

## Incorporating spatial autocorrelation in dasymetric mapping: A hierarchical Poisson spatial disaggregation regression model

Bowen He <sup>a,\*</sup>, Jonathan M. Gilligan <sup>a,b</sup>, Janey V. Camp <sup>c</sup>

<sup>a</sup> Department of Civil and Environmental Engineering, Vanderbilt University, PMB 351831, 2301 Vanderbilt Place, Nashville, TN, 37235-1831, United States

<sup>b</sup> Department of Earth and Environmental Science, Vanderbilt University, PMB 351805, 2301 Vanderbilt Place, Nashville, TN, 37235-1805, United States

<sup>c</sup> Center for Applied Earth Science and Engineering Research (CAESER), The University of Memphis, Wilder Tower 9th Floor, Memphis, TN, 38152, United States

### ARTICLE INFO

#### Keywords:

Population modeling (PM)  
Spatial data analysis  
NLCD (national land cover database)  
Spatial autocorrelation  
Gaussian linear mixed model (GLMM)  
Template model builder (TMB)

### ABSTRACT

The growing demand for spatially detailed population products in various fields continues to rise, as users shift their focus from aggregated areal totals to high-resolution grid estimates. Aggregating demographic data to areas, such as census tracts or block groups, can mask localized heterogeneities within those areas. This paper presents a new pycnophylactic (density-preserving) geospatial model for disaggregating population to high-resolution grids. We describe a Bayesian Hierarchical Poisson Spatial Disaggregation Regression Model (HPSDRM), which incorporates land cover covariates and two levels of spatial autocorrelation. We evaluated the model's predictive ability first with simulation studies, and then by disaggregating census population data for Davidson County, TN, from the census tract-level to a fine grid and comparing predicted to actual block-level population counts. The interpolated population map successfully identified spatial heterogeneities, such as hot- and cold-spots within census tracts. The HPSDRM model out-performed three other types of disaggregation modeling, which suggests the value of incorporating spatial autocorrelation. Based upon this study, HPSDRM has potential for disaggregating other demographic data, such as socioeconomic indicators.

### 1. Introduction

Maps of demographic variables with high spatial resolution are valuable for investigating coupled social-environmental systems (He & Ding, 2021, 2023; He & Guan, 2021b; He et al., 2024b). As mathematical models have been widely applied to evaluate various complex social-environmental systems, population spatial disaggregation models are also urgently needed (He & Guan, 2021a, 2022a, b). Spatial distributions of population and other demographic variables are crucial in analyses of social vulnerability, community resilience, and environmental justice, and many other applications, and because vulnerability and inequality are often concentrated and heterogeneously distributed within census tracts or other spatial units, there is demand for estimates of the distribution of these data at higher resolution (Camp et al., 2023; He, Gilligan, & Camp, 2023; He, Zheng, & Guan, 2023; He, 2023; Nelson et al., 2015; He & Guan, 2022a, 2022b, 2024a, b). Restricting analyses to larger areas, such as tracts, can lead to underestimating spatial variations and non-stationarities (Mennis, 2003; Wu & Murray, 2005), and contributes to the modifiable areal unit problem, in which, fine-scale inhomogeneities can make the statistics of spatially aggregated data

sensitive to small displacements of boundaries, such as when census tracts are modified (Openshaw, 1983). Thus, it is valuable to develop a spatial disaggregation model that can reliably recover spatial variations within administrative units.

Dasymetric mapping, an application of regression models to spatial disaggregation, has seen renewed interest because of rapid progress in algorithms, Geographic Information System (GIS), and remote sensing technologies (Mennis, 2009; Petrov, 2012). The technique uses information such as land cover types or other ancillary data, such as tax parcels, building footprints, or remotely sensed night light intensity, to redistribute aggregated data on finer spatial scales. This is valuable for such purposes as environmental justice analyses and hazard mitigation, because of mismatches between the spatial scales of data aggregation versus the effect of interest, such as flood inundation contours. Mennis and Hultgren (2006) invented an intelligent dasymetric mapping (IDM) approach that combined an analyst's subjective knowledge with empirical sampling to parameterize the relationship between the National Land Cover Database (NLCD) and underlying high-resolution population distributions. IDM proved superior in predictive accuracy to traditional areal weighting and 'binary' dasymetric mapping

\* Corresponding author.

