Published in final edited form as:

Sociol Methods Res. 2020 November 1; 49(4): 1163–1185. doi:10.1177/0049124118769103.

Using universal kriging to improve neighborhood physical disorder measurement

Stephen J Mooney^a, Michael DM Bader^b, Gina S Lovasi^c, Kathryn M Neckerman^d, Andrew G Rundle^e, Julien O Teitler^f

^aHarborview Injury Prevention & Research Center, University of Washington, Seattle, WA

^bCenter on Health, Risk, and Society, American University, Washington, DC

^cDepartment of Epidemiology and Biostatistics, Dornsife School of Public Health, Drexel University, Philadelphia, PA

^dColumbia Population Research Center, Columbia University, New York, NY

^eDepartment of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY

^fSchool of Social Work, Columbia University, New York, NY

Abstract

Ordinary kriging, a spatial interpolation technique, is commonly used in social sciences to estimate neighborhood attributes such as physical disorder. Universal kriging, developed and used in physical sciences, extends ordinary kriging by supplementing the spatial model with additional covariates. We measured physical disorder on 1,826 sampled block faces across 4 US cities (New York, Philadelphia, Detroit, and San Jose) using Google Street View imagery. We then compared leave-one-out cross-validation accuracy between universal and ordinary kriging and used random subsamples of our observed data to explore whether universal kriging could provide equal measurement accuracy with less spatially dense samples. Universal kriging did not always improve accuracy. However, a measure of housing vacancy did improve estimation accuracy in Philadelphia and Detroit (7.9 and 6.8% lower root mean square error, respectively) and allowed for equivalent estimation accuracy with half the sampled points in Philadelphia. Universal kriging may improve neighborhood measurement.

Keywords

Cities; physical disorder; spatial analysis; universal kriging; urban health

Corresponding Author Stephen Mooney, Harborview Injury Prevention & Research Center, University of Washington, sjm2186@u.washington.edu, Phone: (206) 799 3977, Fax: (206) 744 9962.

Conflicts of interest: The authors have no conflicts of interest to declare.

Introduction

Physical Disorder, Health, and Society

Exposure to physical disorder – visual indications of urban deterioration such as litter, graffiti, vacant lots, and abandoned buildings (Raudenbush and Sampson 1999; Skogan 2015) – is of great interest in the social and population health sciences. Most famously, the controversial ‘broken windows’ theory that physical disorder encourages violent crime has been intensely scrutinized in the popular and academic press (Cerdá, Tracy, Messner, Vlahov, Tardiff, and Galea 2009; Sampson and Raudenbush 2004; Wilcox, Quisenberry, Cabrera, and Jones 2004; Wilson and Kelling 1982). Moreover, disorder has been linked to sexually transmitted disease independent of poverty, perhaps because a dilapidated environment encourages high-risk sexual behavior (Cohen, Spear, Scribner, Kissinger, Mason, and Wildgen 2000; Latkin, Curry, Hua, and Davey 2007). Latkin and Curry further proposed that exposure to disorder might cause depression (Latkin and Curry 2003), a finding that has subsequently been explored extensively (Browning, Soller, Gardner, and Brooks-Gunn 2013; Hill, Ross, and Angel 2005; Joshi, Mooney, Rundle, Quinn, Beard, and Cerda ; Mair, Roux, and Galea 2008). More broadly, exposure to physical disorder may induce fear and distress (Ross and Mirowsky 2009; Ross and Mirowsky 2001; Theall, Brett, Shirtcliff, Dunn, and Drury 2013), with consequent harms both to social ties and to personal health.

Policymakers, including public health authorities (Australian Local Government Association 2006; Centers for Disease Control and Prevention 2012), have an interest in physical disorder because it can be addressed through policies that, for example, green vacant lots (Branas, Cheney, MacDonald, Tam, Jackson, and Ten Have 2011) or implement graffiti or blight removal programs (Eagle 2007; Tavares 2014). Evidence of the benefits of program have been limited in part by the difficulty of obtaining accurate measurements of disorder (Skogan 2015), especially in light of potentially serious implications of broken windows policing (e.g., Duneier and Carter 1999). Broadly, the link between physical disorder and health-relevant behaviors has been inconsistent (Hoehner, Ramirez, Elliott, Handy, and Brownson 2005; Laraia, Messer, Evenson, and Kaufman 2007; Mendes de Leon, Cagney, Bienias, Barnes, Skarupski, Scherr, and Evans 2009; Strath, Greenwald, Isaacs, Hart, Lenz, Dondzila, and Swartz 2012) , and thus refinement of disorder estimates using available data may help to resolve or explain such inconsistencies.

Measuring Physical Disorder

One common approach to measuring disorder is to ask subjects to report their perceptions of disorder indicators in their local environment (e.g. (Ross and Mirowsky 1999)). Indeed, understanding perceptions of and behavioral response to different contexts is vital for informing place-based health promotion strategies (Blacksher and Lovasi 2012). Yet because relying on self-report alone to characterize the environment may result in same-source bias (Avolio, Yammarino, and Bass 1991; Duncan and Raudenbush 1999), researchers often complement or supplant self-reported perceived measures with neighborhood measures reported by nearby residents (Mujahid, Diez Roux, Morenoff, and Raghunathan 2007;

Mujahid, Roux, Cooper, Shea, and Williams 2011) or trained neighborhood auditors (Reiss 1971; Sampson and Raudenbush 1999).

