-3).

- Butler, C.D., Harley, D., 2010. Primary, secondary and tertiary effects of eco-climatic change: the medical response. *Postgrad. Med. J.* 86, 230–234. <https://doi.org/10.1136/pgmj.2009.082727>.
- Caldwell, J.T., Ford, C.L., Wallace, S.P., Wang, M.C., Takahashi, L.M., 2016. Intersection of living in a rural versus urban area and race/ethnicity in explaining access to health care in the United States. *Am. J. Public Health* 106 (8), 1463–1469. <https://doi.org/10.2105/ajph.2016.303212>.
- Casetta, B., Videla, A.J., Bardach, A., Morello, P., Soto, N., Lee, K., Camacho, P.A., Hermoza Moquillaza, R.V., Ciapponi, A., 2017. Association between cigarette smoking prevalence and income level: a systematic review and meta-analysis. *Nicotine Tob. Res.* 19 (12), 1401–1407. <https://doi.org/10.1093/ntr/ntw266>.
- Cecchini, M., Sassi, F., Lauer, J.A., Lee, Y.Y., Guajardo-Barron, V., Chisholm, D., 2010. Tackling of unhealthy diets, physical inactivity, and obesity: health effects and cost-effectiveness. *Lancet* 376 (9754), 1775–1784. [https://doi.org/10.1016/S0140-6736\(10\)61514-0](https://doi.org/10.1016/S0140-6736(10)61514-0).
- Corley, B., Bartelt-Hunt, S., Rogan, E., Coulter, D., Sparks, J., Baccaglini, L., Howell, M., Liaquat, S., Commack, R., Kolok, A., 2018. Using watershed boundaries to map adverse health outcomes: examples from Nebraska, USA. *Environ. Health Insights* 12. <https://doi.org/10.1177/1178630217751906>.
- Coutts, C., Hahn, M., 2015. Green infrastructure, ecosystem services, and human health. *Int. J. Environ. Res. Public Health* 12 (8), 9768–9798. <https://doi.org/10.3390/ijerph120809768>.
- Cunningham, T.J., Croft, J.B., Liu, Y., Lu, H., Eke, P.I., Giles, W.H., 2017. Vital signs : racial disparities in age-specific mortality among Blacks or African Americans — United States, 1999–2015. *Morb. Mortal. Wkly Rep.* 66 (17), 444–456.
- Curriero, F.C., Patz, J.A., Rose, J.B., Lele, S., 2001. The association between extreme precipitation and waterborne disease outbreaks in the United States, 1948–1994. *Am. J. Public Health* 91 (8), 1194–1199.
- Dakubo, C.Y., 2010. Chapter 1: Exploring the linkages between ecosystems and human health. *Ecosystems and Human Health*, pp. 1–19. <https://doi.org/10.1007/978-1-4419-0206-1>.
- Dare, S., Mackay, D.F., Pell, J.P., 2015. Relationship between smoking and obesity: a cross-sectional study of 499,504 middle-aged adults in the UK general population. *PLoS ONE* 10 (4), 1–12. <https://doi.org/10.1371/journal.pone.0123579>.
- Dodds, W.K., Rose, K.C., Fei, S., Chandra, S., 2021. Macrosystems revisited: challenges and successes in a new subdiscipline of ecology. *Front. Ecol. Environ.* 19 (1), 4–10. <https://doi.org/10.1002/fee.2286>.
- Doll, R., Peto, R., Boreham, J., Sutherland, I., 2004. Mortality in relation to smoking: 50 years' observations on male British doctors. *Br. Med. J.* 328 (7455), 1519–1528. <https://doi.org/10.1136/bmj.38142.554479.ae>.
- Drummond, M.A., Loveland, T.R., 2010. Land-use pressure and a transition to forest-cover loss in the eastern United States. *Bioscience* 60 (4), 286–298. <https://doi.org/10.1525/bio.2010.60.4.7>.
- Dunham, J.B., Angermeier, P.L., Crausbay, S.D., Cravens, A.E., Gosnell, H., McEvoy, J., Moritz, M.A., Raheem, N., Sanford, T., 2018. Rivers are social–ecological systems: time to integrate human dimensions into riverscape ecology and management. *WIREs Water* 5, e1291. <https://doi.org/10.1002/wat2.1291>.
- Dwyer-Lindgren, L., Bertozzi-Villa, A., Stubbs, R., Morozoff, C., Mackenbach, J., Van Lenthe, F., Mokdad, A., Murray, C., 2017. Inequalities in life expectancy among U.S. counties, 1980 to 2014: Temporal trends and key drivers. *Journal of the American Medical Association Internal Medicine* 177 (7), 1003–1011. <https://doi.org/10.1001/jamainternmed.2017.0918>.
- Ebi, K., Balbus, J., Luber, G., Bole, A., Crimmins, A., Glass, S., Saha, S., Shimamoto, J., Trtanj, J., White-Newsome, J., 2018. Human health. Impacts, Risks, and Adaptation in the United States: Fourth National Climate Assessment. U.S. Global Change Research Program, II, pp. 572–603. <https://doi.org/10.7930/NCA4.2018.CH14>.
- Esteve-Gasset, M.D., Pérez de León, A.A., Romero-Salas, D., Ferial-Arroyo, T.P., Patino, R., Castro-Arellano, I., Gordillo-Pérez, G., Auclair, A., Goolsby, J., Rodriguez-Vivas, R.I., Estrada-Franco, J.G., 2014. Pathogenic landscape of transboundary zoonotic diseases in the Mexico–US border along the Rio Grande. *Front. Public Health* 2, 177. <https://doi.org/10.3389/fpubh.2014.00177>.
- Evans, G.W., Kantrowitz, E., 2002. Socioeconomic status and health: the potential role of environmental risk exposure. *Annu. Rev. Public Health* 23 (1), 303–331. <https://doi.org/10.1146/annurev.publhealth.23.112001.112349>.
- Fan, Y., Chen, J., Shirkey, G., John, R., Wu, S.R., Park, H., Shao, C., 2016. Applications of structural equation modeling (SEM) in ecological studies: an updated review. *Ecol. Process.* 5 (1). <https://doi.org/10.1186/s13717-016-0063-3>.
- Flegal, K.M., 2005. Epidemiologic aspects of overweight and obesity in the United States. *Physiol. Behav.* 86 (5), 599–602. <https://doi.org/10.1016/j.physbeh.2005.08.050>.
- Forget, G., Lebel, J., 2001. An ecosystem approach to human health. *Int. J. Occup. Environ. Health* 7 (2), S3–S38.
- Forsyth, A., Oakes, J.M., Schmitz, K.H., Hearst, M., 2007. Does residential density increase walking and other physical activity? *Urban Stud.* 44 (4), 679–697. <https://doi.org/10.1080/00420980601184729>.
- Frumkin, H., Bratman, G.N., Breslow, S.J., Cochran, B., Kahn, P.H., Lawler, J.J., Levin, P.S., Tandon, P.S., Varanasi, U., Wolf, K.L., Wood, S.A., 2017. Nature contact and human health: a research agenda. *Environ. Health Perspect.* 125, 075001. <https://doi.org/10.1289/EHP1663>.
- Giam, X., Olden, J.D., Simberloff, D., 2018. Impact of coal mining on stream biodiversity in the US and its regulatory implications. *Nat. Sustain.* 1, 176–183.
- Gitelson, A.A., 2004. Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *J. Plant Physiol.* 161 (2), 165–173. <https://doi.org/10.1078/0176-1617-01176>.
- Gomez Isaza, D.F., Cramp, R.L., Franklin, C.E., 2022. Fire and rain: a systematic review of the impacts of wildfire and associated runoff on aquatic fauna. *Glob. Chang. Biol.* 28 (8), 2578–2595. <https://doi.org/10.1111/gcb.16088>.