E-mail address: [bowen.he@vanderbilt.edu](mailto:b Bowen.he@vanderbilt.edu) (B. He).

approaches (Mennis & Hultgren, 2006). The IDM has been widely adopted for environmental justice assessments. Giordano and Cheever (2010) used IDM to produce improved estimates of the spatial distribution of different socioeconomic groups and allowed them to identify communities at risk from hazardous waste generation in San Antonio, Texas. They found that Black and Hispanic people, individuals living below the poverty line, and those who rent were more likely to live near hazardous waste generation sites (Giordano & Cheever, 2010).

Another widely-used dasymetric mapping technique is the Cadastral-based Expert Dasymetric System (CEDS) (Maantay et al., 2007). Rather than using uniform spatial resolution ancillary information such as the land cover data from the NLCD that is a 30m resolution, CEDS uses non-uniformly distributed tax parcel information (e.g. property value, property type, and property area) to delineate the heterogeneous spatial distribution of demographic data in urban areas. The method is widely applied to environmental and health assessment problems in urban regions such as New York City (Maantay et al., 2007). CEDS can ameliorate problems underestimating vulnerable populations exposed to environmental hazards, which often manifest at much finer spatial scales than census tracts or block groups. Nelson et al. (2015) developed a hybrid method for creating a social vulnerability index (SoVI) at a tax parcel level by using supplementary information about tax parcels to link the CEDS technique and the established social vulnerability indexing method to uncover vulnerable groups that were previously masked by areal aggregation. Nonetheless, an important disadvantage of the CEDS approach is that it requires tax-parcel data, which is not always available for the region of interest. To the best of our knowledge, unlike the NLCD database, there is no database of tax parcel data across the U.S. that could be used as ancillary data for CEDS.

Other dasymetric mapping approaches include using built-area and building-height data as ancillary information (Alahmadi et al., 2014), using high-resolution address point datasets to conduct the dasymetric mapping (Zandbergen, 2011); developing multi-layer multi-class dasymetric mapping to estimate population surface (Su et al., 2010); using means of raster pixel maps to rapidly facilitate dasymetric-based population interpolation (Langford, 2007); applying a hybrid model with different ancillary data combination such as land cover data combined with tax parcel data or land cover data combined with NTL (Briggs et al., 2007; Jia & Gaughan, 2016); and applying machine learning models, such as random forests, to combinations of demographic data and remotely-sensed ancillary data (Stevens et al., 2015). While these dasymetric mapping methods provide valuable insights into integrating aggregate demographic data with ancillary data to estimate finer scale distributions of demographic indicators, none of these models incorporated either spatial autocorrelations or non-stationarity. As a result, the dependent variables and their latent field in those models are all treated as identically and independently distributed (iid) observations, although they are spatial statistical data with autocorrelation characteristics.

Historically, several studies have been proposed to address spatial autocorrelation and non-stationarity challenges in the population disaggregation field. For example, Li and Corcoran (2011) suggested dividing the study area into a series of subregions and performing a separate population redistribution within each subregion. The problem with this method is that the strategy of dividing the study area is arbitrary and the newly divided subregions' boundaries are unlikely to represent areas with homogeneous population distribution characteristics. This makes the method vulnerable to the modifiable area unit problem. Other local regression approaches estimated separate coefficients for each population distribution feature using quantile regression (QR) (Cromley et al., 2012) and geographically weighted regression (GWR) (Lin et al., 2011). Although these approaches helped improve prediction accuracy, their estimates of spatial autocorrelation are very sensitive to the configuration of the model's regression covariates and their spatial distribution features (Cockx & Canters, 2015; Lee, 2011).

Based on these literature reviews, we hypothesize that incorporating

spatial autocorrelation into the dasymetric mapping model substantially enhances the precision of grid-based population distribution interpolation. Meanwhile, we also hypothesize a hierarchical Bayesian linear mixed model can successfully encapsulate spatial autocorrelation. Here, we present a Bayesian hierarchical Poisson Spatial Disaggregation Regression Model (HPSDRM) that combines a conditional autoregressive (CAR) model for observed coarsely aggregated areal data with a Gaussian random field for predicting finer-scale grids that explicitly deals with spatial non-stationarity by incorporating two Gaussian random effects to estimate spatial autocorrelation features of the underlying population distribution. Instead of the spatial non-stationarity model that incorporates the model coefficients trend based on locations, this proposed model aims to explain spatial autocorrelation features of population distribution that cannot be explained by the regression covariates' spatial structure or traditional dasymetric sampling strategies. Meanwhile, we prioritized practicality by designing the model to work with the fewest possible predictors without compromising its interpolation effectiveness. This approach also aims to minimizes the model's reliance on external, often inaccessible, ancillary data. Our newly introduced model, the proposed HPSDRM in this study, meets these criteria by internally handling spatial autocorrelation and requiring only readily available national land cover data that is available across the country to make spatial interpolations.