There are several approaches to such independent neighborhood measurement. Some studies assess only the residential block of subjects whose health is being studied (e.g. (Franzini, Caughy, Nettles, and O'Campo 2008; Kimbro, Brooks-Gunn, and McLanahan 2011; Kwarteng, Schulz, Mentz, Zenk, and Opperman 2014)), which can be efficient if a neighborhood audit is added to a home visit protocol, but does not capture the broader surrounding context that is relevant to behaviors and health (Bader and Ailshire 2014) and typically precludes straightforward use of remote imagery for the audit (Bader, Mooney, and Rundle 2016). However, because auditing all blocks describing a subject's neighborhood is usually prohibitively expensive (though not always (Strath et al. 2012)), many studies aiming to characterize neighborhood physical disorder audit a random sample of blocks (King 2008; Sampson and Raudenbush 1999), in some cases using geostatistical tools to interpolate between sampled points (Auchincloss, Roux, Brown, Raghunathan, and Erdmann 2007; Bader and Ailshire 2014; Keyes, McLaughlin, Koenen, Goldmann, Uddin, and Galea 2012; Mooney, Bader, Lovasi, Neckerman, Teitler, and Rundle 2014).

Interpolating Physical Disorder Measures

One common geostatistical interpolation technique is ordinary kriging (Cressie 1988). Ordinary kriging takes advantage of spatial autocorrelation between observations (i.e. that closer measurements tend to be more similar) to estimate both a value and an estimate of error in that value estimate at unobserved locations. However, ordinary kriging does not account for between-location variation in small-scale geographic features such as street size or jurisdictional boundaries that might be correlated with indicators of the attributes of interest (Bader and Ailshire 2014). This can lead to error in imputed neighborhood traits. For example, because pedestrians create litter, urban streets with more retail activity and consequently more pedestrian traffic may have more litter on average, independent of other sources of disorder (Lovasi, Bader, Quinn, Neckerman, Weiss, and Rundle 2012). Ordinary kriging cannot incorporate measures of retail activity such as street classification or retail floor area into disorder interpolations.

One alternative to ordinary kriging is land use regression (Shmool, Kubzansky, Newman, Spengler, Shepard, and Clougherty 2014). Land use regression uses covariates measured at observed points to estimate what would be observed at other locations where covariates have been observed but the measure of interest has not (Ross, Jerrett, Ito, Tempalski, and Thurston 2007). For example, a land use regression model for physical disorder might use observation-specific measurements such as owner-occupancy (because owner-occupied homes tend to be better maintained (Sweeney 1974)) at observed locations to predict physical disorder at unobserved locations. Prior work has shown land use regression to perform better than ordinary kriging for spatial measures with considerable small-scale variation or where the source of the measured construct is known and well-measured (e.g. traffic as a predictor of air pollution) (Briggs, de Hoogh, Gulliver, Wills, Elliott, Kingham, and Smallbone 2000; Hoek, Beelen, De Hoogh, Vienneau, Gulliver, Fischer, and Briggs 2008). However, land use regression does not perform as well in contexts where land use is

not strongly tied to the measure of interest, as may be expected for physical disorder. Moreover, the information that an unobserved location might be near observed locations where high levels of disorder were observed would be lost in a pure land use regression model (Ross, English, Scaif, Gunier, Smorodinsky, Wall, and Jerrett 2006). Thus, both land use regression and ordinary kriging use only part of the information that might be available to base interpolations upon.

Universal kriging is an interpolation technique that incorporates *both* spatial correlation between observations and measured covariates in the same prediction model (Stein and Corsten 1991). For example, a universal kriging model can be fitted to spatial data wherein each point in the sample includes a location, a level of physical disorder, and covariates such as population density in the census block group where the sample was taken. Estimates at unobserved locations would then be made by first estimating the level of disorder as in ordinary kriging, then modeling the residual as a function of measured characteristics at observation locations and ‘correcting’ for the difference. In principle, this approach should improve estimation accuracy at unobserved locations if the measured characteristic incorporates additional information about the level of disorder beyond what can be inferred from autocorrelation of the measured points alone (Hengl, Heuvelink, and Stein 2003). However, to the best of our knowledge, while universal kriging has been used in research regarding health-relevant air quality metrics (Clougherty, Kheirbek, Eisl, Ross, Pezeshki, Gorczynski, Johnson, Markowitz, Kass, and Matte 2013), it has not been explored as a way to improve accuracy of neighborhood disorder estimates.

In this paper, we compare estimation accuracy using ordinary kriging to estimation accuracy using universal kriging to incorporate readily available US Census measures such as population density and housing vacancy to interpolate a measure of neighborhood disorder. We address two aspects of universal kriging. First, we examine the degree to which universal kriging including covariates from free sources of data improves the interpolated values of neighborhood conditions over ordinary kriging. Second, we examine whether universal kriging can reduce the sample size necessary to interpolate values at non-sampled locations. These analyses can help researchers create more accurate and cost-effective designs to measure neighborhood conditions.

Methods

Disorder Observations

For this investigation, we employed measures of neighborhood disorder in New York, NY; Philadelphia, PA; Detroit, MI; and San Jose, CA. The construction and validation of the measure against data from the 2010 decennial US Census has been described in detail elsewhere (Mooney et al. 2014) and validated in Detroit against a neighborhood disorder measure constructed using in-person audits roughly one year earlier (Mooney, Bader, Lovasi, Teitler, Koenen, Aiello, Galea, Goldmann, Sheehan, and Rundle 2017). Briefly, we sampled block faces in a grid across each city with oversamples corresponding to areas of greater researcher interest, generally more densely populated and deprived areas. On each of the 1,826 block faces in the final sample, trained street auditors used the Computer-Assisted Neighborhood Visual Assessment System (CANVAS) (Bader, Mooney, Lee, Sheehan,

Neckerman, Rundle, and Teitler 2015) and imagery from Google Street View (Anguelov, Dulong, Filip, Frueh, Lafon, Lyon, Ogale, Vincent, and Weaver 2010; Badland, Opit, Witten, Kearns, and Mavoa 2010; Rundle, Bader, Richards, Neckerman, and Teitler 2011; Wilson, Kelly, Schootman, Baker, Banerjee, Clennin, and Miller 2012) to measure presence of litter, empty bottles, graffiti, abandoned cars, poor building conditions, burned out buildings, boarded up buildings, vacant lots, and buildings with bars on windows. These items, whose presence on any given block face is time-dependent, but nonetheless reliably recur in more generally disordered spatial contexts (Mooney et al. 2014; Raudenbush and Sampson 1999). These items were combined in an “ecometric” measurement scale using a multilevel item-response model with items nested within street segments with coefficients measuring the “severity” of items (Raudenbush and Sampson 1999). Less severe indicators of disorder such as litter were nearly ubiquitous whereas more severe indicators such as abandoned buildings were very rare, but presence of more severe indicators generally implied presence of less severe indicators as well. Mean disorder was highest in Detroit, lowest in San Jose, and roughly equivalent in New York and Philadelphia (for more detail see Mooney, et al. 2014). The resulting scale had an internal consistency reliability score of 0.93.