- Grace, J.B., Pugeseck, B.H., 1998. On the use of path analysis and related procedures for the investigation of ecological problems. *Am. Nat.* 152 (1), 151–159. <https://doi.org/10.1086/286156>.
- Grace, J.B., Anderson, T.M., Olff, H., Scheiner, S.M., 2010. On the specification of structural equation models for ecological systems. *Ecol. Monogr.* 80 (1), 67–87.
- Grammich, C., Hadaway, K., Housear, R., Jones, D., Krindatch, A., Stanley, R., Taylor, R., 2012. 2010 U.S. Religion Census: Religious Congregations and Membership Study. Association of Statisticians of American Religious Bodies.
- Gregory, S.V., Swanson, F.J., McKee, W.A., Cummins, K.W., 1991. An ecosystem perspective of riparian zones. *Bioscience* 41 (8), 540–551. <https://doi.org/10.2307/1311607>.
- Groffman, P.M., Boulware, N.J., Zipperer, W.C., Pouyat, R.V., Band, L.E., Colosimo, M.F., 2002. Soil nitrogen cycle processes in urban riparian zones. *Environ. Sci. Technol.* 36 (21), 4547–4552. <https://doi.org/10.1021/es020649z>.
- Guernier, V., Hochberg, M.E., Guegan, J.-F., 2004. Ecology drives the worldwide distribution of human diseases. *PLoS Biol.* 2 (6), 740–746. <https://doi.org/10.1371/journal.pbio.0020141>.
- Gundersen, C., Tarasuk, V., Cheng, J., De Oliveira, C., Kurdyak, P., 2018. Food insecurity status and mortality among adults in Ontario/Canada. *PLoS ONE* 13 (8), 1–10. <https://doi.org/10.1371/journal.pone.0202642>.
- Haines, A., Kovats, R.S., Campbell-Lendrum, D., Corvalan, C., 2006. Climate change and human health: impacts, vulnerability and public health. *Public Health* 120, 585–596. <https://doi.org/10.1016/j.puhe.2006.01.002>.
- Halverson, J.A., Ma, L., Harner, E.J., 2004. Analysis of Health Disparities in Health Status and Access to Health Care in the Appalachian Region. Appalachian Regional Commission, September, 1–10.
- Hayashi, M., Rosenberry, D.O., 2002. Effects of ground water exchange on the hydrology and ecology of surface water. *Ground Water* 40 (3), 309–316. <https://doi.org/10.1111/j.1745-6584.2002.tb02659.x>.
- Heffernan, J.B., Soranno, P.A., Angilletta, M.J., Buckley, L.B., Gruner, D.S., Keitt, T.H., Kellner, J.R., Kominski, J.S., Rocha, A.V., Xiao, J., Weathers, K.C., 2014. Macro-systems ecology: Understanding ecological patterns and processes at continental scales. *Frontiers in Ecology and the Environment* 12 (1), 5–14. <https://doi.org/10.1890/1530017>.
- Henley, W.F., Patterson, M.A., Neves, R.J., Dennis Lemly, A., 2000. Effects of sedimentation and turbidity on lotic food webs: a concise review for natural resource managers. *Rev. Fish. Sci.* 8 (2), 125–139. <https://doi.org/10.1080/10641260091129198>.
- Hertel, T.W., Rosch, S.D., 2010. Climate change, agriculture, and poverty. *Appl. Econ. Perspect. Policy* 32 (3), 355–385. <https://doi.org/10.1093/aep/pqp016>.
- Hill, A.R., 2019. Groundwater nitrate removal in riparian buffer zones: a review of research progress in the past 20 years. *Biogeochemistry* 143, 347–369. <https://doi.org/10.1007/s10533-019-00566-5>.
- Hill, R., Weber, M., Leibowitz, S., Olsen, A., Thornbrugh, D., 2016. The stream-catchment (StreamCat) dataset: a database of watershed metrics for the conterminous United States. *J. Am. Water Resour. Assoc.* 52 (1). <https://doi.org/10.1111/1752-1688.12372>.
- Hill, R., Fox, E., Leibowitz, S., Olsen, A., Thornbrugh, D., Weber, M., 2017. Predictive mapping of the biotic condition of conterminous U.S. rivers and streams. *Ecol. Appl.* 27 (8), 2397–2415. <https://doi.org/10.1002/eap.1617>.
- Hitt, N.P., Chambers, D.B., 2014. Temporal changes in taxonomic and functional diversity of fish assemblages downstream from mountaintop mining. *Freshw. Sci.* 33 (3), 915–926. <https://doi.org/10.1086/676997>.
- Hopkins, R.L., Altier, B.M., Haselman, D., Merry, A.D., White, J.J., 2013. Exploring the legacy effects of surface coal mining on stream chemistry. *Hydrobiologia* 713 (1), 87–95. <https://doi.org/10.1007/s10750-013-1494-9>.
- Hoshiko, S., English, P., Smith, D., Trent, R., 2010. A simple method for estimating excess mortality due to heat waves, as applied to the 2006 California heat wave. *Int. J. Public Health* 55 (2), 133–137. <https://doi.org/10.1007/s00038-009-0060-8>.
- Huynh, L.T.M., Gasparatos, S., Lam, R.D., Grant, E., Fukushi, K., 2022. Linking the non-material dimensions of human-nature relations and human well-being through cultural ecosystem services. *Science Advances* 8 (31), 1–22. <https://doi.org/10.1126/sciadv.abn8042>.
- Iwamura, T., Guzman-Holst, A., Murray, K.A., 2020. Accelerating invasion potential of disease vector *Aedes aegypti* under climate change. *Nature Communications* 11 (2130). <https://doi.org/10.1038/s41467-020-16010-4>.
- James, W., 2014. All rural places are not created equal: revisiting the rural mortality penalty in the United States. *Am. J. Public Health* 104 (11), 2122–2129. <https://doi.org/10.2105/AJPH.2014.301989>.
- Jarvis, I., Gergel, S., Koehoorn, M., van den Bosch, M., 2020. Greenspace access does not correspond to nature exposure: measures of urban natural space with implications for health research. *Landsc. Urban Plan.* 194, 103686. <https://doi.org/10.1016/j.landurbplan.2019.103686>.
- Jasechko, S., Kirchner, J.W., Welker, J.M., McDonnell, J., 2016. Substantial proportion of global streamflow less than three months old. *Nat. Geosci.* 9 (2), 126–129. <https://doi.org/10.1038/ngeo2636>.
- Jordan, T.E., Correll, D.L., Weller, D.E., 1997. Relating nutrient discharges from watersheds to land use and stream flow variability. *Water Resour. Res.* 33 (11), 2579–2590.
- Kaczynski, A.T., Manske, S.R., Mannell, R.C., Grewel, K., 2008. Smoking and physical activity: a systematic review. *Am. J. Health Behav.* 32 (1), 93–110.
- Kim, J., Park, D.B., Seo, J.I., 2020. Exploring the relationship between forest structure and health. *Forests* 11 (12), 1–14. <https://doi.org/10.3390/f11121264>.
- Kim-Mozeleski, J.E., Pandey, R., Tsoh, J.Y., 2019. Psychological distress and cigarette smoking among U.S. Households by income: considering the role of food insecurity. *Prev. Med. Rep.* 16, 100983. <https://doi.org/10.1016/j.pmedr.2019.100983>.
- Koricich, A., Chen, X., Hughes, R.P., 2018. Understanding the effects of rurality and socioeconomic status on college attendance and institutional choice in the United States. *Rev. High. Educ.* 41 (2), 281–305. <https://doi.org/10.1353/rhe.2018.0004>.
- Kovats, R.S., Hajat, S., 2008. Heat stress and public health: a critical review. *Annu. Rev. Public Health* 29, 41–55. <https://doi.org/10.1146/annurev.publhealth.29.020907.090843>.
- Lefcheck, J.S., 2016. piecewiseSEM: piecewise structural equation modelling in R for ecology, evolution, and systematics. *Methods Ecol. Evol.* 7, 573–579. <https://doi.org/10.1111/2041-210X.12512>.
- Leitão, P., Andrew, C., Engelhardt, E., Graham, C., Anne, C., Pinkert, S., Schröder, B., Voskamp, A., Hof, C., Fritz, S., 2019. Macroecology as a hub between research disciplines: opportunities, challenges and possible ways forward. *J. Biogeogr.* 1–3. <https://doi.org/10.1111/jbi.13751>.
- Li, X., Tian, H., Pan, S., Lu, C., 2022. Four-century history of land transformation by humans in the United States: 1630–2020. *Earth Syst. Sci. Data* <https://doi.org/10.5194/essd-2022-135>.
- Ligon, F.K., Dietrich, W.E., Trush, W.J., 1995. Downstream ecological effects of dams - a geomorphic perspective. *Bioscience* 45 (3), 183–192. <https://doi.org/10.2307/1312557>.
- Lindberg, T.T., Bernhardt, E.S., Bier, R., Helton, A.M., Brittany Merola, R., Vengosh, A., Di Giulio, R.T., 2011. Cumulative impacts of mountaintop mining on an Appalachian watershed. *Proc. Natl. Acad. Sci. U. S. A.* 108 (52), 20929–20934. <https://doi.org/10.1073/pnas.1112381108>.
- Loopstra, R., Tarasuk, V., 2013. Severity of household food insecurity is sensitive to change in household income and employment status among low-income families. *J. Nutr.* 143 (8), 1316–1323. <https://doi.org/10.3945/jn.113.175414>.
- Lytle, D.A., 2008. Life-history and behavioural adaptations to flow regime in aquatic insects. In: Lancaster, J., Briers, R.A. (Eds.), *Aquatic Insects: Challenges to Populations*. CAB International, pp. 122–138.
- Marani, M., Eltahir, E., Rinaldo, A., 2001. Geomorphic controls on regional base flow. *Water Resour. Res.* 37 (10), 2619–2630. <https://doi.org/10.1029/2000WR001119>.
- Mattei, D., Ruscio, J., 2022. Rita: Automated Transformations, Normality Testing, and Reporting. R Package version 1.1.0. <http://cran.r-project.org/package=RITA>.
- Matthews, T., Lo, A.Y., Byrne, J.A., 2015. Reconceptualizing green infrastructure for climate change adaptation: barriers to adoption and drivers for uptake by spatial planners. *Landsc. Urban Plan.* 138, 155–163.
- McCormack, G.R., Shiell, A., 2011. In search of causality: a systematic review of the relationship between the built environment and physical activity among adults. *Int. J. Behav. Nutr. Phys. Act.* 8, 125.
- Mckay, L., Bondelid, T., Dewald, T., Johnston, C., Moore, R., Rea, A., 2012. NHDPlus Version 2: User Guide (Data Model Version 2.1). Horizon Systems.
- McMichael, A.J., 2013. Globalization, climate change, and human health. *N. Engl. J. Med.* 368 (14), 1335–1343. <https://doi.org/10.1056/nejrml109341>.
- McMichael, A.J., Woodruff, R.E., Hales, S., 2006. Climate change and human health: present and future risks. *Lancet* 367, 859–869.
- Mendelsohn, R., 2008. The impact of climate change on agriculture in developing countries. *J. Nat. Resour. Policy Res.* 1 (1), 5–19. <https://doi.org/10.1080/19390450802495882>.
- Milstein, G., Palitsky, R., Cuevas, A., 2020. The religion variable in community health promotion and illness prevention. *J. Prev. Interv. Community* 48 (1), 1–6. <https://doi.org/10.1080/10852352.2019.1617519>.
- Molenat, J., Gascuel-Oudoux, C., Ruiz, L., Gruau, G., 2008. Role of water table dynamics on stream nitrate export and concentration in agricultural headwater catchment (France). *J. Hydrol.* 348 (3–4), 363–378. <https://doi.org/10.1016/j.jhydrol.2007.10.005>.
- Muldavin, E., Milford, E.R., Grunau, L., Rondeau, R., 2019. Middle Rio Grande Conservation Action Plan: Framework and Status Assessment. Natural Heritage New Mexico Report No. 19-421. Museum of Southwestern Biology, University of New Mexico, Albuquerque, NM. https://nhnm.unm.edu/sites/default/files/nonsensitive/publications/MRGCAP_Ver_1.2_2019.pdf.
- Myers, S.S., Bernstein, A., 2011. The coming health crisis: Indirect health effects of global climate change. *F1000 Biology Reports*. 3. <https://doi.org/10.3410/B3-3> Issue 3.
- Myers, S.S., Gaffikin, L., Golden, C.D., Ostfeld, R.S., Redford, K.H., Ricketts, T.H., Turner, W.R., Ososky, S.A., 2013. Human health impacts of ecosystem alteration. *Proc. Natl. Acad. Sci. U. S. A.* 110 (47), 18753–18760. <https://doi.org/10.1073/pnas.1218656110>.
- Nelson, K.S., Nguyen, T.D., Brownstein, N.A., Garcia, D., Walker, H.C., Watson, J.T., Xin, A., 2021. Definitions, measures, and uses of rurality: a systematic review of the empirical and quantitative literature. *J. Rural. Stud.* 82, 351–365. <https://doi.org/10.1016/j.jrurstud.2021.01.035>.
- Nguyen, P.-Y., Astell-Burt, T., Rahimi-Ardabili, H., Feng, X., 2021. Green space quality and health: a systematic review. *Int. J. Environ. Res. Public Health* 18 (21), 11028. <https://doi.org/10.3390/ijerph182111028>.
- Nislow, K.H., Lowe, W.H., 2006. Influences of logging history and riparian forest characteristics on macroinvertebrates and brook trout (*Salvelinus fontinalis*) in headwater streams (New Hampshire, U.S.A.). *Freshw. Biol.* 51 (2), 388–397. <https://doi.org/10.1111/j.1365-2427.2005.01492.x>.
- Noteboom, M., Seidou, O., Lapen, D.R., 2021. Predicting water quality trends resulting from forest cover change in an agriculturally dominated river basin in eastern Ontario, Canada. *Water Quality Research Journal* 56 (4), 218–238. <https://doi.org/10.2166/wqrj.2021.010>.
- Omerik, J.M., 1987. Ecoregions of the conterminous United States. *Ann. Assoc. Am. Geogr.* 77 (1), 118–125.
- Paffenbarger, R., Hyde, R., Alvin, W., Lee, I.-M., Jung, D., Kampert, J., 1993. The association of changes in physical-activity level and other lifestyle characteristics with mortality among men. *N. Engl. J. Med.* 328 (8), 538–545.
- Palmer, M., Bernhardt, E., Schlesinger, W., Eshleman, K., Foufloula-Georgiou, E., Hendryx, M., Lemly, A., Likens, G., Loucks, O., Power, M., White, P., Wilcock, P., 2010. Mountaintop mining consequences. *Science* 327 (5962), 148–149. <https://doi.org/10.1021/es1036018>.
- Pate, R.R., Pratt, M., Blair, S.N., Haskell, W.L., Macera, C.A., Bouchard, C., Buchner, D., Ettinger, W., Heath, G.W., King, A.C., Wilmore, J.H., 1995. Public health and prevention and the American College of Sports Medicine. *Journal of American Medical Association* 273 (5), 402–407.
- Patrick, C.J., McGarvey, D.J., Larson, J.H., Cross, W.F., Allen, D.C., Benke, A.C., Brey, T., Huryn, A.D., Jones, J., Murphy, C.A., Woodward, G., 2019. Precipitation and temperature

- drive continental-scale patterns in stream invertebrate production. *Science Advances* 5 (4), 1–10. <https://doi.org/10.1126/sciadv.aav2348>.
- Pellow, D.N., 2016. Environmental justice and rural studies: a critical conversation and invitation to collaboration. *J. Rural. Stud.* 47, 381–386. <https://doi.org/10.1016/j.jrurstud.2016.06.018>.
- Pimentel, D., Kounang, N., 1998. Ecology of soil erosion in ecosystems. *Ecosystems* 1 (5), 416–426. <https://doi.org/10.1007/s100219900035>.
- Piontak, J.R., Schulman, M.D., 2014. Food insecurity in rural America. *Contexts* 13 (3), 75–77. <https://doi.org/10.1177/1536504214545766>.
- Poff, L.N., Allan, J.D., 1995. Functional organization of stream fish assemblages in relation to hydrological variability. *Ecology* 76 (2), 606–627.
- Poff, N.L., Allan, J.D., Bain, M.B., Karr, J.R., Prestegard, K.L., Richter, B.D., Sparks, R.E., Stromberg, J.C., 1997. The natural flow regime. *Bioscience* 47 (11), 769–784. <https://doi.org/10.2307/1313099>.
- Pond, G.J., Passmore, M.E., Borsuk, F.A., Reynolds, L., Rose, C.J., 2008. Downstream effects of mountaintop coal mining: comparing biological conditions using family- and genus-level macroinvertebrate bioassessment tools. *J. N. Am. Benthol. Soc.* 27 (3), 717–737. <https://doi.org/10.1899/08-015.1>.
- Portier, C.J., Thigpen, K., Carter, S.R., Dilworth, C.H., Grambsch, A.E., Gohlke, J., Hess, J., Howard, S.N., Luber, G., Lutz, J.T., Whung, P.-Y., 2013. A human health perspective on climate change: a report outlining the research needs on the human health effects of climate change. *J. Curr. Issues Glob.* 6 (4), 621–710.
- Pyne, M.J., Poff, L., 2017. Vulnerability of stream community composition and function to projected thermal warming and hydrologic change across ecoregions in the western United States. *Glob. Chang. Biol.* 23, 77–93. <https://doi.org/10.1111/gcb.13437>.
- R Core Team, 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing. <https://www.r-project.org/>.
- Rapport, D.J., Costanza, R., McMichael, A.J., 1998. Assessing ecosystem health. *Trends Ecol. Evol.* 13 (10), 397–402. [https://doi.org/10.1016/S0169-5347\(98\)01449-9](https://doi.org/10.1016/S0169-5347(98)01449-9).
- Remington, P.L., Catlin, B.B., Gennuso, K.P., 2015. The county health rankings: rationale and methods. *Popul. Health Metrics* 13, 11. <https://doi.org/10.1186/s12963-015-0044-2>.
- Riseng, C.M., Wiley, M.J., Black, R.W., Munn, M.D., 2011. Impacts of agricultural land use on biological integrity: a causal analysis. *Ecol. Appl.* 21 (8), 3128–3146. <https://doi.org/10.1890/11-0077.1>.
- Romanello, M., McGushin, A., Di Napoli, C., Drummond, P., Hughes, N., Jamart, L., Kennard, H., Lampard, P., Solano Rodriguez, B., Hamilton, I., 2021. The 2021 report of the lancet countdown on health and climate change: code read for a healthy future. *Lancet* 398 (10311), 1619–1662.
- Rosi-Marshall, E.J., Royer, T.V., 2012. Pharmaceutical compounds and ecosystem function: an emerging research challenge for aquatic ecologists. *Ecosystems* 15, 867–880. <https://doi.org/10.1007/s10021-012-9553-z>.
- Rosseel, Y., 2012. Lavaan : an R package for structural equation modeling. *J. Stat. Softw.* 48 (2), 1–36. <https://doi.org/10.18637/jss.v048.i02>.
- Roth, N., Allan, J., Erickson, D., 1996. Landscape influences on stream biotic integrity assessed at multiple spatial scales. *Landsc. Ecol.* 11 (3), 141–156.
- Ruaro, R., Gubiani, É.A., Hughes, R.M., Mormul, R.P., 2020. Global trends and challenges in multimetric indices of biological condition. *Ecol. Indic.* 110, 105862. <https://doi.org/10.1016/j.ecolind.2019.105862>.
- Russette, H., Graham, J., Holden, Z., Semmens, E., Williams, E., Landguth, E., 2021. Greenspace exposure and COVID-19 mortality in the United States: january-july 2020. *Environ. Res.* 198 (111195).
- Schade, J.D., Fisher, S.G., Grimm, N.B., Seddon, J.A., 2001. The influence of a riparian shrub on nitrogen cycling in a Sonoran Desert stream. *Ecol. Soc. Am.* 82 (12), 3363–3376.
- Sindall, R., Mecrow, T., Queiroga, A.C., Boyer, C., Koon, W., Peden, A.E., 2022. Drowning risk and climate change: a state-of-the-art review. *Injury Prevent.* 28, 185–191. <https://doi.org/10.1136/injuryprev-2021-044486>.
- Skurnik, Y., Shoenfeld, Y., 1998. Health effects of cigarette smoking. *Clin. Dermatol.* 16 (5), 545–556. [https://doi.org/10.1016/S0738-081X\(98\)00037-6](https://doi.org/10.1016/S0738-081X(98)00037-6).
- Solomon, G., Manson, E., 1997. Obesity and mortality: a review of the epidemiologic data. *Am. J. Clin. Nutr.* 66, 1044S–1050S.
- Sophocleous, M., 2002. Interactions between groundwater and surface water: the state of the science. *Hydrogeol. J.* 10 (1), 52–67. <https://doi.org/10.1007/s10040-001-0170-8>.
- Sponseller, R.A., Benfield, E.F., Valett, H.M., 2001. Relationships between land use, spatial scale and stream macroinvertebrate communities. *Freshw. Biol.* 46, 1409–1424. <https://doi.org/10.1046/j.1365-2427.2001.00758.x>.
- Stoddard, J.L., Herlihy, A.T., Peck, D.V., Hughes, R.M., Whittier, T.R., Tarquinio, E., 2008. A process for creating multimetric indices for large-scale aquatic surveys. *J. N. Am. Benthol. Soc.* 27 (4), 878–891. <https://doi.org/10.1899/08-053.1>.
- Strayer, D.L., Beighley, R.E., Thompson, L.C., Brooks, S., Nilsson, C., Pinay, G., Naiman, R.J., 2003. Effects of land cover on stream ecosystems: roles of empirical models and scaling issues. *Ecosystems* 6 (5), 407–423. <https://doi.org/10.1007/s10021-002-0170-0>.
- Sun, Y., Liu, B., Rong, S., Du, Y., Xu, G., Snetselaar, L.G., Wallace, R.B., Bao, W., 2020. Food insecurity is associated with cardiovascular and all-cause mortality among adults in the United States. *J. Am. Heart Assoc.* 9 (19), e014629. <https://doi.org/10.1161/JAHA.119.014629>.
- Sunderlin, W.D., Dewi, S., Puntodewo, A., 2007. Poverty and forests: Multi-country analysis of spatial association and proposed policy solutions. Center for International Forestry Research. <https://www.cifor.org/knowledge/publication/2297/>.
- Triska, F.J., Sedell, J.R., Gregory, S.V., 1982. Coniferous forest streams. In: Edmonds, R.L. (Ed.), *Analysis of Coniferous Forest Ecosystems in the Western United States*. Hutchinson Ross Publishing Company, pp. 292–332.
- Tromboni, F., Liu, J., Ziaco, E., Breshears, D.D., Thompson, K.L., Dodds, W.K., Dahlin, K.M., LaRue, E.A., Thorp, J.H., Viña, A., Laguë, M.M., Fei, S., 2021. Macrosystems as metacoupled human and natural systems. *Frontiers in Ecology and the Environment* 19 (1), 20–29. <https://doi.org/10.1002/fee.2289>.
- Tweeten, L., 1999. The economics of global food security. *Appl. Econ. Perspect. Policy* 21 (2), 473–488. <https://doi.org/10.2307/1349892>.
- Tzoulas, K., Korpela, K., Venn, S., Yli-Pelkonen, V., Kaźmierczak, A., Niemela, J., James, P., 2007. Promoting ecosystem and human health in urban areas using green infrastructure: a literature review. *Landsc. Urban Plan.* 81 (3), 167–178. <https://doi.org/10.1016/j.landurbplan.2007.02.001>.
- Urban, M.C., Skelly, D.K., Burchsted, D., Price, W., Lowry, S., 2006. Stream communities across a rural-urban landscape gradient. *Divers. Distrib.* 12 (4), 337–350. <https://doi.org/10.1111/j.1366-9516.2005.00226.x>.
- Van Dyne, G.M., 1966. Ecosystems, Systems Ecology, and Systems Ecologists. ORNL-3957. Oak Ridge National Laboratory, Oak Ridge, Tennessee <https://doi.org/10.2172/4516278>.
- Verberk, W.C.E.P., Bilton, D.T., 2013. Respiratory control in aquatic insects dictates their vulnerability to global warming. *Biol. Lett.* 9, 1–4.
- Violin, C.R., Cada, P., Sudduth, E.B., Hassett, B.A., Penrose, D.L., Bernhardt, E.S., 2011. Effects of urbanization and urban stream restoration on the physical and biological structure of stream ecosystems. *Ecol. Appl.* 21 (6), 1932–1949. <https://doi.org/10.1890/10-1551.1>.
- Walls, F.N., McGarvey, D.J., 2023. Building a macrosystems ecology framework to identify links between environmental and human health: a random forest modeling approach. *People and Nature* 5 (1), 183–197.
- Webster, J.R., Meyer, J.L., 1997. Organic matter budgets for streams: a synthesis. *J. N. Am. Benthol. Soc.* 16 (1), 141–161.
- Wheeler, B.W., Lovell, R., Higgins, S.L., White, M.P., Alcock, I., Osborne, N.J., Husk, K., Sabel, C.E., Depledge, M.H., 2015. Beyond greenspace: an ecological study of population general health and indicators of natural environment type and quality. *Int. J. Health Geogr.* 14 (1), 17. <https://doi.org/10.1186/s12942-015-0009-5>.
- Wood, P.J., Armitage, P.D., 1997. Biological effects of fine sediment in the lotic environment. *Environ. Manag.* 21 (2), 203–217. <https://doi.org/10.1007/s002679900019>.
- Zhang, M., Ning, L., Harper, R., Li, Q., Liu, K., Wei, X., Ning, D., Hou, Y., Liu, S., 2017. A global review on hydrological responses to forest change across multiple spatial scales: importance of scale, climate, forest type and hydrological regime. *J. Hydrol.* 44–59.
- Zhang, J., Xu, W., Liao, X., Zong, S., Liu, B., 2021. Global mortality risk assessment from river flooding under climate change. *Environ. Res. Lett.* 16 (6). <https://doi.org/10.1088/1748-9326/abff87>.
- Zhu, Z., Evans, D.L., 1994. US forest types and predicted percent forest cover from AVHRR data. *Photogramm. Eng. Remote Sens.* 60 (5), 525–531.
- Zhu, P., Zhang, Y., 2008. Demand for urban forests in United States cities. *Landsc. Urban Plan.* 84 (3), 293–300. <https://doi.org/10.1016/j.landurbplan.2007.09.005>.