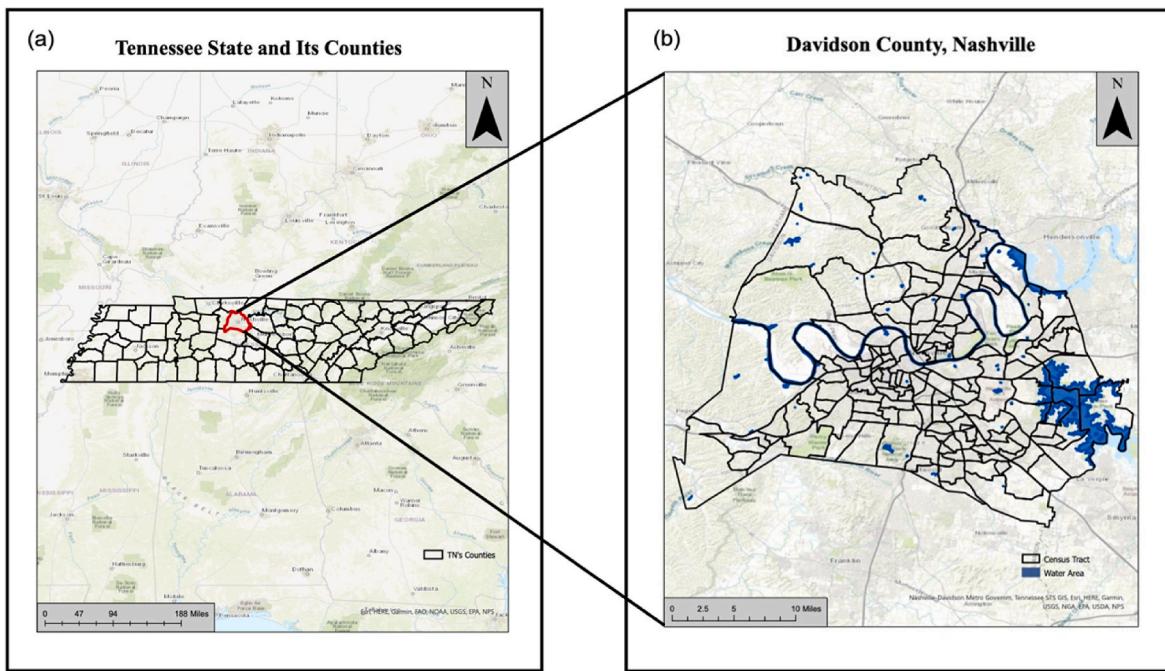
Our model uses a hierarchical generalized linear mixed-effects modeling structure (Gelman and Hill, 2006; Woltman et al., 2012). A linear predictor is included to link the areal aggregated observations and the prediction grids by two-level latent Gaussian random processes, and then combining these with gridded land cover data, so that the spatial autocorrelation pattern of the population distribution is incorporated in the dasymetric mapping process. We evaluate this model by comparing its predictive accuracy to three other dasymetric methods: IDM, a traditional ordinary least squares linear regression model (OLS), and the conventional areal weighting mapping (AWM) method. We compare the performance of each of these four models in predicting observed census block populations in Davidson County, TN, USA, from census tract data.

The remainder of this paper is structured as follows: In Section 2, we describe and discuss the study area, its associated population data sets, and the ancillary land cover data. In Section 3, we briefly describe and discuss the IDM, the OLS, and AWM methods, and present the hierarchical Poisson spatial disaggregation regression model (HPSDRM) and its accompanying Bayesian inferential process, which uses a Stochastic Partial Differential Equation (SPDE) approach (Lindgren et al., 2011), realized in the R programming language with the Integrated Nested Laplace Approximation package (Bakka et al., 2018) and the template model builder (TMB) package (Kristensen et al., 2016). Section 4 describes using a simulation study to evaluate the predictive performance of the HPSDR model. Section 5 presents the comparative assessment of HPSDRM to IDM, OLS, and AWM, using a case study of census tract and block population data for Davidson County, TN. Finally, section 6 discusses this study and presents our conclusions.