Covariate Observations

Because our primary goal was to assess whether we could improve accuracy of spatial models by incorporating a free, readily available source of data, our primary spatially located covariates were economic and demographic measures taken from the US Census. We obtained 2010 Decennial US Census measures of population density, housing vacancy, and proportion of owner-occupied housing units occupied for each census block group (United States Census Bureau 2013). We selected population density because it is a driver of high pedestrian volume, which contributes to some indicators of disorder (Lovasi et al. 2012). We selected housing vacancy because vacant homes are frequently poorly maintained or subject to defacement, and because vacant buildings are sometimes considered an indicator of disorder in themselves (Skogan 2015). Finally, we selected owner occupied housing because of evidence from real estate economics that owner-occupied housing tends to be better maintained (Sweeney 1974). For the census measures, we selected block groups because they were the smallest spatial unit for which these census data were available. To conduct a sensitivity analysis, we also obtained all measures at the census tract level. As compared with block groups, tracts are larger spatial units, each containing at least one and usually several block groups. Tracts tend to be more heterogeneous than block groups (Iceland and Steinmetz 2003), and thus may convey less information for a universal kriging model.

Kriging

Ordinary kriging assumes intrinsic stationarity – that is, the spatial process (i.e. the social processes generating disorder) has a constant mean across the area for which the data are being interpolated, while the covariance between pairs of points depends only on the distance between the points (Cressie 1988). By contrast, universal kriging allows the spatial process to include an underlying trend (or ‘drift’ in the geostatistical literature) that can be described by a linear function of predictors and assumes intrinsic stationarity only after accounting for this underlying trend (Pebesma 2006). For physical disorder, as noted in the introduction, we anticipate that social processes such as neighborhood abandonment are

spatially patterned and may be explained by measured covariates, thus suggesting universal kriging may improve interpolation performance. All kriging techniques allow estimation of error surrounding the estimated value.

To quantify the incremental accuracy improvement afforded by incorporating covariates into the kriging model, we used leave-one-out (also known as jackknife) cross-validation to estimate the root mean squared error (RMSE) from each model (Bivand, Pebesma, Gomez-Rubio, and Pebesma 2008). That is, for each observation, we used all other observations to predict what that observation would be, then considered the difference between the model's prediction and the actual observation to represent the error in that location. Lower RMSE scores thus indicate improved model accuracy. The kriging process requires fitting a function estimating variance between observations as a function of distance between observations (the 'lag distance') (Bader and Ailshire 2014). Such functions can be fit visually or algorithmically; fitting visually allows more flexibility to minimize error at different spatial distances but fitting algorithmically limits subjectivity and error due to human input (Cressie 1985). We fit all variograms algorithmically using the Levenberg-Marquardt algorithm (Moré 1978).

Because kriging uses point locations, we assigned each block face observation the coordinates of the point midway between the blocks endpoints. We then assigned each point the characteristics of the census block group in which it was embedded. However, not all census measures were available for each sampled location. For example, the census provides no information regarding owner-occupancy of the non-residential tract that comprises Philadelphia International Airport. While this census data was missing for only 17 block faces (1% of all audited block faces), we wanted to retain maximal comparability between models. We decided that since only observations in residential census tracts without missing data would be used for our universal kriging, this should be balanced by a version of the ordinary kriging restricted to the same subset of observations. We therefore computed two RMSEs for ordinary kriging: one using all sampling points to represent the real-world ordinary kriging error and one using only the sample points for which census data were available to be more directly comparable to universal kriging models.

In all, we computed five cross-validated measures for each city using: (1) ordinary kriging, (2) ordinary kriging only for the sample points where all census data was available, (3) universal kriging incorporating population density as a predictor, (4) universal kriging incorporating housing vacancy as a predictor, and (5) universal kriging incorporating proportion of housing units occupied by owners as a predictor.

Sampling Density

In addition to improving estimation accuracy for a fixed sample density, incorporating covariates might also allow equally accurate models at decreased sample density, because the incorporated covariates provide additional information at unmeasured locations. To assess the benefits of universal kriging for the sensitivity of model accuracy to sampling density, we compared the decrease in accuracy from each kriging approach with smaller samples within Philadelphia. Specifically, we first randomly selected a 90% subset of sample points and computed the RMSE for the resulting model. We then repeated this

procedure for progressively smaller random subsets from 80% to 10%. We repeated this random subsetting procedure 100 times, for a total of 800 models fitted for each approach. We then used Wilcoxon rank-sum tests to identify presence of improvement in accuracy at each level of sample density. For this analysis, we tested accuracy of prediction using two covariates: population density and housing vacancy.

Software

All analyses used R for Windows version 3.2.3 (Vienna, Austria), including the ‘geoR’, ‘gstat’, and ‘rgdal’ packages for spatial analysis and ‘ggplot2’ to construct the figure (Pebesma 2004; Pebesma and Bivand 2015; Ribeiro Jr and Diggle 2001; Wickham 2009).

Results

The characteristics of each block face sample in each city are presented in Table 1. The four cities covered a diverse profile of urban form and economic prosperity. Whereas New York and San Jose represent extremes of population density, both are only minimally abandoned. Philadelphia represented a middle ground on the measures of interest in this study.