Chapter 3

Systems-level models accounting for racial and regional differences in links between the environment, socioeconomic factors, and human mortality

By: Felisha Nicole Marie Walls

Major Adviser: Dr. Daniel J. McGarvey, Center for Environmental Studies

Abstract

Continental scale models of linked human–environmental health have shown that environmental disturbances, such as land use alterations and climate change, can have significant consequences for human health. However, these human–environment links may vary among demographic groups and geographic regions that experience different environmental and socioeconomic conditions. In this study, we develop a macrosystems ecology framework to evaluate differences in environmental and socioeconomic influences on human health outcomes among Black and White residents within three subregions of the conterminous United States. We begin by compiling a large database of environmental variables, socioeconomic variables, and age-adjusted mortality rates within U.S. counties. Next, we parse the data into six distinct subsets, representing three regions (Northeast, Southeast, and Southern Plains) and two demographic groups (Black and White residents). Six independent random forest models are then built to identify the strongest predictors of mortality for each regional and demographic subgroup (3 regions \times 2 races). Finally, six independent path analysis models are built and used to explore multi-step pathways among the environmental and socioeconomic factors that influence mortality rates for each race and in each region. Results among regions and races are highly variable but include several key insights: (1) socioeconomic influences, particularly education and income, are consistently the strongest predictors of mortality rate; (2) environmental variables, particularly land cover characteristics, have predictable influences on education and income; and (3) the climate variables air temperature and precipitation have critical, regulatory effects on all environmental, socioeconomic, and human health variables.

Introduction

Faced with growing evidence that environmental perturbations, such as climate change and deforestation, can have significant consequences for human health (Ebi et al., 2018; Guernier et al., 2004; Romanello et al., 2021), many ecologists and environmental scientists have extended their work to incorporate indicators and drivers of human well-being. These expanded efforts range from conceptual models and general advice for applying ecological tools to new human health questions (e.g., Burnside et al., 2012; Ulrich et al., 2022) to granular, empirical studies of linked human-environmental health in a specific region (e.g., Angermeier et al., 2021) or natural resource context (e.g., Engemann et al., 2019). They also seek to break down traditional, disciplinary boundaries between social and ecological systems research (e.g., Dunham et al., 2018; Fleishman & Brown, 2019).

Two recent studies used a macrosystems ecology approach (*sensu* Heffernan et al., 2014), combining large-scale, multidisciplinary data sets at the continental scale, to identify links between environmental conditions, socioeconomic factors and human health. Walls and McGarvey (2023a) began by assembling a set of 76 socioeconomic and environmental covariates, then using them as predictors in a random forest (RF) model of age-adjusted mortality rate (AAMR) in counties throughout the conterminous United States (CONUS). Their model accounted for 76% of the CONUS-scale variation in AAMR and identified a shortlist of socioeconomic, climatic, geologic, and land use covariates with the strongest individual effects on AAMR. Walls and McGarvey (2023b) then extended the RF modeling results by organizing the most influential AAMR predictors within a path analysis (PA) model. This produced a quantitative, systems-level map of direct and indirect effects of each covariate on AAMR. For instance, their path model showed that smoking has a stronger effect on AAMR than poverty,

food insecurity, obesity, or level of physical activity, but smoking prevalence was itself influenced by poverty and food insecurity, as well as land cover and air temperature.

One limitation of the national-scale AAMR model of Walls and McGarvey (2023b) is the potential for CONUS-level data to obscure key differences in environment—health relationships among subregions and/or demographic subgroups. Clear spatial trends in human mortality rate have previously been reported among U.S. counties, with highest rates clustered in the Southeast and Appalachian regions (Dwyer-Lindgren et al., 2017; Ezzati et al., 2008). This spatial clustering is likely to reflect regional differences in the environmental and socioeconomic factors that influence mortality. For example, high mortality rates in Appalachia are linked to the historical legacies of coal mining and industrial logging, which capitalized on the region’s abundant natural resources but left many local communities isolated and impoverished (Borak et al., 2012; Partridge et al., 2013).

Furthermore, consistent differences in AAMR have been documented among racial groups, with lowest mortality rates most often exhibited by Whites and significantly higher rates exhibited by Blacks (Cunningham et al., 2017). This racial disparity is attributed to multiple socioeconomic and environmental causes, such as unequal healthcare access (Sudano & Baker, 2006), differing exposure to heat stress (Graff et al., 2021), and a discriminatory history of toxic waste disposal (Bullard, 1993). It is also persistent among studies that focus on state (Harper et al., 2014), county (Davids et al., 2014), and city-level demographics (Benjamins et al., 2021).

Regional and racial trends in AAMR suggest that sub-models may be necessary to fully understand environmental-health relationships within the CONUS. Our present objective is therefore to identify potential differences in the effects that environmental and socioeconomic factors have on AAMR, comparing results among CONUS subregions and among White and

Black populations. We accomplish this through a three-stage process. First, we delineate three subregions – the Northeast, Southeast, and Southern Plains (Fig. 1) – that allow us to focus on finer-scale environmental and socioeconomic effects on AAMR than the CONUS-scale data used by Walls and McGarvey (2023b). Next, we assemble a set of 85 covariates, representing five key themes of information (geologic, climatic, hydrologic, land cover, and socioeconomic covariates), and use RF modeling to identify the strongest predictors of AAMR. Separate RF models are built for White and Black AAMR values in each of the three subregions (i.e., six independent models). Finally, independent PA models are built for White and Black residents of each subregion, testing for racial or regional differences in the multi-step pathways that link environmental and socioeconomic factors to AAMR.

Materials and Methods

Data collection and management

A complete record of all environmental, socioeconomic, and mortality data used herein is provided in Walls and McGarvey (2023a, b). In summary, environmental data were downloaded (Feb. 2020) from two sources: (1) the 1:100,000 scale National Hydrography Dataset (NHD), Version 2 digital stream network (Mckay et al., 2012); and (2) the US Environmental Protection Agency Stream-Catchment (StreamCat) database (Hill et al., 2016). The StreamCat data represent a broad range of physicochemical characteristics (water quality, land cover, geology and morphology, etc.) and were linked to discrete NHD stream/river segments with unique ‘COMID’ index codes. The environmental data were superimposed on a digital map of US counties, then three county-level summary statistics – the minimum, median, and maximum of multiple stream/river segments within a given county – were calculated for each environmental

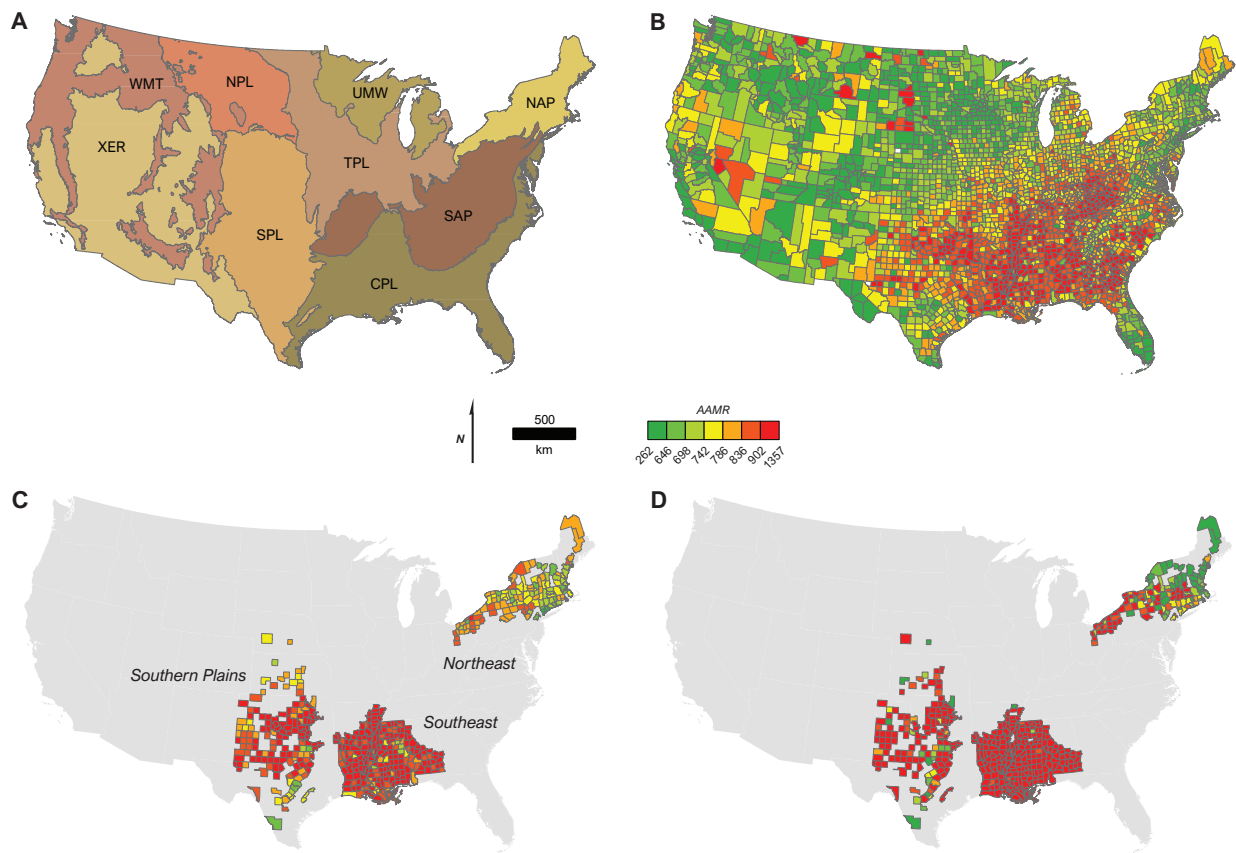


Fig. 1. Ecoregion and age-adjusted mortality rate (AAMR) maps. Panel A shows the spatial boundaries of the nine aggregate ecoregions of Herlihy et al. (2008). Ecoregion labels are: WMT – Western Mountains (Northwest); XER – Xeric; NPL – Northern Plains; SPL – Southern Plains; TPL – Temperate Plains; UMW – Upper Midwest; CPL – Coastal Plain (Southeast); SAP – Southern Appalachians; NAP – Northern Appalachians (Northeast). Panel B shows the county-level distribution of total AAMR (all demographic groups combined) throughout the conterminous U.S. Panel C shows AAMR for White residents of subset counties in the Northeast, Southeast, and Southern Plains subregions. Panel D shows AAMR for Black residents in each of the three subregions. All maps are shown at a common scale. The color scheme used to illustrate AAMR data is identical in panels B-D.

covariate (see Fig. 1 in Walls & McGarvey, 2023b). This resulted in a list of 596 environmental covariates for every county within the CONUS. Principal component analysis was then used to select a subset of 85 covariates (Table S1) that represented much of the information in the complete set of environmental covariates while minimizing collinearity ($|r| < 0.7$) among covariates.