## 2. Data and study area

Davidson County, TN, USA, is used as the case study region (Fig. 1). Nashville is the county seat, and the Cumberland River flows through it, from east to west (Fig. 1(b)). It encompasses 174 census tracts, 487 census block groups, and 9097 census blocks with a total population of 715,884 as of the 2020 decennial census (United States Census Bureau, 2021). Davidson County encompasses both rural and urban areas, with the population clustered in urban tracts. This produces a strongly autocorrelated spatial distribution, which was evaluated using a Moran's I analysis (see Supplemental Materials).

The National Land Cover Database (NLCD) was used as ancillary information to serve as the covariate predictors. The data was accessed through the U.S. Geological Survey (USGS)'s Multi-Resolution Land Characteristics (MRLC) Consortium. We used the NLCD 2019 land-cover



**Fig. 1.** Study area: (a) the state of Tennessee and its counties; (b) Davidson County, Nashville.

classification data, which represents land cover for the 48 contiguous states, using 16 categories, with 30m spatial resolution (Anderson et al., 1976; Dewitz and Geological Survey, 2021).

Using these population and land cover data, a series of population dasymetric maps were created using the IDM, OLS, AWM, and HPSDRM models. We used a set of cross-validated penalized regression algorithms to select the statistically significant land cover classes as covariates in the OLS and HPSDRM. Detailed information about the cross-validated penalized regression algorithms and its associated results are presented in the [Supplementary Material section 4 Cross-Validated Penalized Regression Analysis](#).

### 3. Methods

The research design of this study consists of three stages. The first stage is simulation studies that test the predictive ability of the proposed model in a designed areal and grid configuration using random realizations generated from the proposed model structure. Data are generated at high resolution by sampling from parametric distributions with known parameters. Then these data are aggregated at coarse resolution and the HPSDRM model is assessed for its ability to accurately recover the original high-resolution data from the aggregated data. We conducted two simulation studies: (1) CAR model simulations; (2) integrated Matérn and CAR hierarchical model simulations. The simulation studies are described in detail in Section 3.1.

The second stage is a case study that predicts high-resolution gridded populations from 2020 census tract populations and NLCD land cover for Davidson County. Specifically, a Moran's I analysis was implemented on the Nashville block population density data to test the existence of spatial autocorrelation characteristics of the area's population distribution. It should be noted that the Moran's I analysis was just used to demonstrate that the population is highly autocorrelated in the block unit level and the block level data was not used to calibrate the model coefficients. The detailed information regarding Moran's I analysis and its results are summarized in the [Supplementary Material section 3 Moran's I Analysis](#). Then, the disaggregated grid population can be aggregated into the Decennial census blocks population to compute three performance metrics of RMSE,  $R^2$ , and MAE to evaluate the model's predictive ability. Finally, three other dasymetric mapping

models, including the linear regression model, the IDM, and the AWM, were included to perform the same population disaggregation regression task in the Nashville study area, and their performance metrics were compared with the proposed HPSDRM. The research design of this study is revealed in [Fig. 2](#), and detailed information about each research stage is illustrated in the following sections.

#### 3.1. Simulation study

The simulation study tests the spatial interpolation ability of the proposed HPSDRM by generating simulated high-resolution spatial data, which is drawn at random from parametric distributions with known parameters and hyperparameters, aggregating this data at coarser resolution, and measuring the accuracy with which the HPSDRM model can estimate both the high-resolution data and the parameters and hyperparameters that characterize the probability distributions from which it was drawn. Data were generated on a  $150 \times 150$  grid ( $n_p = 22500$ ), and then aggregated to a coarser  $30 \times 30$  "observation" grid ( $n_A = 900$ ). The two grid configurations in this simulation study are shown in [Fig. 3](#). Details about the simulation study configuration and results are presented in the [Supplementary Material section 8 Simulation Study](#).

#### 3.2. Davidson County, Nashville case study

In the Nashville case study, the population was disaggregated from census tract populations on a  $150m \times 150m$  grid, using each of the four dasymetric mapping models. We briefly describe the four models here, and further details are presented in the [Supplementary Material section 4 - 7](#).