Kriging accuracy varied much more between cities than between kriging methods (Table 2). New York had the best prediction accuracy for all models (RMSE=0.288 for ordinary kriging), while Detroit had the worst (RMSE=0.595 for ordinary kriging). Universal kriging incorporating block group-level measures modestly improved estimation accuracy in most, though not all, instances. Of the covariates tested, housing vacancy was the most effective at increasing accuracy, particularly in Detroit and Philadelphia (6.8% and 7.9% improved accuracy, respectively), and population density added the least information (no more than 1.4% improved accuracy) in each city. Estimation accuracy improvement was more modest using covariates taken from the point’s census tract rather than its census block group (Appendix Table 1). Maps generated from ordinary kriging resulted in fewer, smoother clusters of disorder than those generated using universal kriging (Figure 1).

In the context in which universal kriging was most successful at reducing error in the overall sample, for housing vacancy in Philadelphia, universal kriging also substantially reduced the sampling density necessary to achieve a given accuracy (Figure 2, Appendix Table 2). For example, with only 200 sample points (40% of the full sample) incorporated in the kriging model, the median RMSE in the universal kriging model was 0.438, lower than the 0.439 RMSE for ordinary kriging for the full sample (Wilcoxon rank-sum test p-value < 0.01).

However, incorporating population density, a covariate that did not substantially improve the model in the full sample, into the universal kriging model in Philadelphia did not reduce the sample size needed to achieve a given accuracy (Appendix Figure 1).

Discussion

Reducing error in the measures used to quantify neighborhood environment characteristics can increase researchers’ ability to identify associations between neighborhood conditions and health outcomes. In this analysis of small-scale spatial patterns of physical disorder, we

compared the estimation accuracy of universal kriging incorporating readily available US Census measures to that of ordinary kriging. Incorporating a census measure of housing vacancy modestly improved estimation accuracy in Philadelphia and Detroit, cities with substantial areas of abandonment, but universal kriging did not uniformly improve accuracy across multiple measures in four cities. A measure that was effective in improving accuracy in Philadelphia, housing vacancy, was also effective in achieving equal or better accuracy with less dense samples in Philadelphia.

We found that both the theoretical relevance of the covariates included in models and the context in which disorder was being estimated had strong impacts on model accuracy. Housing vacancy, the covariate that most improved accuracy, is conceptually similar to neighborhood abandonment, which is a key driver of physical disorder (Sampson and Raudenbush 1999; Skogan 2015). The fact that housing vacancy measures did not substantially improve accuracy in New York and San Jose likely is consistent with very low vacancy rates throughout those cities, as shown in Table 1. Future research might productively explore the availability of and benefits of various forms of spatially located covariates more systematically, analogous to a genome-wide association study (GWAS) (Ioannidis, Loy, Poulton, and Chia 2009). Understanding which covariates are most informative in which contexts may be particularly valuable as spatially located administrative datasets become increasingly available, and as spatial tools develop to make ‘Big Data’ analyses incorporating such datasets simpler (Mooney, Westreich, and El-Sayed 2015). ‘Blind kriging’ is one approach to interpolation using empirical rather than a priori covariate selection, and has been explored in the computer simulation and materials design literature but to our knowledge not explored in social sciences (Joseph, Hung, and Sudjianto 2008).

The present study focused on demonstrating the advantages of universal kriging for a measure of neighborhood physical disorder. In principle, universal kriging is well suited to measures with micro-scale variation that ordinary kriging cannot account for but that also demonstrate spatial autocorrelation that covariates incorporated in land use regression cannot account for. For example, universal kriging might be appropriate for estimating prevalence of informal street vendors (Lucan, Varona, Maroko, Bumol, Torrens, and Wylie-Rosett 2013), whose locations are both affected by street scale and by broader neighborhood conditions, or tree pollen counts (Weinberger, Kinney, and Lovasi 2015), which are affected by proximity to individual trees, proximity to parks (which typically have different tree species than streets), and prevailing wind patterns. Broadly speaking, accurate measurement of contextual factors that vary within a city can be used to inform spatially targeted public health interventions and to coordinate a multi-sectoral response to address any health hazards in the environment.

However, it is important to note that universal kriging models performed worse than ordinary kriging models in some cases. In kriging models as in all statistical models, incorporating additional information increases the risk of model overfitting (Box 1976). Additionally, universal kriging, unlike ordinary kriging, allows interpolations to exceed the observed range of the data and could, therefore, increase the error of the interpolations if the covariates are not well specified. A prior study of air pollution found ordinary kriging to be

prone to biases due to over-fit models (Parmentier, McGill, Wilson, Regetz, Jetz, Guralnick, Tuanmu, Robinson, and Schildhauer 2014); by incorporating more information, universal kriging may raise this risk. A poorly specified measurement model will affect the accuracy of measures. For example, in San Jose, population density does not vary much across observed locations and does not inform the model. Universal kriging, unlike ordinary kriging, allows for estimates outside of the observed range of observations and thus introduces error not included in the ordinary kriging model. Broadly, we expect that universal kriging will improve measure accuracy when a) the covariate varies across spatial units, b) that variation is spatially correlated with the measure being interpolated, and c) the spatial correlation between the covariate and the measure exists at a smaller scale than the spatial density of the sample.

Our analysis showed that universal kriging with appropriate covariates helped reduce the sampling error and sampling variance over ordinary kriging. In practice this indicates that well-selected covariates can reduce the spatial density of samples required to interpolate values of disorder and other characteristics. The reduction in sampling density could be especially important for studies that rely on in-person audits where transportation to and from sampled locations makes up a substantial fraction of study costs (Mooney et al. 2017).