County-level socioeconomic data were downloaded (Dec. 2021) from four sources. Prevalence of adults with a bachelor's degree, adults without health insurance, and the percentage of renter occupied housing were obtained from the U.S. Census Bureau, American Community Survey (www.census.gov/programs-surveys/acs/). Prevalence of adult smokers, physical inactivity, food insecurity, violent crime, and housing shortages were downloaded from the County Health Rankings and Roadmaps website (www.countyhealthrankings.org; Remington et al., 2015). Degree of ruralness and social association rate were also obtained from the County Health Rankings. Data on religious activity and affiliation were obtained from the U.S. Religious Census (Grammich et al., 2012). Finally, the county-level Leaf Area Index (LAI) was downloaded from (Russette et al., 2021) and used as a measure of green space. All socioeconomic covariates are listed with units of measure in Table S1.

Age-adjusted mortality rate, representing all documented causes of mortality, was downloaded (July 2022) for all CONUS counties from the US Centers for Disease Control and Prevention, Wide-ranging Online Data for Epidemiologic Research platform (<https://wonder.cdc.gov/>). Separate mortality rates were downloaded for two demographic or racial subgroups: non-Hispanic Whites (W-AAMR) and Blacks or African Americans (B-AAMR). All AAMR values were calculated as the average number of deaths per 100,000 White or Black residents within the county between 1999-2017.

When all environmental, socioeconomic, and AAMR data were aligned within CONUS counties, we used imputation to replace missing values. Approximately 0.2% of all county \times covariate entries were missing values. Imputation was performed with the ‘missForest’ package in R (Stekhoven & Bühlmann, 2012). All covariates were then transformed to improve normality, using the ‘Rita’ package (Mattei & Ruscio, 2022) in R. The B-AAMR and W-AAMR data were already normally distributed and not transformed.

Delineating CONUS subregions

We used the nine ecoregions of (Herlihy et al., 2008) as a starting point to delineate subregions (Fig. 1A). The nine ecoregions are aggregates of the original level III ecoregions of (Omernik, 1987). They provide a simplified, regional classification scheme while still capturing many of the historic, climatic, and geologic influences that regulate terrestrial and aquatic environments throughout the CONUS. The nine aggregate ecoregions have also been used to account for underlying biogeographical influences in national-scale studies of physical and biological systems. For example, Hill et al. (2017) modeled the health of macroinvertebrate assemblages in freshwater systems throughout the CONUS, taking care to specify unique models for each of the nine ecoregions. Similarly, Kaufmann et al. (2022) modeled physical habitat in a national sample of streams and rivers, with separate models developed for each ecoregion.

From the nine ecoregions, three were selected to represent a broad range of environmental, socioeconomic, and human health conditions: the Northeast (Northern Appalachians in Fig. 1A), Southeast (Coastal Plain in Fig. 1A), and Southern Plains. Summary statistics comparing a suite of environmental and socioeconomic covariates among the three selected ecoregions are provided in Table 1. Next, we removed any counties that did not have

Table 1. Summary statistics (minimum, median, maximum, and transformation) for select variables in each of the three subregions. Variables are organized by class (Human health, Geology/Morphology, Climate, Land cover/Land use, and Socioeconomic) with units of measure shown in parentheses. Separate age-adjusted mortality rate (AAMR) statistics are shown for Black and White residents. The 20 environmental and socioeconomic variables listed here were used in the subregional path analysis models of AAMR.

Variable class	Variable	Statistic	Subregion		
			So. Plains	Southeast	Northeast
Human health	AAMR Black (deaths per 100,000)	Min.	0.50	0.64	0.41
		Median	1.04	1.11	0.78
		Max.	3.13	1.39	1.42
		S.D.	0.28	0.11	0.19
		Transformation	NA	NA	NA
	AAMR White (deaths per 100,000)	Min.	0.65	0.71	0.61
		Median	0.90	0.95	0.78
		Max.	1.15	1.21	0.91
		S.D.	0.11	0.09	0.06
		Transformation	NA	NA	NA
Geology	Elevation (m)	Min.	-2.09	-2.24	-2.53
		Median	0.24	0.03	0.05
		Max.	1.43	4.14	2.57
		S.D.	1	1	1
		Transformation	Log	Arc-sine	Arc-sine
Climate	Air temperature (C°)	Min.	-2.64	-3.48	-2.29
		Median	-0.16	0.11	-0.18
		Max.	3.63	1.98	3.97
		S.D.	1	1	1
		Transformation	Logit	Arc-sine	Logit
	Precipitation (mm per yr)	Min.	-1.57	-3.86	-1.84
		Median	-0.22	-0.01	-0.13
		Max.	1.77	1.78	2.05
		S.D.	1	1	1
		Transformation	Logit	Arc-sine	Logit
Land cover	Crop cover (% by area)	Min.	-1.50	-1.04	-1.07
		Median	0.19	0.02	-0.04
		Max.	2.25	3.24	3.58
		S.D.	1	1	1
		Transformation	Arc-sine	Logit	Logit
	Forest cover (% by area)	Min.	-1.17	-1.45	-2.87
		Median	-0.20	0.39	0.21
		Max.	2.38	2.07	2.41
		S.D.	1	1	1
		Transformation	Log	Arc-sine	Logit

Table 1. Continued.

	Pesticide use (kg/km ² /yr)	Min.	-2.29	-3.51	-3.70
		Median	0.15	0.33	0.15
		Max.	3.20	2.51	1.80
		S.D.	1	1	1
		Transformation	Arc-sine	Log	Log
	Road density (km/km ²)	Min.	-2.18	-3.05	-2.22
		Median	-0.15	0.09	-0.19
		Max.	5.69	2.91	4.45
		S.D.	1	1	1
		Transformation	Sqrt	Log	Log
Socioeconomic	College degree (% by population)	Min.	-3.27	-4.44	-2.93
		Median	-0.15	-0.26	-0.04
		Max.	3.15	3.79	2.13
		S.D.	1	1	1
		Transformation	Sqrt	Sqrt	Logit
	Food insecure (% by population)	Min.	-2.25	-3.99	-2.87
		Median	-0.22	-0.10	0.19
		Max.	2.83	3.02	2.89
		S.D.	1	1	1
		Transformation	Logit	Arc-sine	Logit
	High school education (% by population)	Min.	-2.40	-5.09	-2.86
		Median	-0.07	0.12	0.08
		Max.	4.38	2.21	1.85
		S.D.	1	1	1
		Transformation	Logit	Arc-sine	Arc-sine
	Physically inactive (% by population)	Min.	-4.75	-2.28	-2.01
		Median	-0.13	-0.00	-0.01
		Max.	2.67	4.27	3.13
		S.D.	1	1	1
		Transformation	Arc-sine	Logit	Logit
	No high school diploma (% by population)	Min.	-2.71	0.00	-2.78
		Median	-0.16	32.02	0.09
		Max.	2.31	74.53	2.42
		S.D.	1	24.59	1
		Transformation	Sqrt	Arc-sine	Arc-sine
	Obesity (% by population)	Min.	-4.86	-3.30	-4.12
		Median	-0.09	0.13	0.14
		Max.	2.32	3.91	1.68
		S.D.	1	1	1
		Transformation	Arc-sine	Arc-sine	Arc-sine

Table 1. Continued.

Rural (% by population)	Min.	-2.81	-2.65	-2.23
	Median	0.09	-0.06	0.13
	Max.	1.62	1.49	2.47
	S.D.	1	1	1
	Transformation	Sqrt	Arc-sine	Arc-sine
Physical illness (avg. no./month)	Min.	-2.08	-3.84	-2.18
	Median	-0.07	0.06	-0.10
	Max.	3.13	2.47	2.83
	S.D.	1	1	1
	Transformation	Logit	Log	Logit
Population (no./km ²)	Min.	-1.57	-1.87	-4.97
	Median	-0.25	-0.20	0.12
	Max.	2.88	3.26	2.33
	S.D.	1	1	1
	Transformation	Sqrt	Sqrt	Log
Poverty (% by population)	Min.	-3.73	-3.78	-3.34
	Median	-0.03	-0.01	0.02
	Max.	4.09	2.87	2.23
	S.D.	1	1	1
	Transformation	Arc-sine	Sqrt	Arc-sine
Religion (% by population)	Min.	-3.57	-4.12	-2.71
	Median	-0.08	-0.00	-0.03
	Max.	3.34	3.79	3.27
	S.D.	1	1	1
	Transformation	Sqrt	Logit	Logit
Smoking (% by population)	Min.	-3.56	-4.54	-3.81
	Median	-0.04	0.01	0.17
	Max.	2.82	3.47	1.91
	S.D.	1	1	1
	Transformation	Arc-sine	Arc-sine	Arc-sine
Violent crime (% by population × 100,000)	Min.	-3.08	-3.07	-2.49
	Median	-0.12	-0.09	-0.09
	Max.	2.93	3.29	3.43
	S.D.	1	1	1
	Transformation	Sqrt	Sqrt	Sqrt

paired W-AAMR and B-AAMR data (some counties did not have enough Black residents to report B-AAMR). Counties that crossed ecoregion boundaries were partitioned with a ‘majority rule’ spatial overlay process: all counties that were at least 50% contained (by surface area) within a given ecoregion were included in the respective query. Given the particularly large spatial extents of the Southeast and Southern Plains ecoregions (Fig. 1A), we further subset their constituent counties by states. In the Southeast, we included counties within Arkansas, Louisiana, Mississippi, and Alabama. This provided a large sample of counties ($n = 223$ counties) within the lower Mississippi and Mobile River Basins. In the Southern Plains, we included counties within Texas, Oklahoma, Kansas, and Nebraska ($n = 155$). The Northeast subregion was inclusive of all counties in the Northern Appalachian ecoregion ($n = 124$) between Maine and Ohio.

Next, we tested for spatial autocorrelation in AAMR. This was prudent because significant autocorrelation in county-level AAMR was previously reported for the entire CONUS ($p < 0.01$ for Moran’s I ; see Walls & McGarvey, 2023b) and the linear model equations used in path analysis (see below) are sensitive to autocorrelation. Separate autocorrelation tests were performed for each of the three subregions, using B-AAMR and W-AAMR as independent response variables. We first calculated centroid coordinates ([maximum longitude/latitude – minimum longitude/latitude] $\times 0.5$) for each county within the three subregions. We then compiled a list of neighbors for each county, where neighbors were defined as all counties with centroids that occurred within a 25 km radius of the centroid of a focal county. We did not define neighbors as counties with shared edges because some counties in the Northeast subregion were much larger (particularly in Maine) than counties in the Southeast or Southern Plains (Fig. 1B). A global Moran’s I statistic was calculated for each subregion and a threshold of $p < 0.05$ was

used with the I statistic to test for significant autocorrelation. Notably, we did not detect significant autocorrelation in the county-level B-AAMR or W-AAMR data in any of the three subregions (Table S2). All queried counties in each of the subregions were therefore used in RF and PA models.

Finally, we performed a paired samples t -test comparing B-AAMR and W-AAMR in each of the three subregions. In each test, the null hypothesis of no difference was compared to the 1-sided alternative: B-AAMR > W-AAMR. The a priori expectation that B-AAMR > W-AAMR was consistent with prior work documenting higher mortality rates for Blacks than Whites (e.g., Cunningham et al., 2017). Counties with paired B-AAMR and W-AAMR data (Fig. 1C, D) were used as replicates in each of the subregional tests; counties lacking paired B-AAMR and W-AAMR data were excluded.

Random forest modeling

The RF modeling process is fully described in Walls and McGarvey (2023a). Briefly, in each of the three subregions, independent RF models were built for B-AAMR and W-AAMR (i.e., six independent models). The complete list of 85 covariates (Table S1) was used to fit each model. All RF models included an arbitrarily large number of trees ($n_{tree} = 1000$) and the number of covariates tested at each branching node ($m_{try} = 28$) was approximately 1/3 of the total number of covariates. RF models were fit with the ‘randomForest’ package in R (Liaw & Wiener, 2002).

For each RF model, we recorded the percent increase in mean squared error (%IncMSE) statistic for each of the 85 covariates. We then used the %IncMSE values to determine which covariates had the strongest influences on B-AAMR and W-AAMR in each subregion. The

%IncMSE values were subsequently used to guide the selection of covariates for inclusion in PA models.

Path analysis

The general PA process of Walls and McGarvey (2023b) was adapted to model B-AAMR and W-AAMR within each of the three subregions. We began with a metamodel of nested links among environmental influences, socioeconomic factors, and AAMR (Fig. 2). This metamodel was similar to the metamodel of Walls and McGarvey (2023b; see their Fig. 1), but the present model focused exclusively on human health and did not include a direct measure of ecosystem health. This was prudent because the smaller numbers of counties within subregions constrained the sample sizes in subregional PA models, limiting our ability to fit complex models with multiple endpoints and many links.

Next, an exploratory process was used to model B-AAMR and W-AAMR in each of the three subregions, with independent path models fit to each of the six datasets (2 races \times 3 subregions). We selected individual covariates from each of five classes of covariates (geology/morphology, climate, hydrology, land cover/land use, socioeconomic) and oriented them within a path model, using the metamodel (Fig. 2) as a guide. Initial covariate selections were guided by %IncMSE results from each of the respective RF models; within covariate classes, we prioritized covariates with the highest %IncMSE values. Iterative decisions to retain, replace, or remove paths and covariates in PA models were informed by path significance levels (p -values), path modifications indices, and three indices of overall model fit: the Standardized Root Mean Squared Residual (SRMR), Tucker-Lewis Index (TLI) and Comparative Fit Index (CFI).