##### 3.2.1. Hierarchical Poisson spatial disaggregation regression model (HPSDRM)

The HPSDRM model begins with a study area  $\mathcal{A} \in \mathbb{R}^2$ , which can be partitioned into  $n_{\mathcal{A}}$  areal units  $\mathcal{A}_1, \dots, \mathcal{A}_{n_{\mathcal{A}}}$  of arbitrary size and shape, for which there are population observations  $Y_{\mathcal{A}_1}, \dots, Y_{\mathcal{A}_{n_{\mathcal{A}}}}$ . Underlying these observations are population distributions on a finer grid  $S$ , consisting of  $n_p$  grid cells  $s_1, \dots, s_{n_p}$ , where each cell  $s_i$  has an unobserved population count  $Y_i$ . The model incorporates two levels of spatial

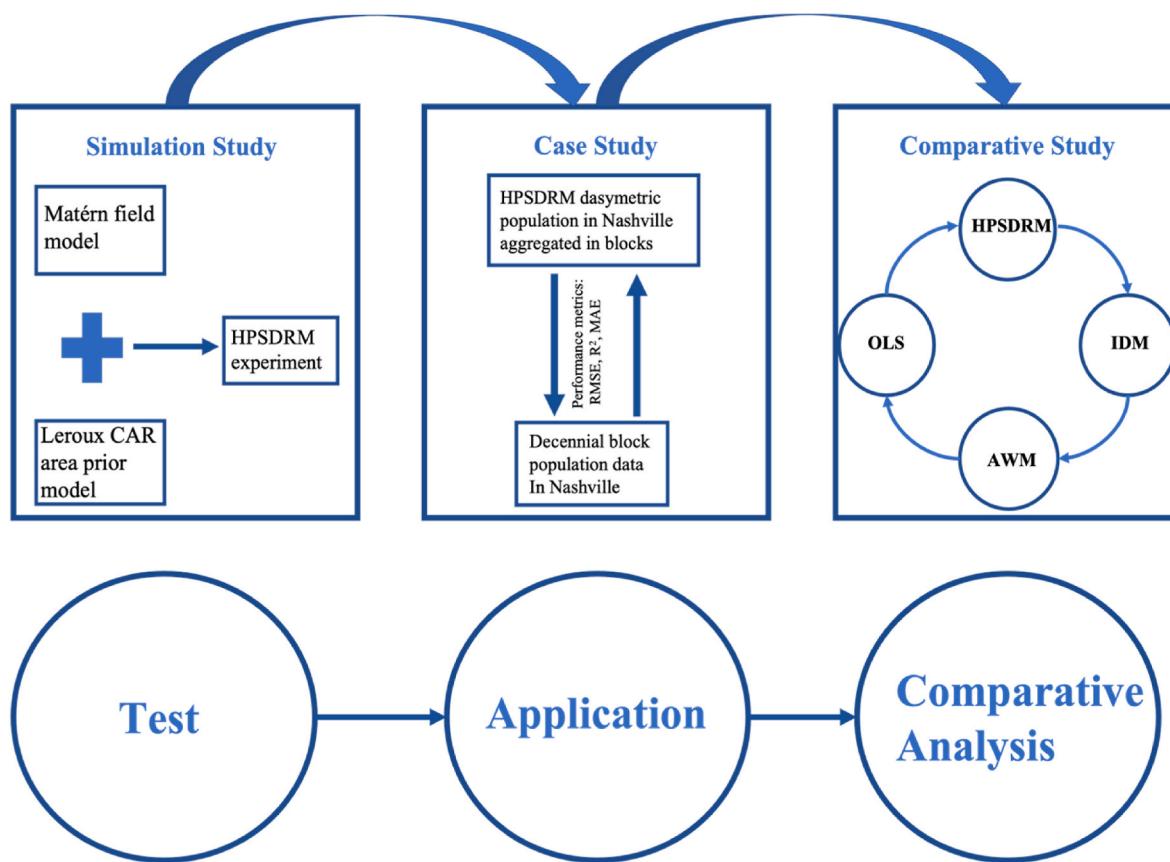


Fig. 2. The research design of the study.