In data science research focused on accurate prediction, recent studies suggest promising fixes to overfitting by incorporating machine learning-derived algorithmic model fitting with tunable penalty parameters (Tibshirani 1996). Tunable models are particularly appropriate when estimation rather than inference is the goal of the modeling exercise (Breiman 2001), as is the case for spatial estimation of neighborhood measures. Penalized universal kriging models with many covariates? may be a promising area for future research.

This study had several notable strengths. Most importantly, our dataset included data from four cities which have very different economic and spatial profiles, so our results are relatively robust to city-specific artifacts. More broadly, the relatively large spatial sample of assessed block faces provided a solid base with which to explore sample size variability.

Limitations of this work include the lack of a gold standard benchmark. In practice there is no single optimal measure of disorder (Skogan 2015); future work on disorder measurement might focus on alternate methods of triangulation that incorporate ‘Big Data’ disorder indicators such as 311 calls (Mooney, Westreich, and El-Sayed 2015), and consider the potential for locale-dependent measurement error (as when an auditor familiar with a given context can distinguish better between a street mural and graffiti in that context than he or she might in another context) (Bader et al. 2015; Phillips, Engelberg, Geremia, Zhu, Kurka, Cain, Sallis, Conway, and Adams 2017). Future work estimating the value of various types of kriging models would also benefit from simulation studies using observational data for which true values are known. Another limitation is that the relevant scale for neighborhood disorder is likely different from that of other health-relevant constructs such as air pollution or pedestrian activity; thus our finding on the value of universal kriging may not generalize to other neighborhood measures. Our use of census data brings several limitation. First, census data captures only variability between block groups, leaving variability within block groups as a source of error, Second, census data likely captures only some of the underlying

drift in physical disorder levels; barriers such as highways or rivers might be larger determinants of spatial variation in disorder. Future work should investigate the use of distance metrics for kriging variograms that incorporate a distance penalty for crossing such a barrier. Finally, we used an algorithm to fit kriging functions to variograms in order to avoid biases in the comparison, but in practice, fitting variograms by eye is often recommended. Thus, our comparison may not replicate the true trade-off facing a spatial analyst interested in interpolating a physical disorder measure.

Conclusions

Universal kriging incorporating relevant spatially located covariates improved model estimation accuracy for one measure of neighborhood physical disorder. This improved accuracy leveraged pre-existing data and required no additional audit time. Figure 3 is a flow chart of how secondary data, universal kriging, and cross-validation could be used to improve neighborhood measure interpolation. Broadly, universal kriging holds promise for improved interpolation of neighborhood environmental measures in future research on social conditions. Our work serves as a demonstration of universal kriging in general, but additional investigation will be needed to identify the measures, covariates, and spatial scales for which universal kriging can most improve measurement of neighborhood conditions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

- Angelov Dragomir, Dulong Carole, Filip Daniel, Frueh Christian, Lafon Stéphane, Lyon Richard, Ogale Abhijit, Vincent Luc, and Weaver Josh. 2010. "Google street view: Capturing the world at street level." *Computer*:32–38.
- Auchincloss Amy H, Roux Ana V Diez, Brown Daniel G, Raghunathan Trivellore E, and Erdmann Christine A. 2007. "Filling the gaps: spatial interpolation of residential survey data in the estimation of neighborhood characteristics." *Epidemiology (Cambridge, Mass.)* 18:469.
- Australian Local Government Association. 2006. Age-friendly built environments: Opportunities for local government: Australian Local Government Association.
- Avolio Bruce J, Yammarino Francis J, and Bass Bernard M. 1991. "Identifying common methods variance with data collected from a single source: An unresolved sticky issue." *Journal of Management* 17:571–587.
- Bader Michael D M and Ailshire Jennifer A. 2014. "Creating Measures of Theoretically Relevant Neighborhood Attributes at Multiple Spatial Scales [Available online ahead of print February 7, 2014]." *Sociological Methodology*:(doi:10.1177/0081175013516749).
- Bader Michael D M, Mooney Stephen J, Lee Yeon Jin, Sheehan Daniel, Neckerman Kathryn M, Rundle Andrew G, and Teitler Julien O. 2015. "Development and deployment of the Computer Assisted Neighborhood Visual Assessment System (CANVAS) to measure health-related neighborhood conditions." *Health & place* 31:163–172. [PubMed: 25545769]
- Bader Michael DM, Mooney Stephen J, and Rundle Andrew G. 2016. "Protecting Personally Identifiable Information When Using Online Geographic Tools for Public Health Research." *American journal of public health* 106:206–208. [PubMed: 26794375]