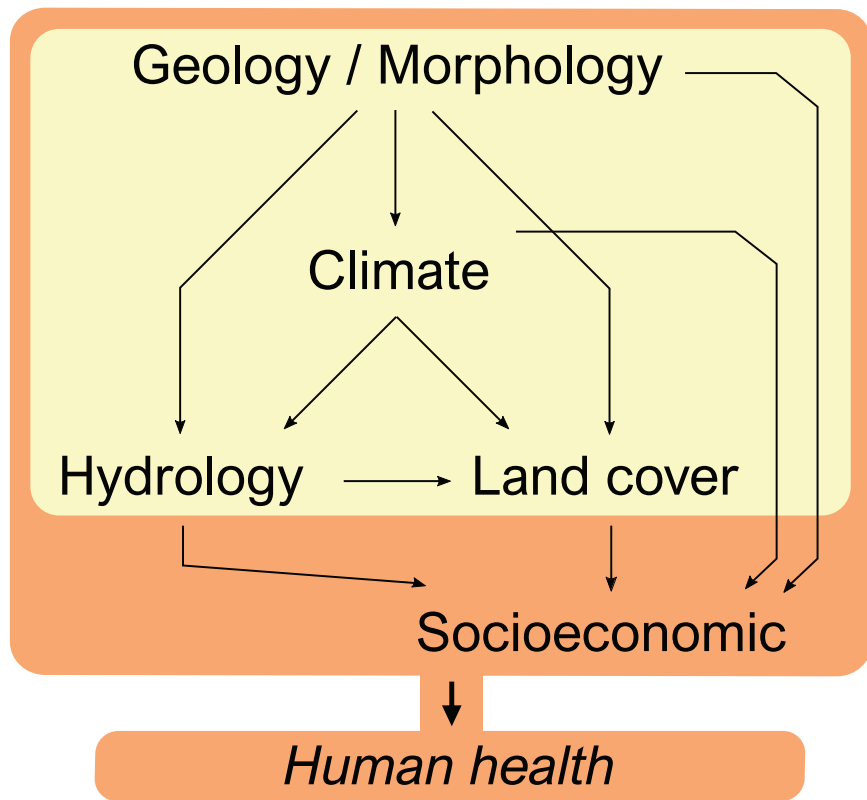


Fig. 2. Metamodel of hypothesized links between the five classes of covariates (geology/morphology, climate, hydrology, land cover/land use, socioeconomic) and human health.

Standard thresholds for ‘good’ model fit were used with each of the three fit indices (SRMR < 0.09; TLI \geq 0.90; CFI \geq 0.95; see Fan et al., 2016).

For each of the final PA models, we recorded the standardized path coefficient and p-value for every path, the coefficient of determination (R^2) for every endogenous variable, and the three indices of overall model fit. Prior to fitting PA models, all predictor variables were standardized (z -transformed) to prevent convergence errors that can result from large discrepancies in covariance (>3 orders of magnitude; see Walls and McGarvey, 2023b). PA modeling was performed with the ‘lavaan’ package in R (Rosseel, 2012).

Results

In each of the three subregions, B-AAMR was significantly higher than W-AAMR ($p < 0.05$; see Table 1).

Random forest models showed that in all subregions, socioeconomic variables were the strongest predictors of B-AAMR and W-AAMR. Lack of a high school diploma (No_HS_Edu), lack of health insurance (No_HI_cov_Percent), poverty (IncomeBelowPov_Percent), physical inactivity (Pct_Physically_Inactive), and mental illness (Mentally_Unhealthy_Days) were particularly important and consistently ranked among the top-10 predictors for a given race \times subregion model (Fig. 3). Among the remaining classes of covariates, geologic and land cover variables were generally the most influential. Nine geology variables and nine land cover variables were ranked in the top-10 for at least one of the six sub-models. Hydrology and climate covariates were each represented among the top-10 by at least a single variable: base flow (BFIWs_max) and air temperature (TmeanCat_median).

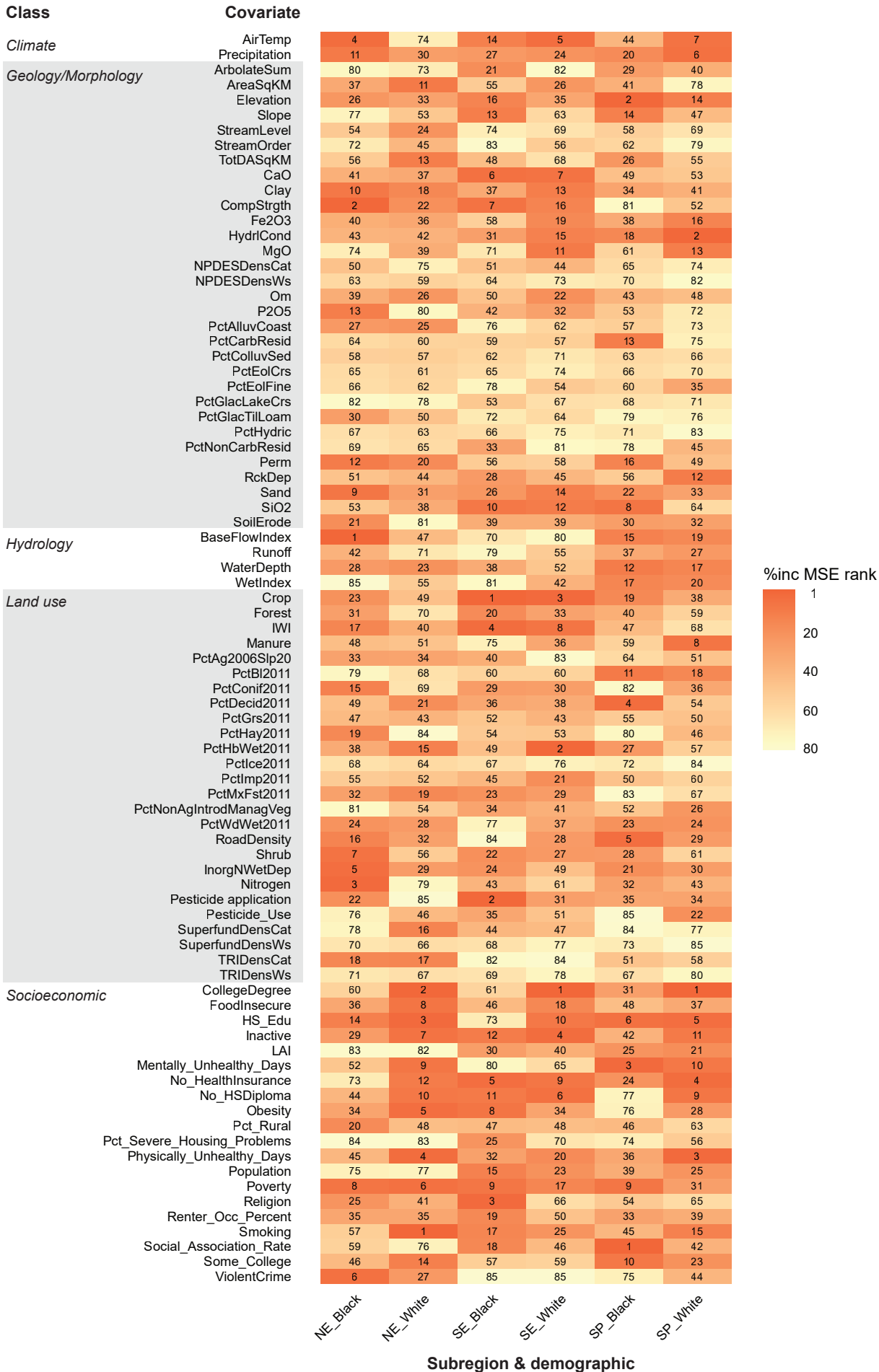


Fig. 3. Heatmap of covariate importance values, measured as the percent increase in mean squared error (%IncMSE) from random forest models. Each of the 85 covariates is ranked by %IncMSE (rank 1 = most important, or highest %IncMSE), with rankings shown independently (i.e., one set per column) for the six random forest models (3 subregions \times 2 demographic groups) tested in this study: Black residents (NE_Black) and White residents (NE_White) in Northeast counties; Black and White residents in Southeast counties (SE_Black and SE_White, respectively); and Black and White residents in Southern Plains counties (SP_Black and SP_White, respectively). Numbers shown in each cell are the %IncMSE ranks for a given model, with darker colors indicating more important covariates.

Within subregions, the strongest predictors of B-AAMR and W-AAMR differed. For example, each of the top-10 (ranked by %IncMSE) predictors of W-AAMR in the Northeast was a socioeconomic variable (Fig. 3), but only two of the top-10 B-AAMR predictors – violent crime and poverty – were socioeconomic. Instead, hydrologic and geologic variables, such as base flow and compressive strength (CompStrgthCat_Max), had the strongest overall influences on B-AAMR. In the Southeast, the land cover variables crop cover (PctCrop2011Ws_median) and pesticide use (Pestic97Ws_max) were the strongest predictors of B-AAMR. The socioeconomic and land cover variables, percentage of adults with a college degree (College_degree) and herbaceous wetland cover (PctHbWet2011Ws_median), were the highest ranked predictors of W-AAMR in the Southeast. In the Southern Plains, B-AAMR was most strongly influenced by a mix of socioeconomic (e.g., social association, Social_Association_Rate), geologic (e.g., elevation, ElevCat_median), and land cover (e.g., deciduous forest cover, PctDecid2011Ws_max) variables.

Guided by the RF model %IncMSE results, we constrained covariate selection in each of the six PA models to a pool of one geologic, two climatic, four land cover, and 13 socioeconomic covariates (see Table 1). In the final PA models, all paths were significant at $p \leq 0.05$ and overall model fit was consistently good, as indicated by the CFI, TLI, and SRMR statistics (Table 2). For B-AAMR models, R^2 values ranged from 0.09 (Southern Plains) to 0.43 (Northeast). For W-AAMR models, R^2 values ranged from 0.19 (Southeast) to 0.72 (Northeast).

Across all subregions, air temperature, poverty, and smoking (Pct_Smokers) had the strongest overall influences on B-AAMR. Air temperature had a direct, positive association with B-AAMR in the Northeast (Fig. 4B) and Southeast regions (Fig. 5B). In the Southern Plains, air

Table 2. Summary statistics for the six final path analysis models (see Figs. 4-6). Results include the number of counties (n) selected by the ‘majority rule’ spatial overlay process, three indices of model fit (SRMR, TLI, and CFI), and coefficients of determination (R^2) for the model response variable AAMR. Separate results are included for the the six demographic subgroup (Black and White) \times subregions (Northeast, Southeast, and Southern Plains).

Statistic	Northeast		Southeast		Southern Plains	
	Black	White	Black	White	Black	White
n	124	124	223	223	155	155
SRMR (< 0.09)	0.06	0.06	0.04	0.04	0.05	0.05
TLI (≥ 0.90)	0.95	0.97	0.98	0.98	1.00	0.97
CFI (≥ 0.95)	0.97	0.98	0.99	0.99	1.00	0.99
R^2 AAMR	0.43	0.72	0.26	0.19	0.09	0.48

temperature, poverty, and smoking were each significant predictors of physical illness (Physically_Unhealthy_Days), the only covariate to have a direct, positive effect on B-AAMR (Fig. 6B). For W-AAMR, physical inactivity was the only consistent predictor among subregions. Physical inactivity had a direct, positive association with W-AAMR in the Southeast (Fig. 5A) and Southern Plains (Fig. 6A). In the Northeast, physical inactivity was a significant predictor of obesity (Pct_Obese), which in turn had a direct effect on W-AAMR (Fig. 4A).

Other effects on AAMR were highly variable among subregions and demographic groups. In the Northeast, B-AAMR was significantly influenced by violent crime (Violent_Crime_Rate), physical inactivity, and forest cover (Fig. 4B). Religion, poverty, and obesity were significant predictors of B-AAMR in the Southeast (Fig. 5B). In the Southern Plains, B-AAMR was directly influenced by physical illness (Fig. 6B). In the Southeast, food insecurity (Pct_Food_Insecure) was the strongest predictor of W-AAMR (Fig. 5A). In the Southern Plains, college degree, physical inactivity, and precipitation (PrecipCat_median) were direct predictors of W-AAMR. (Fig. 6A) Smoking was the strongest predictor of W-AAMR in the Northeast, followed by obesity and poverty (Fig.4A).

Land cover variables played a critical role in each subregion, where they mediated indirect pathways between environmental and socioeconomic variables. In the Northeast, air temperature and precipitation were significant predictors of forest cover, while forest cover was a key predictor of socioeconomic covariates that directly influenced B-AAMR and W-AAMR (Fig. 4). In the Southeast, indirect links from elevation (ElevCat_median) and air temperature to socioeconomic variables were mediated by crop cover (Fig. 5). In the Southern Plains, road density (RdDensCat_median) mediated indirect links between the environmental variables and

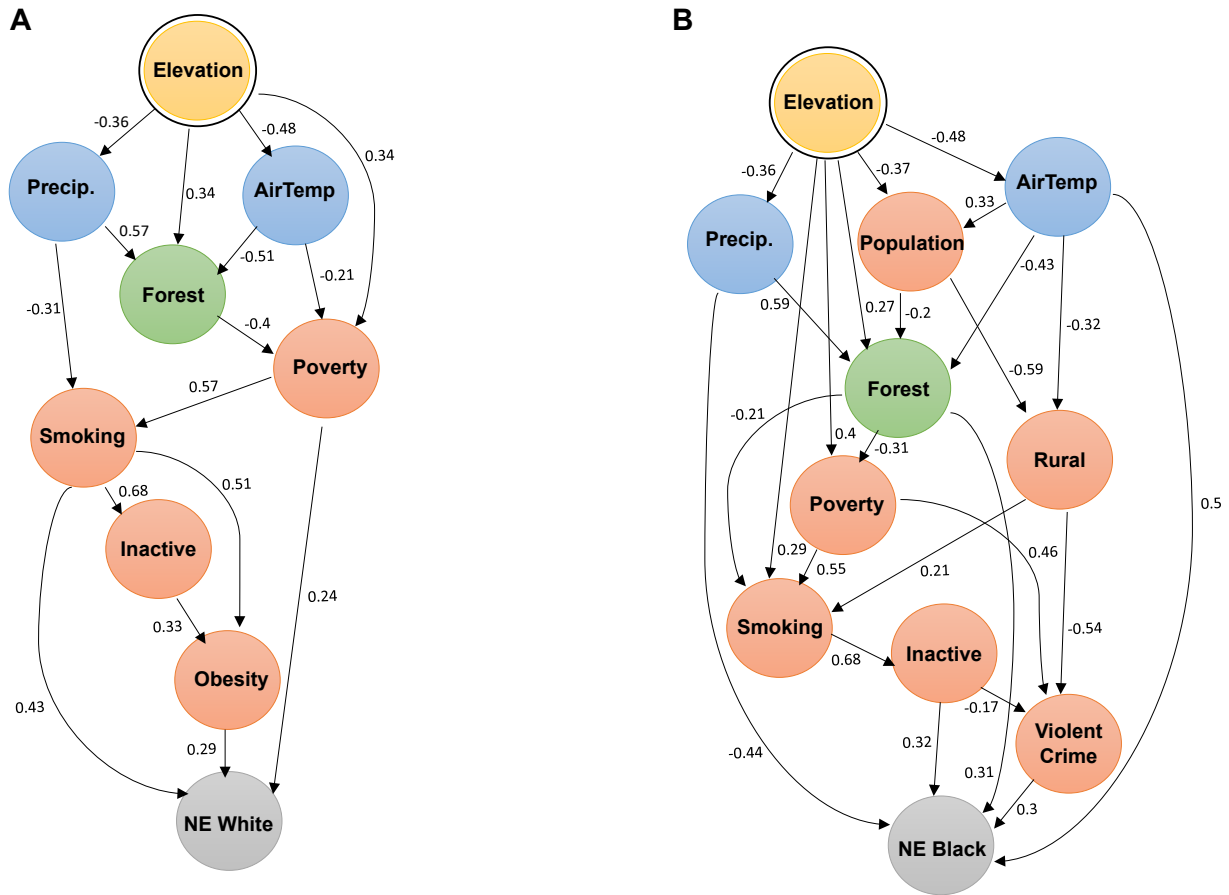


Fig. 4. Path model diagrams of age-adjusted mortality rates (AAMR) for (A) White and (B) Black residents in Northeast counties. The four classes of predictor variables are indicated by color: yellow = geologic; blue = climatic; green = land cover; pink = socioeconomic. Variables shown in grey are indices of human health (AAMR). Exogenous variables are shown with an outer ring; all remaining variables are endogenous. Standardized path coefficients are listed along each path and all paths shown in the diagram are significant at $p \leq 0.05$. Complete path analysis results (model fit statistics and coefficients of determination) are listed in Table 2.