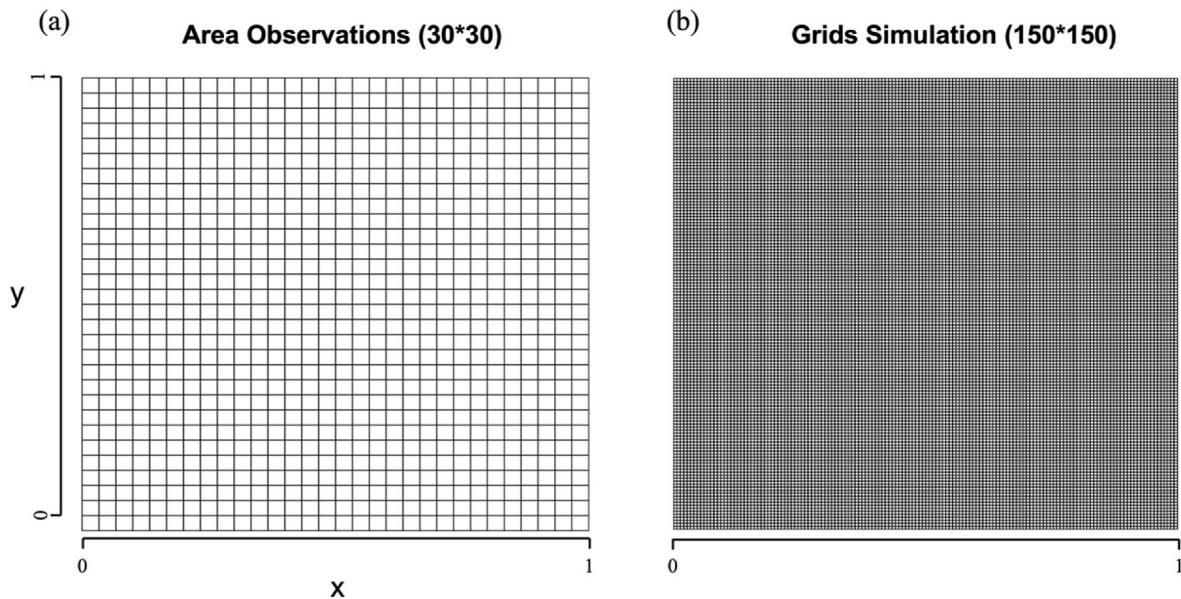


Fig. 3. Plots of the (a) observational ( $30 \times 30$ ) and (b) underlying grid ( $150 \times 150$ ) configurations used in the simulation study.

random effects to characterize the spatial autocorrelation features at both coarse (observed) and fine (unobserved) scales. Populations are estimated on the target grid using generalized linear Poisson regression:

$$Y_i \sim \text{Poisson}(\mu_i), i = 1, \dots, n_p \quad (1)$$

$$\mu_i(\gamma) = \exp(\gamma), i = 1, \dots, n_p \quad (2)$$

where the linear predictor  $\gamma$  for fine-grid level is characterized as follows:

$$\gamma_i = \mathbf{x}'_i \boldsymbol{\beta} + \eta(s_i) + \varphi_{\omega_i}, i = 1, \dots, n_p, \quad (3)$$

Where  $\mathbf{x}'_i$  are indicator vectors of length  $k$ , representing selected land cover covariates for the  $i$  th grid cell (the vector element corresponding to the cell's land cover is 1 and the others are 0) and  $\boldsymbol{\beta}$  is the

corresponding regression coefficient vector. Here, the link function is a log function, which is typical when modeling count data. The spatial noise is modeled as two sets which are all realizations of Gaussian random processes. The first set of spatial random effects  $\eta = (\eta(s_1), \dots, \eta(s_{n_p}))$  characterize spatial autocorrelation at the grid point level in the model that assumes the population in closer grids is more similar than that in grids far apart. These random effects are assumed to be a zero-mean stationary gaussian random field, that is  $\eta \sim N(\mathbf{0}, \Sigma)$ , where  $\Sigma$  is a spatially structured, positive definite variance-covariance matrix that is assumed to follow the Matérn family of covariance functions such that for generic grid point  $s_x$  and  $s_y \in \mathbb{R}^2$ , we have

$$\Sigma_{xy} = \text{Cov}(\eta(s_x), \eta(s_y)) = \frac{\sigma_\eta^2}{2^{\nu-1}\Gamma(\nu)} (\kappa \|s_x - s_y\|)^\nu K_\nu(\kappa \|s_x - s_y\|) \quad (4)$$

where  $\|\cdot\|$  denotes the Euclidean distance between the spatial grid point  $s_x$  and  $s_y$ ,  $\sigma_\eta^2$  is the marginal variance of the process,  $\kappa$  is a scaling parameter that controls the range  $\rho = \left(\frac{\sqrt{8\nu}}{\kappa}\right)$  which is the distance for which spatial correlation is approximately 0.1 (Matérn, 1986).  $K_\nu$  is the modified Bessel function of the second kind, and  $\nu$  is the smoothness parameter that is often set as a constant due to identifiability issues (Abramowitz & Stegun, 1972). Here, we set  $\nu = 1$  following Lindgren et al. (2011).