- Badland HM, Opit S, Witten K, Kearns RA, and Mavoa S. 2010. "Can Virtual Streetscape Audits Reliably Replace Physical Streetscape Audits?" *Journal of Urban Health-Bulletin of the New York Academy of Medicine* 87:1007–1016. [PubMed: 21104331]
- Bivand Roger S, Pebesma Edzer J, Gomez-Rubio Virgilio, and Pebesma Edzer Jan. 2008. *Applied spatial data analysis with R*, vol. 747248717: Springer.
- Blacksher E and Lovasi GS. 2012. "Place-focused physical activity research, human agency, and social justice in public health: taking agency seriously in studies of the built environment." *Health Place* 18:172–9. [PubMed: 21940195]
- Box George EP. 1976. "Science and statistics." *Journal of the American Statistical Association* 71:791–799.
- Branas Charles C, Cheney Rose A, MacDonald John M, Tam Vicky W, Jackson Tara D, and Have Thomas R Ten. 2011. "A Difference-in-Differences Analysis of Health, Safety, and Greening Vacant Urban Space." *American Journal of Epidemiology* 174:1296. [PubMed: 22079788]
- Breiman Leo. 2001. "Statistical modeling: The two cultures (with comments and a rejoinder by the author)." *Statistical Science* 16:199–231.
- Briggs David J, de Hoogh Cornelis, Gulliver John, Wills John, Elliott Paul, Kingham Simon, and Smallbone Kirsty. 2000. "A regression-based method for mapping traffic-related air pollution: application and testing in four contrasting urban environments." *Science of the Total Environment* 253:151–167.
- Browning Christopher R, Soller Brian, Gardner Margo, and Brooks-Gunn Jeanne. 2013. "'Feeling Disorder' as a Comparative and Contingent Process Gender, Neighborhood Conditions, and Adolescent Mental Health." *Journal of health and social behavior* 54:296–314. [PubMed: 24026534]
- Centers for Disease Control and Prevention. 2012. "Healthy Aging & the Built Environment." vol. 2014.
- Cerdá Magdalena, Tracy Melissa, Messner Steven F, Vlahov David, Tardiff Kenneth, and Galea Sandro. 2009. "Misdemeanor policing, physical disorder, and gun-related homicide: a spatial analytic test of 'broken-windows' theory." *Epidemiology* 20:533–541. [PubMed: 19525689]
- Clougherty Jane E, Kheirbek Iyad, Eisl Holger M, Ross Zev, Pezeshki Grant, Gorczynski John E, Johnson Sarah, Markowitz Steven, Kass Daniel, and Matte Thomas. 2013. "Intra-urban spatial variability in wintertime street-level concentrations of multiple combustion-related air pollutants: the New York City Community Air Survey (NYCCAS)." *Journal of Exposure Science and Environmental Epidemiology* 23:232–240. [PubMed: 23361442]
- Cohen D, Spear S, Scribner R, Kissinger P, Mason K, and Wildgen J. 2000. "'Broken windows' and the risk of gonorrhea." *Am J Public Health* 90:230–6. [PubMed: 10667184]
- Cressie Noel. 1985. "Fitting variogram models by weighted least squares." *Journal of the International Association for Mathematical Geology* 17:563–586.
- . 1988. "Spatial prediction and ordinary kriging." *Mathematical Geology* 20:405–421.
- Duncan Greg J and Raudenbush Stephen W. 1999. "Assesing the effects of context in studies of child and youth development." *Educational psychologist* 34:29–41.
- Duneier Mitchell and Carter Ovie. 1999. *Sidewalk*: Macmillan.
- Eagle Steven J. 2007. "Does Blight Really Justify Condemnation?" *The Urban Lawyer*:833–858.
- Franzini Luisa, Caughy Margaret O'Brien, Nettles Sandra Murray, and O'Campo Patricia. 2008. "Perceptions of disorder: Contributions of neighborhood characteristics to subjective perceptions of disorder." *Journal of Environmental Psychology* 28:83–93.
- Hengl TOMISLAV, Heuvelink GERARDBM, and Stein ALFRED. 2003. "Comparison of kriging with external drift and regression-kriging." *Technical note, ITC* 51.
- Hill Terrence D, Ross Catherine E, and Angel Ronald J. 2005. "Neighborhood disorder, psychophysiological distress, and health." *Journal of Health and Social Behavior* 46:170–186. [PubMed: 16028456]
- Hoehner Christine M, Ramirez Laura K Brennan, Elliott Michael B, Handy Susan L, and Brownson Ross C. 2005. "Perceived and objective environmental measures and physical activity among urban adults." *American Journal of Preventive Medicine* 28:105–116. [PubMed: 15694518]

- Hoek Gerard, Beelen Rob, De Hoogh Kees, Vienneau Danielle, Gulliver John, Fischer Paul, and Briggs David. 2008. "A review of land-use regression models to assess spatial variation of outdoor air pollution." *Atmospheric environment* 42:7561–7578.
- Iceland John and Steinmetz Erika. 2003. "The effects of using census block groups instead of census tracts when examining residential housing patterns." Bureau of the Census.
- Ioannidis John PA, Loy En Yun, Poulton Richie, and Chia Kee Seng. 2009. "Researching genetic versus nongenetic determinants of disease: a comparison and proposed unification." *Science Translational Medicine* 1:7ps8–7ps8.
- Joseph V Roshan, Hung Ying, and Sudjianto Agus. 2008. "Blind kriging: A new method for developing metamodels." *Journal of mechanical design* 130:031102.
- Joshi Spruha, Mooney Stephen J, Rundle Andrew G, Quinn James W, Beard John B, and Cerda Magdalena. "Pathways from neighborhood poverty to depression among older adults." *Under Review*.
- Keyes Katherine M, McLaughlin Katie A, Koenen Karestan C, Goldmann Emily, Uddin Monica, and Galea Sandro. 2012. "Child maltreatment increases sensitivity to adverse social contexts: neighborhood physical disorder and incident binge drinking in Detroit." *Drug and alcohol dependence* 122:77–85. [PubMed: 21981990]
- Kimbrow Rachel Tolbert, Brooks-Gunn Jeanne, and McLanahan Sara. 2011. "Young children in urban areas: links among neighborhood characteristics, weight status, outdoor play, and television watching." *Social science & medicine* 72:668–676. [PubMed: 21324574]
- King Diane. 2008. "Neighborhood and individual factors in activity in older adults: results from the neighborhood and senior health study." *Journal of Aging and Physical Activity* 16:144–170. [PubMed: 18483439]
- Kwarteng Jamila L., Schulz Amy J., Mentz Graciela B., Zenk Shannon N., and Opperman Alisha A.. 2014. "Associations between observed neighborhood characteristics and physical activity: findings from a multiethnic urban community." *Journal of Public Health* 36:358–367. [PubMed: 24159053]
- Laraia Barbara, Messer Lynne, Evenson Kelly, and Kaufman Jay S.. 2007. "Neighborhood Factors Associated with Physical Activity and Adequacy of Weight Gain During Pregnancy." *Journal of Urban Health* 84:793–806. [PubMed: 17710552]
- Latkin Carl A and Curry Aaron D. 2003. "Stressful neighborhoods and depression: a prospective study of the impact of neighborhood disorder." *Journal of health and social behavior*:34–44. [PubMed: 12751309]
- Latkin Carl A, Curry Aaron D, Hua Wei, and Davey Melissa A. 2007. "Direct and indirect associations of neighborhood disorder with drug use and high-risk sexual partners." *American journal of preventive medicine* 32:S234–S241. [PubMed: 17543716]
- Lovasi Gina S., Bader Michael D. M., Quinn James, Neckerman Kathryn, Weiss Christopher, and Rundle Andrew. 2012. "Body Mass Index, Safety Hazards, and Neighborhood Attractiveness." *American Journal of Preventive Medicine* 43:378–384. [PubMed: 22992355]
- Lucan Sean C, Varona Monica, Maroko Andrew R, Bumol Joel, Torrens Luis, and Wylie-Rosett Judith. 2013. "Assessing mobile food vendors (aka street food vendors)—methods, challenges, and lessons learned for future food-environment research." *Public health* 127:766–776. [PubMed: 23891280]
- Mair Christina F, Roux Ana V Diez, and Galea Sandro. 2008. "Are neighborhood characteristics associated with depressive symptoms? A critical review." *Journal of epidemiology and community health*:jech. 2007.066605.
- Mendes de Leon Carlos F, Cagney Kathleen A, Bienias Julia L, Barnes Lisa L, Skarupski Kimberly A, Scherr Paul A, and Evans Denis A. 2009. "Neighborhood Social Cohesion and Disorder in Relation to Walking in Community-Dwelling Older Adults A Multilevel Analysis." *Journal of Aging and Health* 21:155–171. [PubMed: 19144973]
- Mooney Stephen J, Bader Michael DM, Lovasi Gina S, Neckerman Kathryn M, Teitler Julien O, and Rundle Andrew G. 2014. "Validity of an ecometric neighborhood physical disorder measure constructed by virtual street audit." *American journal of epidemiology* 180:626–635. [PubMed: 25122584]