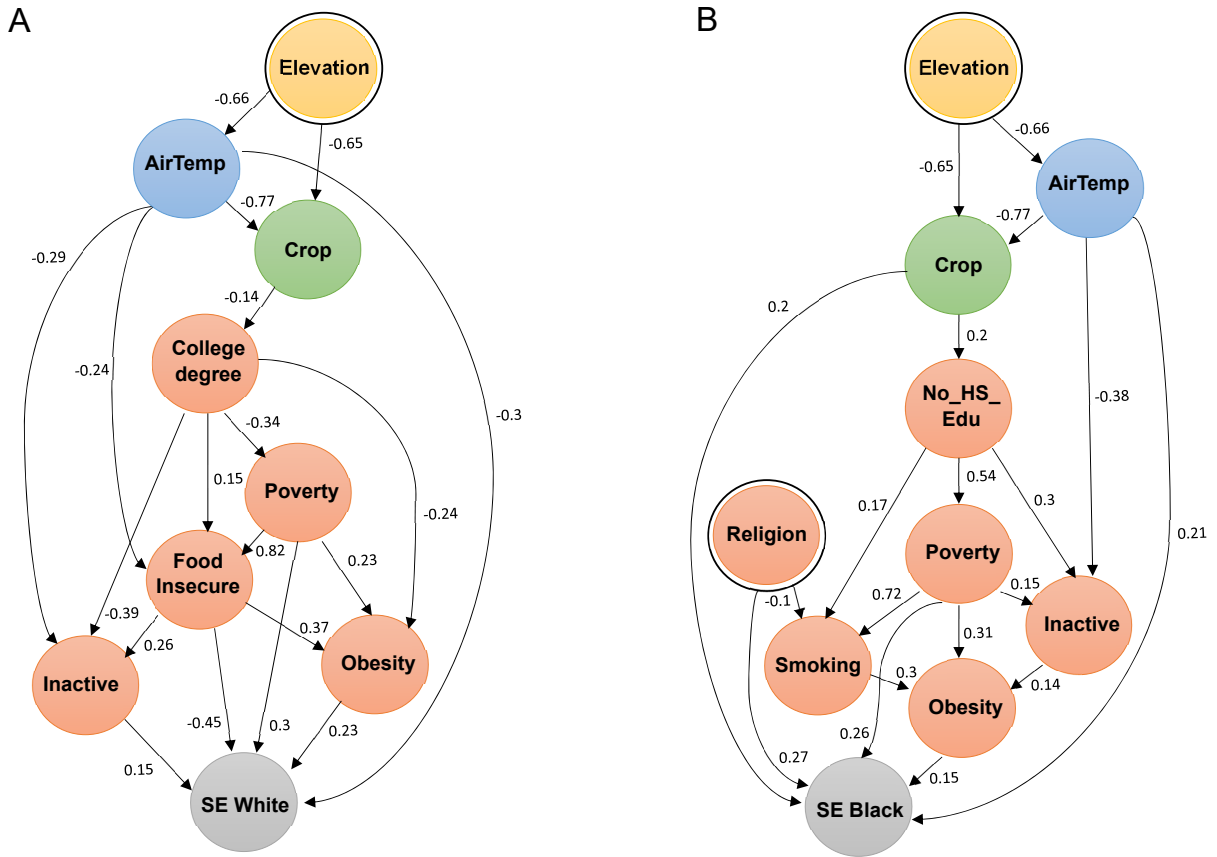


Fig. 5. Path model diagrams of age-adjusted mortality rates (AAMR) for (A) White and (B) Black residents in Southeast counties. The four classes of predictor variables are indicated by color: yellow = geologic; blue = climatic; green = land cover; pink = socioeconomic. Variables shown in grey are indices of human health (AAMR). Exogenous variables are shown with an outer ring; all remaining variables are endogenous. Standardized path coefficients are listed along each path and all paths shown in the diagram are significant at $p \leq 0.05$. Complete path analysis results (model fit statistics and coefficients of determination) are listed in Table 2.

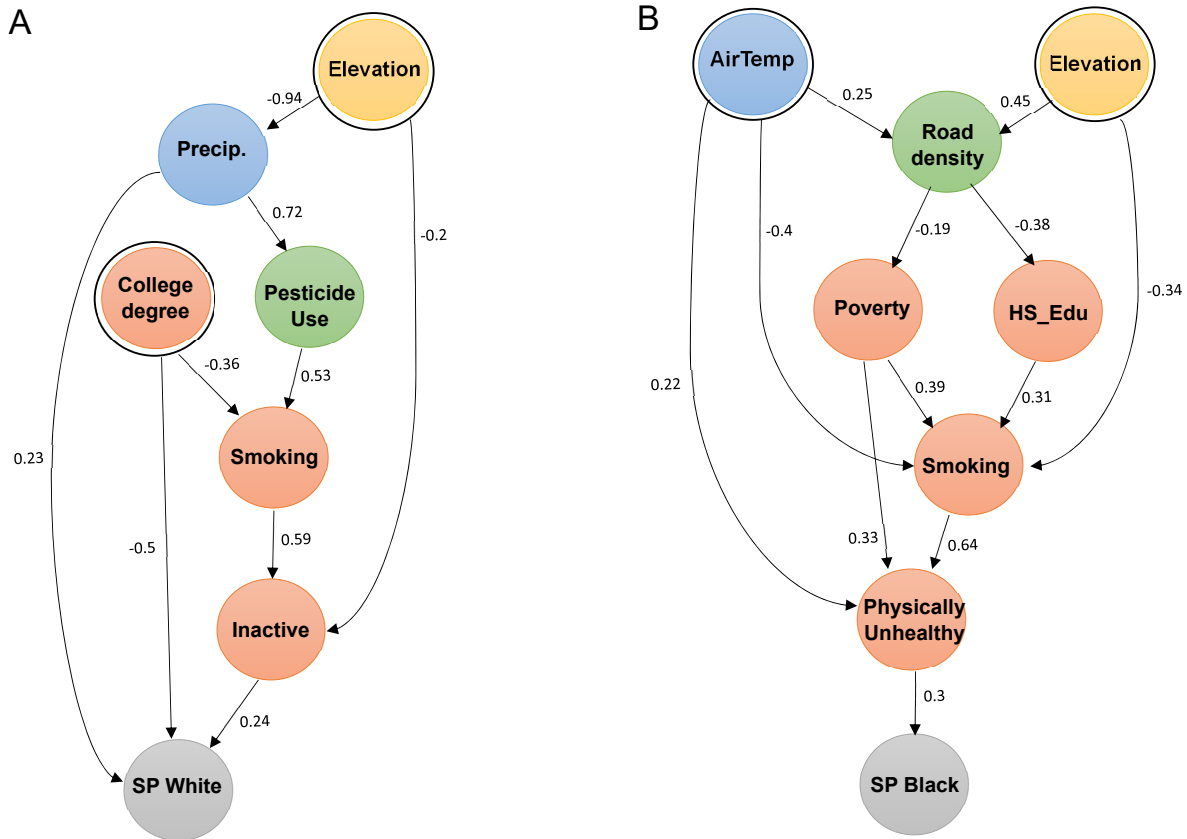


Fig. 6. Path model diagrams of age-adjusted mortality rates (AAMR) for (A) White and (B) Black residents in Southern Plains counties. The four classes of predictor variables are indicated by color: yellow = geologic; blue = climatic; green = land cover; pink = socioeconomic. Variables shown in grey are indices of human health (AAMR). Exogenous variables are shown with an outer ring; all remaining variables are endogenous. Standardized path coefficients are listed along each path and all paths shown in the diagram are significant at $p \leq 0.05$. Complete path analysis results (model fit statistics and coefficients of determination) are listed in Table 2.

B-AAMR, while pesticide use mediated indirect links between precipitation and W-AAMR (Fig. 6).

Discussion

This study examined regional and racial differences in environmental and socioeconomic factors that influence AAMR. In general, the six region \times race PA sub-models did not predict AAMR as well as the CONUS-scale model of Walls and McGarvey (2023b). For instance, while indicators of overall path model fit were generally good (see Table 2), five of the six sub-models had AAMR R^2 values < 0.5 . Only the Northeast W-AAMR model had a higher R^2 (0.72) than the Walls and McGarvey (2023b) model (AAMR $R^2 = 0.63$). Lower R^2 values may be due to the relatively simple structures of the sub-models; sample-size limitations ($n = 124\text{--}233$) prevented us from including as many environmental and socioeconomic covariates in the sub-models as the CONUS-scale model ($n > 450$; Walls & McGarvey, 2023b). However, lower R^2 values may also indicate that our county-level environmental and socioeconomic data may be less well-suited to study local-scale health indicators than finer-scale information. For instance, Wheeler et al. (2015) used 1-km resolution environmental data to model drinking water contamination throughout the state of Iowa. We therefore predict that revisiting the pathways detected in the region \times race sub-models with finer-scale data will reveal new insight to the environmental and socioeconomic drivers of B-AAMR and W-AAMR.

Despite their modest R^2 values, the sub-models did indicate the metamodel structure (Fig. 2), which mirrored the underlying structure of the CONUS-scale model (see Fig. 1 in Walls & McGarvey, 2023b), remains a useful tool for studying environmental and socioeconomic effects on human health. Geology and landscape morphology variables served as master, exogenous

influences on the climate, hydrology, and land cover variables that regulated socioeconomic conditions (Figs. 4-6). This consistency in general metamodel structure suggests that our systems-level approach to the science of linked human-environmental health is applicable among spatial scales, regions, and racial demographics.

Similar to the national AAMR model of Walls and McGarvey (2023b), two socioeconomic variables, smoking and physical inactivity, had consistently strong influences on AAMR (Fig. 3). Notably, these socioeconomic effects were pervasive among all regions and for both B-AAMR and W-AAMR (Figs. 4-6). Strong effects of smoking and physical inactivity were not surprising, as they have previously been linked to cancer, diabetes, and a variety of cardiovascular and respiratory diseases (Burns, 2003; Cecchini et al., 2010; Paffenbarger et al., 1993).

Other key predictors of AAMR included covariates reflecting education (attainment of high school diploma or a college degree) and income (poverty, lack of health insurance, and food insecurity). Links between education, income, and human health have previously been attributed to the mediating effects of social-psychological well-being and behavior (Ross & Wu, 1995). Level of education is a strong predictor of income, while income is directly associated with health insurance and access to preventative medical care (Hargraves, 2004; Raghupathi & Raghupathi, 2020). Similarly, education level is positively associated with psychological well-being, access to robust social support networks, and intentional self-care behaviors, such as avoidance of smoking and the pursuit of active lifestyles (Raghupathi & Raghupathi, 2020; Ross & Wu, 1995; Zajacova & Lawrence, 2018).

One key difference among the strongest predictors of B-AAMR and W-AAMR was the importance of social cohesion. While the top predictors of W-AAMR (e.g., college degree, smoking and physical inactivity; Fig. 3) were similar to the leading predictors of overall AAMR

(all races combined) at the CONUS-scale (see Walls & McGarvey, 2023a, 2023b), B-AAMR was particularly responsive to social cohesion variables in the Southeast and Southern Plains. Specifically, the percent of adults who actively participate in an organized religion (Religion) was the third strongest predictor of B-AAMR in the Southeast, while the percent of adults with formal links to community organizations (Social_Association_Rate) was the top predictor of B-AAMR in the Southern Plains (Fig. 3).

The potential health benefits of social cohesion have been explored in research on religious activity. Higher, more consistent levels of participation in religious functions are associated with robust social support structures and marital stability, both of which facilitate measurable reductions in stress (Hummer et al., 1999; Idler et al., 2017; Sissamis et al., 2022). Conversely, measurable harm may result when individuals with strong spiritual roots become estranged from organized religion (Abu-Raiya et al., 2016); effects may include depression, increased suicide risk, and increased mortality rates (Exline et al., 2000; Pargament et al., 2001). This suggests that additional research on social cohesion may be valuable in efforts to improve health outcomes for Black or other minoritized communities.

Both the RF (Fig. 3) and PA results (Figs. 4-6) indicate that climate, represented by air temperature and precipitation, are fundamental determinants of AAMR. Direct effects of climate may be realized through extreme precipitation or temperature events. For instance, flooding may increase the risk of drowning (Haines et al., 2006; Zhang et al., 2021) or exposure to untreated sewage overflows (Jagai et al., 2015; Wade et al., 2014), while heat waves lead directly to heat stroke (Hoshiko et al., 2010; Kovats & Hajat, 2008). Exposure to heat waves may be particularly problematic for Black residents in the Southeast, where air temperature was positively correlated with B-AAMR (Fig. 5B), but negatively correlated with W-AAMR (Fig. 5A). This difference may

reflect inequal access to efficient, temperature-controlled housing (Kontokosta et al., 2020; Kovats & Hajat, 2008; Romanello et al., 2021) or a disproportionate percentage of Black individuals who are vulnerable to heat stress, due to pre-existing cardiovascular or chronic respiratory diseases (Basu, 2009; Jesdale et al., 2013; McMichael et al., 2006).