- Mooney Stephen J, Westreich Daniel J, and El-Sayed Abdulrahman M. 2015. "Commentary: Epidemiology in the Era of Big Data." *Epidemiology* 26:390–394. [PubMed: 25756221]
- Mooney Stephen J., Bader Michael D. M., Lovasi Gina S., Teitler Julien O., Koenen Karestan C., Aiello Allison E., Galea Sandro, Goldmann Emily, Sheehan Daniel M., and Rundle Andrew G.. 2017. "Street Audits to Measure Neighborhood Disorder: Virtual or In-Person?" *American Journal of Epidemiology* 186:265–273. [PubMed: 28899028]
- Moré Jorge J. 1978. "The Levenberg-Marquardt algorithm: implementation and theory." Pp. 105–116 in *Numerical analysis*: Springer.
- Mujahid MS, Diez Roux AV, Morenoff JD, and Raghunathan T. 2007. "Assessing the measurement properties of neighborhood scales: from psychometrics to econometrics." *Am J Epidemiol* 165:858–67. [PubMed: 17329713]
- Mujahid Mahasin S, Roux Ana V Diez, Cooper Richard C, Shea Steven, and Williams David R. 2011. "Neighborhood stressors and race/ethnic differences in hypertension prevalence (the Multi-Ethnic Study of Atherosclerosis)." *American journal of hypertension* 24:187–193. [PubMed: 20847728]
- Parmentier Benoit, McGill Brian, Wilson Adam M, Regetz James, Jetz Walter, Guralnick Robert P, Tuanmu Mao-Ning, Robinson Natalie, and Schildhauer Mark. 2014. "An assessment of methods and remote-sensing derived covariates for regional predictions of 1 km daily maximum air temperature." *Remote Sensing* 6:8639–8670.
- Pebesma Edzer J. 2004. "Multivariable geostatistics in S: the gstat package." *Computers & Geosciences* 30:683–691.
- Pebesma Edzer J and Bivand Roger. 2015. "Classes and methods for spatial data."
- Pebesma Edzer J. 2006. "The Role of External Variables and GIS Databases in Geostatistical Analysis." *Transactions in GIS* 10:615–632.
- Phillips Christine B, Engelberg Jessa K, Geremia Carrie M, Zhu Wenfei, Kurka Jonathan M, Cain Kelli L, Sallis James F, Conway Terry L, and Adams Marc A. 2017. "Online versus in-person comparison of Microscale Audit of Pedestrian Streetscapes (MAPS) assessments: reliability of alternate methods." *International journal of health geographics* 16:27. [PubMed: 28778205]
- Raudenbush SW and Sampson RJ. 1999. "Econometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods." *Sociological Methodology* 1999, Vol 29 29:1–41.
- Reiss Albert J. 1971. "Systematic observation of natural social phenomena." *Sociological methodology* 3:3–33.
- Ribeiro Paulo J Jr and Diggle Peter J. 2001. "geoR: A package for geostatistical analysis." *R news* 1:14–18.
- Ross Catherine E and Mirowsky John. 2009. "Neighborhood disorder, subjective alienation, and distress." *Journal of Health and Social Behavior* 50:49–64. [PubMed: 19413134]
- Ross Catherine E. and Mirowsky J. 2001. "Neighborhood disadvantage, disorder, and health." *Journal of Health and Social Behavior* 42:258–276. [PubMed: 11668773]
- Ross Catherine E. and Mirowsky John. 1999. "Disorder and decay the concept and measurement of perceived neighborhood disorder." *Urban Affairs Review* 34:412–432.
- Ross Zev, English Paul B, Scalf Rusty, Gunier Robert, Smorodinsky Svetlana, Wall Steve, and Jerrett Michael. 2006. "Nitrogen dioxide prediction in Southern California using land use regression modeling: potential for environmental health analyses." *Journal of Exposure Science and Environmental Epidemiology* 16:106–114. [PubMed: 16047040]
- Ross Zev, Jerrett Michael, Ito Kazuhiko, Tempalski Barbara, and Thurston George D. 2007. "A land use regression for predicting fine particulate matter concentrations in the New York City region." *Atmospheric Environment* 41:2255–2269.
- Rundle AG, Bader MD, Richards CA, Neckerman KM, and Teitler JO. 2011. "Using Google Street View to audit neighborhood environments." *Am J Prev Med* 40:94–100. [PubMed: 21146773]
- Sampson RJ and Raudenbush SW. 1999. "Systematic social observation of public spaces: A new look at disorder in urban neighborhoods." *American Journal of Sociology* 105:603–651.
- . 2004. "Seeing disorder: Neighborhood stigma and the social construction of "Broken windows"." *Social Psychology Quarterly* 67:319–342.

- Shmool Jessie LC, Kubzansky Laura D, Newman Ogonnaya Dotson, Spengler John, Shepard Peggy, and Clougherty Jane E. 2014. "Social stressors and air pollution across New York City communities: a spatial approach for assessing correlations among multiple exposures." *Environmental Health* 13:91. [PubMed: 25374310]
- Skogan Wesley. 2015. "Disorder and Decline The State of Research." *Journal of Research in Crime and Delinquency* 52:464–485.
- Stein A and Corsten LCA. 1991. "Universal kriging and cokriging as a regression procedure." *Biometrics*:575–587.
- Strath Scott J., Greenwald Michael J., Isaacs Raymond, Hart Teresa L., Lenz Elizabeth K., Dondzila Christopher J., and Swartz Ann M.. 2012. "Measured and perceived environmental characteristics are related to accelerometer defined physical activity in older adults." *International Journal of Behavioral Nutrition and Physical Activity* 9:40.
- Sweeney James L. 1974. "Housing unit maintenance and the mode of tenure." *Journal of Economic Theory* 8:111–138.
- Tavares Samantha Silva. 2014. "California Graffiti Removal Programs: Benchmarking San José's Graffiti Abatement Program against Best Practices in the Cities of Long Beach, San Diego, and Santa Ana."
- Theall Katherine P, Brett Zoë H, Shirtcliff Elizabeth A, Dunn Erin C, and Drury Stacy S. 2013. "Neighborhood disorder and telomeres: Connecting children's exposure to community level stress and cellular response." *Social Science & Medicine* 85:50–58. [PubMed: 23540366]
- Tibshirani Robert. 1996. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society. Series B (Methodological)*:267–288.
- United States Census Bureau. 2013. "TIGER/Line Shapefiles."
- Weinberger Kate R, Kinney Patrick L, and Lovasi Gina S. 2015. "A Review of Spatial Variation of Allergenic Tree Pollen Within Cities." *Arboriculture & Urban Forestry* 41.
- Wickham Hadley. 2009. *ggplot2: elegant graphics for data analysis*: Springer Science & Business Media.
- Wilcox Pamela, Quisenberry Neil, Cabrera Debra T, and Jones Shayne. 2004. "Busy places and broken windows? Toward defining the role of physical structure and process in community crime models." *The Sociological Quarterly* 45:185–207.
- Wilson Jeffrey S, Kelly Cheryl M, Schootman Mario, Baker Elizabeth A, Banerjee Aniruddha, Clennin Morgan, and Miller Douglas K. 2012. "Assessing the Built Environment Using Omnidirectional Imagery." *American Journal of Preventive Medicine* 42:193–199. [PubMed: 22261217]
- Wilson JQ and Kelling GL. 1982. "Broken Windows." *Atlantic Monthly* 249:29–36.

Figure 1.

Maps of census tracts in Philadelphia. The left panel shows mean physical disorder in each tract computed using ordinary kriging, the middle panel shows physical disorder computed using universal kriging incorporating 2010 US Census data on housing vacancy as a covariate, and the right panel shows 2010 US Census data on housing vacancy for each tract.

Figure 2:

Comparison of the empirical error in kriging models and size of the subset sample.

Universal kriging with % vacancy was not only more accurate overall but also reduced the number of sample points needed to achieve a given accuracy as compared with ordinary kriging.

Figure 3:

A potential work flow for researchers considering using universal kriging to improve neighborhood measurement. This workflow could be used either with pilot data to estimate the sample density needed using a suitable universal kriging covariate or used with final data to improve the final measurement model.

Table 1:

Selected characteristics of the block face sample for each city

	Audited Segments	Segments with census data (% of total)	Population Density per square mile (Median, IQR)	% Vacant (Median, IQR)	% Owner- occupied (Median, IQR)
New York	532	520 (98)	51,982 (28,411-81,251)	6 (4-8)	23 (9-47)
Detroit	503	500 (100)	6,567 (5,193-7,712)	22 (17-27)	56 (49-63)
Philadelphia	502	500 (100)	18,983 (14,047-24,595)	11 (7-15)	54 (45-63)
San Jose	289	289 (100)	2,903 (2,408-3,549)	3 (3-4)	61 (40-73)

Cross-validation results evaluating improvement in in-sample prediction accuracy achieved when incorporating covariates in universal kriging model

Table 2:

Geographic Feature included as a Universal Kriging Covariate	New York, NY		Detroit, MI		Philadelphia, PA		San Jose, CA	
	RMSE	Percent improvement	RMSE	Percent improvement	RMSE	Percent improvement	RMSE	Percent improvement
Ordinary Kriging								
All sample points	0.288	--	0.595	--	0.441	---	0.326	---
Points with census data	0.286	0.7%	0.581	2.3%	0.439	0.3%	0.326	0.0%
Population Density in Block Group	0.286	0.8%	0.587	1.4%	0.440	0.1%	0.325	0.4%
Percent Vacant in Block Group	0.283	1.7%	0.555	6.8%	0.406	7.9%	0.342	-4.8%
Owner-occupied Housing	0.289	-0.3%	0.570	4.2%	0.432	1.9%	0.318	2.6%

RMSE: Root Mean Squared Error (lower is better)