Despite the potential harm caused by direct climate-health links, we believe the most pervasive effects of climate on human health may be indirect. As predicted by the metamodel (Fig. 2), the PA results confirmed air temperature and precipitation have significant, indirect effects AAMR that are mediated by land cover and socioeconomic factors. For example, in the Northeast, temperature and precipitation are essential predictors of forest cover (Fig. 4). Forest cover, in turn, predicts poverty, while poverty is linked to multiple socioeconomic variables (e.g., smoking and violent crime) that directly influence both B-AAMR and W-AAMR. This sequence of influences makes sense in a historical context. Climate and local geology promoted growth of the large, dense timber stands that once sustained a robust logging industry in the Northeast (Thompson et al., 2013). However, the subsequent decline of industrial logging and the rise of urban and agricultural land conversion created an inverse relationship between contemporary forest cover and poverty, as exhibited in the Northeast PA models (Fig. 4).

The observed negative link between air temperature and crop cover in the Southeast (Fig. 5) is likely a reflection of heat stress. Increasing prevalence of extreme heat events is a major concern for the agricultural industry in the Southeast, where the most productive plots often occur in relatively cool areas (Asseng, 2013). Local crop production and its associated revenues are in turn linked to wages and unemployment rates (Piontak & Schulman, 2014), as well as high prevalence of food insecurity and malnutrition in the Southeast (McMichael et al., 2006).

This work is an early step in the effort to understand how the health of Black and White communities may respond to environmental and/or socioeconomic changes. The six PA models lay the groundwork for identifying distinct multi-step pathways between environmental and socioeconomic variables that play a key role in driving regional and racial mortality trends within the CONUS. Our results highlight three major findings. First, education and income are consistently the strongest predictors of B-AAMR and W-AAMR within all three subregions (Northeast, Southeast, and Southern Plains). Second, mortality trends among subregions are driven by differences in environmental influences on socioeconomic predictors. For instance, forest cover had a significant influence on income in the Northeast. But crop cover had a direct influence on education in the Southeast subregion. Third, climate variables, air temperature and precipitation, have the most widespread influence on linked human-environmental health systems among regions and demographic subgroups.

The relatively low R^2 values observed for the PA models, combined with the modest number of variables used (relative to the CONUS model of Walls & McGarvey, 2023b) to represent the overall interconnections between the environment, socioeconomic factors and human health indicate that we did not capture all relevant factors in this study. There are likely finer scale, predictor of mortality trends that were not considered in this analysis. For instance, links between increased risk of vector-borne diseases and differing socioeconomic and climate change conditions have been well documented in previous studies (Guernier et al., 2004; Springer & Johnson, 2018), but were not evaluated in this study. We therefore encourage future studies to incorporate new, finer-scales variables to the dataset. We predict these additions to future macrosystem ecology studies would be useful for addressing uncertainties surrounding the widespread and varied consequences of global environmental change.

References

- Abu-Raiya, H., Pargament, K. I., & Krause, N. (2016). Religion as problem, religion as solution: religious buffers of the links between religious/spiritual struggles and well-being/mental health. *Quality of Life Research, 25*, 1265–1274.
- Angermeier, P., Krometis, L., Stern, M., & Hemby, T. (2021). Exploring relationships among stream health, human well-being, and demographics in Virginia, USA. *Ecological Indicators, 121*, 107194.
- Asseng, S. (2013). *Agriculture and climate change in the Southeast USA, In climate of the Southeast United States: variability, change, impacts, and vulnerability.* (K. T. Ingram, K. Dow, L. Carter, & J. Anderson, Eds.). Island Press/Center for Resource Economics.
- Basu, R. (2009). High ambient temperature and mortality : a review of epidemiologic studies from 2001 to 2008. *Environmental Health, 8*(40), 1–13.
- Benamins, M. R., Silva, A., Saiyed, N. S., & De Maio, F. G. (2021). Comparison of all-cause mortality rates and inequities between Black and White populations across the 30 most populous US cities. *JAMA Network Open, 4*(1).
- Borak, J., Salipante-Zaidel, C., Slade, M. D., & Fields, C. A. (2012). Mortality disparities in Appalachia: Reassessment of major risk factors. *Journal of Occupational and Environmental Medicine, 54*(2), 146–156.
- Bullard, R. D. (1993). Race and environmental justice in the United States. *Yale Journal of International Law, 18*(1), 319–336.
- Burns, D. M. (2003). Epidemiology of smoking-induced cardiovascular disease. *Progress in Cardiovascular Diseases, 46*(1), 11–29.

- Burnside, W., Brown, J., Burger, O., Hamilton, M., Moses, M., & Bettencourt, L. (2012). Human macroecology : linking pattern and process in big-picture human ecology. *Biological Reviews*, 87(2012), 194–208.
- Cecchini, M., Sassi, F., Lauer, J. A., Lee, Y. Y., Guajardo-Barron, V., & Chisholm, D. (2010). Tackling of unhealthy diets, physical inactivity, and obesity: Health effects and cost-effectiveness. *The Lancet*, 376(9754), 1775–1784.
- Cunningham, T. J., Croft, J. B., Liu, Y., Lu, H., Eke, P. I., & Giles, W. H. (2017). Vital signs : racial disparities in age-specific mortality among Blacks or African Americans — United States , 1999–2015. *Morbidity and Mortality Weekly Report*, 66(17), 444–456.
- Davids, B.-O., Hutchins, S. S., Jones, C. P., & Hood, J. R. (2014). Disparities in life expectancy across US counties linked to county social factors, 2009 Community Health Status Indicators (CHSI). *Journal of Racial and Ethnic Health Disparities*, 1(1), 2–11.
- Dunham, J., Angermeier, P., Crausbay, S., Cravens, A., Gosnell, H., Mcevoy, J., Moritz, M., Raheem, N., & Sanford, T. (2018). Rivers are social-ecological systems: Time to integrate human dimensions in riverscape ecology and management. *WIREs Water*, 1–10.
- Dwyer-Lindgren, L., Bertozzi-Villa, A., Stubbs, R., Morozoff, C., Mackenbach, J., Van Lenthe, F., Mokdad, A., & Murray, C. (2017). Inequalities in life expectancy among US counties, 1980 to 2014: Temporal trends and key drivers. *Journal of the American Medical Association Internal Medicine*, 177(7), 1003–1011.
- Ebi, K., Balbus, J., Luber, G., Bole, A., Crimmins, A., Glass, S., Saha, S., Shimamoto, J., Trtanj, J., & White-Newsome, J. (2018). Human health. In impacts, risks, and adaptation in the United States: Fourth national climate assessment. *U.S. Global Change Research Program*, II, 572–603.

- Engemann, K., Pedersen, C. B., Arge, L., Tsirogiannis, C., Mortensen, P. B., & Svenning, J. C. (2019). Residential green space in childhood is associated with lower risk of psychiatric disorders from adolescence into adulthood. *Proceedings of the National Academy of Sciences*, *116*, 5188–5193.
- Exline, J. J., Yali, A. M., & Sanderson, W. C. (2000). Guilt, discord, and alienation: The role of religious strain in depression and suicidality. *Journal of Clinical Psychology*, *56*(12), 1481–1496.
- Ezzati, M., Friedman, A. B., Kulkarni, S. C., & Murray, C. J. L. (2008). *The reversal of fortunes: Trends in county mortality and cross-county mortality disparities in the United States*.
- Fan, Y., Chen, J., Shirkey, G., John, R., Wu, S. R., Park, H., & Shao, C. (2016). Applications of structural equation modeling (SEM) in ecological studies: an updated review. *Ecological Processes*, *5*(1).
- Fleishman, E., & Brown, H. (2019). Use of macroecology to integrate social justice and conservation. *Global Ecology and Biogeography*, *00*, 1–7.
- Graff, M., Carley, S., Konisky, D. M., & Memmott, T. (2021). Which households are energy insecure? An empirical analysis of race, housing conditions, and energy burdens in the United States. *Energy Research and Social Science*, *79*.
- Grammich, C., Hadaway, K., Houseal, R., Jones, D., Krindatch, A., Stanley, R., & Taylor, R. (2012). *2010 U.S. religion census: religious congregations and membership study*. Association of Statisticians of American Religious Bodies.
- Guernier, V., Hochberg, M. E., & Guegan, J.-F. (2004). Ecology drives the worldwide distribution of human diseases. *PLoS Biology*, *2*(6), 740–746.

- Haines, A., Kovats, R. S., Campbell-Lendrum, D., & Corvalan, C. (2006). Climate change and human health: Impacts, vulnerability and public health. *Public Health, 120*, 585–596.
- Hargraves, J. L. (2004). *Trends in health insurance coverage and access among black, Latino and white Americans, 2201-2003*. Center for Studying Health System Change.
- Harper, S., MacLehose, R. F., & Kaufman, J. S. (2014). Trends in the black-white life expectancy gap among US States, 1990-2009. *Health Affairs, 33*(8), 1375–1382.
- Heffernan, J. B., Soranno, P. A., Angilletta, M. J., Buckley, L. B., Gruner, D. S., Keitt, T. H., Kellner, J. R., Kominoski, J. S., Rocha, A. V., Xiao, J., Harms, T. K., Goring, S. J., Koenig, L. E., McDowell, W. H., Powell, H., Richardson, A. D., Stow, C. A., Vargas, R., & Weathers, K. C. (2014). Macrosystems ecology: Understanding ecological patterns and processes at continental scales. *Frontiers in Ecology and the Environment, 12*(1), 5–14.
- Herlihy, A. T., Paulsen, S. G., Van Sickle, J., Stoddard, J. L., Hawkins, C. P., & Yuan, L. L. (2008). Striving for consistency in a national assessment: the challenges of applying a reference-condition approach at a continental scale. *Journal of The North American Benthological Society, 27*(4), 860–877.
- Hill, R., Fox, E., Leibowitz, S., Olsen, A., Thornbrugh, D., & Weber, M. (2017). Predictive mapping of the biotic condition of conterminous U.S. rivers and streams. *Ecological Applications, 27*(8), 2397–2415.
- Hill, R., Weber, M., Leibowitz, S., Olsen, A., & Thornbrugh, D. (2016). The Stream-Catchment (StreamCat) Dataset: A Database of watershed metrics for the Conterminous United States. *Journal of the American Water Resources Association, 52*(1).

- Hoshiko, S., English, P., Smith, D., & Trent, R. (2010). A simple method for estimating excess mortality due to heat waves, as applied to the 2006 California heat wave. *International Journal of Public Health, 55*(2), 133–137.
- Hummer, R. A., Rogers, R. G., Nam, C. B., & Ellison, C. G. (1999). Religious involvement and U.S. adult mortality. *Demography, 36*(2), 273–285.
- Idler, E., Blevins, J., Kiser, M., & Hogue, C. (2017). Religion, a social determinant of mortality? A 10-year follow-up of the Health and Retirement Study. *PLoS ONE, 12*(12), 1–15.
- Jagai, J., Quanlin, L., Wang, S., Messier, K., Wade, T., & Hilborn, E. (2015). Extreme precipitation and emergency room visits for gastrointestinal illness in areas with and without combined sewer systems : An analysis of Massachusetts data, 2003-2007. *Environmental Health Perspectives, 123*(9), 873–879.
- Jesdale, B. M., Morello-frosch, R., & Cushing, L. (2013). Ethnic racial/ethnic distribution of heat risk – related land cover in relation to residential segregation. *Environmental Health Perspectives, 121*(07), 811–817.
- Kaufmann, P. R., Hughes, R. M., Paulsen, S. G., Peck, D. V., Seeliger, C. W., Weber, M. H., & Mitchell, R. M. (2022). Physical habitat in conterminous US streams and rivers, Part 1: Geoclimatic controls and anthropogenic alteration. *Ecological Indicators, 141*, 109046.
- Kontokosta, C. E., Reina, V. J., & Bonczak, B. (2020). Energy Cost Burdens for Low-Income and Minority Households: Evidence From Energy Benchmarking and Audit Data in Five U.S. Cities. *Journal of the American Planning Association, 86*(1), 89–105.
- Kovats, R. S., & Hajat, S. (2008). Heat stress and public health: A critical review. *Annual Review of Public Health, 29*, 41–55.

- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3), 18–22.
- Mattei, D., & Ruscio, J. (2022). Automated transformations, normality testing, and reporting. R package ‘Rita’ version 1.2.0. In *Testing For Normality*. CRC Press.
- Mckay, L., Bondelid, T., Dewald, T., Johnston, C., Moore, R., & Rea, A. (2012). NHDPlus Version 2 : User guide (data model version 2.1). *Horizon Systems*.
- McMichael, A. J., Woodruff, R. E., & Hales, S. (2006). Climate change and human health: present and future risks. *The Lancet*, 367, 859–869.
- Omernik, J. M. (1987). Ecoregions of the conterminous United States. *Annals of the Association of American Geographers*, 77(1), 118–125.
- Paffenbarger, R., Hyde, R., Alvin, W., Lee, I.-M., Jung, D., & Kampert, J. (1993). The association of changes in physical-activity level and other lifestyle characteristics with mortality among men. *New England Journal of Medicine*, 328(8), 538–545.
- Pargament, K. I., Koenig, H. G., Tarakeshwar, N., & Hahn, J. (2001). Religious Struggle as a predictor of mortality among medically ill elderly patients: A 2-year longitudinal study. *Archives of Internal Medicine*, 161(125), 1881–1885.
- Partridge, M. D., Betz, M. R., & Lobao, L. (2013). Natural resource curse and poverty in appalachian America. *American Journal of Agricultural Economics*, 95(2), 449–456.
- Piontak, J. R., & Schulman, M. D. (2014). Food insecurity in rural America. *Contexts*, 13(3), 75–77.
- Raghupathi, V., & Raghupathi, W. (2020). The influence of education on health: an emperical assessment of OECD countries for the period 1995-2015. *Archives of Public Health*, 78(20), 1–18.

- Remington, P. L., Catlin, B. B., & Gennuso, K. P. (2015). The County Health Rankings: rationale and methods. *Population Health Metrics*, 13, 11.
- Romanello, M., McGushin, A., Di Napoli, C., Drummond, P., Hughes, N., Jamart, L., Kennard, H., Lampard, P. M., Solano Rodriguez, B., Arnell, N., Ayeb-Karlsson, S., Belesova, K., Cai, W., Campbell-Lendrum, D., Capstick, S., Chambers, J., Chu, L., Ciampi, L., Dalin, C., ... Hamilton, I. (2021). The 2021 report of the Lancet Countdown on health and climate change: code red for a healthy future. *The Lancet*, 398(10311), 1619–1662.
- Ross, C. E., & Wu, C. (1995). The links between education and health. *American Sociological Review*, 60, 719–7454.
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more Version 0.5-12 (BETA). *Journal of Statistical Software*, 48.
- Russette, H., Graham, J., Holden, Z., Semmens, E., Williams, E., & Landguth, E. (2021). Greenspace exposure and COVID-19 mortality in the United States: January-July 2020. *Environmental Research*, 198(111195).
- Sissamis, F., Villalba, K., Garcia, J., Melus, V., Markentell, E. J., Perez, L. D., & Ramirez, G. (2022). Religion and death in the United States: A meta-regression comparative assessment of between-county mortality heterogeneity in the United States. *International Journal of Environmental Research and Public Health*, 19(2), 1–16.
- Springer, Y. R., & Johnson, P. T. J. (2018). Large-scale health disparities associated with Lyme disease and human monocytic ehrlichiosis in the United States , 2007–2013. *PLoS ONE*, 13(9), e0204609.
- Stekhoven, D., & Bühlmann, P. (2012). Missforest-non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1), 112–118.

- Sudano, J. J., & Baker, D. W. (2006). Explaining US racial/ethnic disparities in health declines and mortality in late middle age: The roles of socioeconomic status, health behaviors, and health insurance. *Social Science and Medicine*, 62(4), 909–922.
<https://doi.org/10.1016/j.socscimed.2005.06.041>
- Thompson, J. R., Carpenter, D. N., Cogbill, C. V., & Foster, D. R. (2013). Four centuries of change in northeastern United States forests. *PLoS ONE*, 8(9).
- Ulrich, W., Batary, P., Baudry, J., Beaumelle, L., Bucher, R., Cerevkova, A., de la Riva, E. G., Felipe-Lucia, M. R., Galle, R., Kesse-Guyot, E., Rembialkowska, E., Rusch, A., Stanley, D., & Birkhofer, K. (2022). From biodiversity to health: Quantifying the impact of diverse ecosystems on human well-being. *People and Nature*, 5, 69–83.
- Wade, T., Lin, C., Jagai, J., & Hilborn, E. (2014). Flooding and emergency room visits for gastrointestinal illness in Massachusetts : A case-crossover study. *PloS ONE*, 9(10).
- Walls, F. N., & McGarvey, D. J. (2023a). A systems-level model of direct and indirect links between environmental health, socioeconomic factors, and human mortality. *Science of the Total Environment*, 874, 162486.
- Walls, F. N., & McGarvey, D. J. (2023b). Building a macrosystems ecology framework to identify links between environmental and human health: A random forest modelling approach. *People and Nature*, 5(1), 183–197.
- Wheeler, D., Nolan, B., Flory, A., DellaValle, C., & Ward, M. (2015). Modeling groundwater nitrate concentrations in private wells in Iowa. *Science of the Total Environment*, 536, 481–488.

Zajacova, A., & Lawrence, E. M. (2018). The relationship between education and health: reducing disparities through a contextual approach. *Annual Review of Public Health, 39*, 273–289.

Zhang, J., Xu, W., Liao, X., Zong, S., & Liu, B. (2021). Global mortality risk assessment from river flooding under climate change. *Environmental Research Letters, 16*(6).

Table S1. List of the 85 covariates and 2 response variables included in the path analysis models. Covariate descriptions, variable classes, units of measure, and original data source citations (see references in main text) are shown.

Variable	Description	Class	Units	Source
ArbolateSum_min	Total length of all upstream segments (minimum)	geology/morphology	km	Mckay et al. (2012)
AreaSqKM_median	Area of watershed (median)	geology/morphology	km2	Mckay et al. (2012)
ElevCat_median	Elevation (median within catchment)	geology/morphology	m	Mckay et al. (2012)
SLOPE_max	Slope (maximum)	geology/morphology	m/m	Mckay et al. (2012)
StreamLevel_min	Stream level (minimum)	geology/morphology	unitless	Mckay et al. (2012)
StreamOrder_max	Stream order (maximum)	geology/morphology	unitless	Mckay et al. (2012)
TotDASqKM_median	Total drainage area (median)	geology/morphology	km2	Mckay et al. (2012)
AgKffactWs_max	Soil erodibility factor (maximum within watershed)	geology/morphology	unitless	Hill et al. (2016)
CaOWs_max	Lithologic calcium oxide content (maximum within watershed)	geology/morphology	% by volume	Hill et al. (2016)
ClayCat_median	Soil clay content (median within catchment)	geology/morphology	% by mass	Hill et al. (2016)
CoalMineDensCat_max	Coal mine density (maximum within catchment)	geology/morphology	no./km2	Hill et al. (2016)
CoalMineDensWs_median	Coal mine density (median within watershed)	geology/morphology	no./km2	Hill et al. (2016)
CompStrgthCat_max	Lithologic compressive strength (maximum within catchment)	geology/morphology	mega Pascals	Hill et al. (2016)
Fe2O3Ws_median	Lithologic ferric oxide content (median within watershed)	geology/morphology	% by volume	Hill et al. (2016)
HydrCondWs_median	Lithologic hydraulic conductivity (median within watershed)	geology/morphology	?m/sec	Hill et al. (2016)
MgOWs_median	Lithologic magnesium oxide content (median within watershed)	geology/morphology	% by volume	Hill et al. (2016)
NWS_median	Lithologic nitrogen content (median within watershed)	geology/morphology	% by mass	Hill et al. (2016)
OmWs_max	Soil organic matter content (maximum within watershed)	geology/morphology	% by mass	Hill et al. (2016)
P2O5Ws_median	Lithologic phosphorus oxide content (median within watershed)	geology/morphology	% by volume	Hill et al. (2016)
PctAlluvCoastWs_max	Lithologic alluvium and fine texture coastal zone sediment (maximum within watershed)	geology/morphology	% by volume	Hill et al. (2016)
PctCarbResidWs_median	Lithologic carbonate residual content (median within watershed)	geology/morphology	% by volume	Hill et al. (2016)
PctColluvSedWs_median	Lithologic colluvial sediment content (median within watershed)	geology/morphology	%	Hill et al. (2016)
PctEolCrsWs_median	Lithologic fine-textured sediment (median within watershed)	geology/morphology	% by volume	Hill et al. (2016)
PctEolFineWs_max	Lithologic fine-textured sediment (maximum within watershed)	geology/morphology	% by volume	Hill et al. (2016)
PctGlaclakeCrsWs_max	Lithologic glacial till coarse-textured content (maximum within watershed)	geology/morphology	% by volume	Hill et al. (2016)
PctGlactilloamWs_max	Lithologic glacial till loamy content (maximum within watershed)	geology/morphology	% by volume	Hill et al. (2016)
PctHydricWs_median	Lithologic peat and muck content (median within watershed)	geology/morphology	% by volume	Hill et al. (2016)
PctNonCarbResidWs_median	Lithologic non-carbonate residual content (median within watershed)	geology/morphology	% by volume	Hill et al. (2016)
PermCat_max	Soil permeability (maximum within catchment)	geology/morphology	cm/hr	Hill et al. (2016)
RckDepWs_min	Depth to bedrock (minimum within watershed)	geology/morphology	cm	Hill et al. (2016)
SandWs_median	Sand content of soils (median within watershed)	geology/morphology	% by mass	Hill et al. (2016)
SiO2Ws_min	Lithologic silicon dioxide content (minimum within watershed)	geology/morphology	% by volume	Hill et al. (2016)
PrecipCat_median	Annual precipitation (median within catchment)	climate	mm/yr	Hill et al. (2016)
TmeanCat_median	Mean annual air temperature (median within catchment)	climate	degrees C	Hill et al. (2016)
BFIWs_max	Baseflow index (groundwater influence) (maximum within watershed)	hydrology	% of total flow	Hill et al. (2016)
RunoffWs_max	Runoff (maximum within watershed)	hydrology	mm/yr	Hill et al. (2016)
WetIndexCat_max	Soil wetness index (maximum within catchment)	hydrology	unitless	Hill et al. (2016)
WtDepWs_median	Water table depth of soils (median within watershed)	hydrology	cm	Hill et al. (2016)
AllForest_median	Combined deciduous + evergreen + mixed forest land cover (median within watershed)	land cover/land use	% by area	Hill et al. (2016)
InorgNWetDep_2008Ws_median	Inorganic nitrogen wet deposition (median within watershed)	land cover/land use	kg of N/ha/yr	Hill et al. (2016)
IWI_v2_1_median	Index of Watershed Integrity version 2 (median within watershed)	land cover/land use	unitless	Hill et al. (2016)
ManureWs_median	Nitrogen manure application (median within watershed)	land cover/land use	kg/ha/yr	Hill et al. (2016)
NPDESdensCat_max	National Pollutant Discharge Elimination System sites (maximum within catchment)	land cover/land use	no./km2	Hill et al. (2016)
NPDESdensWs_median	National Pollutant Discharge Elimination System sites (median within watershed)	land cover/land use	no./km2	Hill et al. (2016)
PctAg2006Slp20Cat_max	Agricultural land cover (maximum within catchment)	land cover/land use	% by area	Hill et al. (2016)
PctBl2011Ws_median	Barren land cover (median within watershed)	land cover/land use	% by area	Hill et al. (2016)
PctConif2011Cat_max	Evergreen forest land cover (maximum within catchment)	land cover/land use	% by area	Hill et al. (2016)
PctCrop2011Ws_median	Crop land use (median within watershed)	land cover/land use	% by area	Hill et al. (2016)
PctDecid2011Ws_max	Deciduous forest land cover (maximum within watershed)	land cover/land use	% by area	Hill et al. (2016)
PctGr2011Ws_max	Grassland herbaceous cover (maximum within watershed)	land cover/land use	% by area	Hill et al. (2016)
PctHay2011Ws_median	Hay land use cover (median within watershed)	land cover/land use	% by area	Hill et al. (2016)
PctHbWet2011Ws_median	Herbaceous wetland cover (median within watershed)	land cover/land use	% by area	Hill et al. (2016)
PctIce2011Ws_median	Ice/snow land cover (median within watershed)	land cover/land use	% by area	Hill et al. (2016)
PctImp2011Ws_median	Impervious surface cover (median within watershed)	land cover/land use	% by area	Hill et al. (2016)
PctMxFst2011Ws_median	Mixed deciduous/evergreen forest (median within watershed)	land cover/land use	% by area	Hill et al. (2016)
PctNonAgIntrodManagVegWs_median	Managed nonnative vegetation cover (median within watershed)	land cover/land use	% by area	Hill et al. (2016)
PctShrb2011Ws_median	Shrub/scrub land cover (median within watershed)	land cover/land use	% by area	Hill et al. (2016)

Table S1. Continued .

PctWdWet2011Cat_median	Woody wetland cover (median within catchment)	land cover/land use	% by area	Hill et al. (2016)
Pestic97Ws_max	Pesticide use (maximum within watershed)	land cover/land use	kg/km2/yr	Hill et al. (2016)
Pestic97Ws_median	Pesticide application (median within watershed)	land cover/land use	kg/km2	Hill et al. (2016)
RdDensCat_median	Road density (median within catchment)	land cover/land use	km/km2	Hill et al. (2016)
SuperfundDensCat_max	Superfund sites (maximum within catchment)	land cover/land use	no./km2	Hill et al. (2016)
SuperfundDensWs_min	Superfund sites (minimum within watershed)	land cover/land use	no./km2	Hill et al. (2016)
TRIDensCat_max	Toxic Release Inventory sites (maximum within catchment)	land cover/land use	no./km2	Hill et al. (2016)
TRIDensWs_median	Toxic Release Inventory sites (median within watershed)	land cover/land use	no./km2	Hill et al. (2016)
PopDen2010Ws_max	Human population density (maximum within watershed)	socioeconomic	no./km2	Hill et al. (2016)
LAI	Leaf area index (surface area of green vegetation per unit ground surface area; mean within county)	socioeconomic	unitless	Russette et al. (2021)
Religion	Individuals who adhere to any organized religion, as percent of adult population	socioeconomic	% by population	Grammich et al. (2012)
No_HS_Edu	Individuals with a less than a high school diploma, as percent of adult population	socioeconomic	% by population	American Community Survey (US Census Bureau)
HS_Edu	Individuals with a high school diploma only, as percent of adult population	socioeconomic	% by population	American Community Survey (US Census Bureau)
Some_College	Individuals with some college or associates degree, as percent of adult population	socioeconomic	% by population	American Community Survey (US Census Bureau)
College_degree	Individuals with a bachelor's degree or higher, as percent of adult population	socioeconomic	% by population	American Community Survey (US Census Bureau)
IncomeBelowPov_Percent	Individuals whose income in the last 12 months is below the poverty level, as percent of adult population	socioeconomic	% by population	American Community Survey (US Census Bureau)
No_HI_cov_Percent	Individuals lacking health insurance, as percent of adult population	socioeconomic	% by population	American Community Survey (US Census Bureau)
Renter_Occ_Percent	Individuals renting their main residence, as percent of adult population	socioeconomic	% by population	American Community Survey (US Census Bureau)
Pct_Smokers	Active smokers, as percent of adult population	socioeconomic	% by population	Remington et al. (2015)
Pct_Physically_Inactive	Inactive individuals, as percent of adult population	socioeconomic	% by population	Remington et al. (2015)
Physically_Unhealthy_Days	Poor physical health days, as average number reported per month	socioeconomic	avg. no./month	Remington et al. (2015)
Mentally_Unhealthy_Days	Poor mental health days, as average number reported per month	socioeconomic	avg. no./month	Remington et al. (2015)
Pct_Obese	Obesity (reported BMI >= 30), as percent of adult population	socioeconomic	% by population	Remington et al. (2015)
Social_Association_Rate	Individuals with formal links to community programs, as percent of adult population ÷ 10,000	socioeconomic	% by population ÷ 10,000	Remington et al. (2015)
Violent_Crime_Rate	Number of violent crimes, as percent of adult population × 100,000	socioeconomic	% by population × 100,000	Remington et al. (2015)
Pct_Severe_Housing_Problems	Individuals lacking secure and fully functional housing, as percent of adult population	socioeconomic	% by household	Remington et al. (2015)
Pct_Food_Insecure	Individuals lacking adequate access to food, as percent of adult population	socioeconomic	% by population	Remington et al. (2015)
Pct_Rural	Individuals living in designated rural communities, as percent of total adult population	socioeconomic	% by population	Remington et al. (2015)
W-AAMR	Human age-adjusted mortality rate for non-Hispanic whites	response variable	no./100,000	US Centers for Disease Control and Prevention
B-AAMR	Human age-adjusted mortality rate for non-Hispanic blacks	response variable	no./100,000	US Centers for Disease Control and Prevention

Table S2. Results from Moran's I tests of spatial autocorrelation. Separate results are listed for White and Black age-adjusted morality rate in each of the three study regions.

Statistic	Northeast		Southeast		Southern Plains	
	White	Black	White	Black	White	Black
n	124		223		155	
I	-0.004	-0.069	-0.027	0.022	-0.033	-0.004
variance	0.033	0.033	0.016	0.016	0.065	0.007
p -value	0.463	0.603	0.553	0.401	0.491	0.339