The second set of areal-level spatial random effects  $\varphi = (\varphi_1, \dots, \varphi_{n_s})$  characterizes spatial autocorrelation in the observed areal data and is assigned a conditional autoregressive (CAR) prior. CAR models consider spatial autocorrelations in terms of adjacency among polygons, rather than distance, which makes them useful in analyzing spatial autocorrelation when observations are taken over irregular spatial units, such as counties or census tracts. We apply the CAR model proposed by Leroux et al. (2000), which found that this CAR model outperformed other models in recent studies (Lee, 2011; Utazi et al., 2019). These spatial random effects are also assumed to follow a zero-mean Gaussian random process  $\varphi \sim N(\mathbf{0}, \sigma_\varphi^2 \mathbf{Q}^{-1}(\mathbf{W}))$ , where  $\mathbf{Q}(\bullet)_{n_s \times n_s}$  is a precision matrix and  $\sigma_\varphi^2$  is the marginal variance parameter of the Gaussian process. Specifically,  $\mathbf{Q}(\mathbf{W}) = \lambda(\text{diag}(\mathbf{W}\mathbf{1}) - \mathbf{W}) + (1 - \lambda)\mathbf{I}_{n_s}$ , where  $\lambda$  is a spatial autocorrelation mixing parameter,  $\mathbf{1}$  is an  $n_s$  vector of 1's,  $\mathbf{I}_{n_s}$  is the identity matrix and  $\mathbf{W}$  is a binary adjacency matrix, which captures the neighborhood information of the areas:  $W_{ij} = 1$  if areas  $\mathcal{A}_i$  and  $\mathcal{A}_j$  are adjacent (share at least one vertex), and  $W_{ij} = 0$  otherwise.  $\varphi_{\mathcal{A}_i}$  is the spatial random effect of the area  $\mathcal{A}_i$  to which the  $i$  th grid cell belongs.

These spatial noise terms can be thought of as random effects influencing population distribution and they can be measured by observations. The population count  $Y_j$  observed in tract  $j$ , must be the sum of all the unobserved counts of population in each grid cell  $i$  inside the tract  $j$ .

$$Y_j = \sum_{i=1}^{N_j} y_i \quad (5)$$

where  $N_j$  is the number of grid cells in tract  $j$ , and  $y_i$  is the unobserved population count for cell  $i$ . Conditional on the underlying latent population surface, the Poisson processes in each pixel are independent. This sum also follows a Poisson distribution with mean equal to the sum of the means of each grid-cell Poisson process, that is,

$$Y_j \sim \text{Poisson}\left(\sum_{i=1}^{N_j} \lambda_i\right), j = 1, \dots, n_A \quad (6)$$

Where  $\lambda_i$  is the mean expected population for cell  $i$  within tract  $j$ , and  $n_A$  is the number of tracts. This allows us to compute the likelihood function for a set of model parameters  $\theta = (\beta, \sigma_\eta^2, \kappa, \sigma_\varphi^2, \rho)$ . The Bayesian model is completed by specifying a set of hyperpriors that characterize the prior

probability distributions for each of these model parameters. The model was implemented in R using the R-INLA (Bakka et al., 2018) and Template Model Builder (TMB) packages (Kristensen et al., 2016), and the parameters were estimated by maximizing the likelihood estimators (MLEs). A statistical review of the TMB package is elaborated in the following section.

### 3.2.2. Scale function to preserve the pycnophylactic property

To preserve the pycnophylactic (the sum of the spatial disaggregated finer grids value equal to the original value in the coarse spatial unit) property of the dasymetric mapping process, a scale function (Eqn. (7)) was applied to the interpolated grids population so that the grid populations sum to the tract population,

$$Pop_i = Pop_j \times \frac{p_i}{\sum_{i \in j} (p_i)} \quad (7)$$

where  $Pop_i$  is the estimated number of persons in the grid  $i$ ,  $Pop_j$  is the observed number of persons in the tract  $j$  that the grid  $i$  resides in, and  $p_i$  is the predicted grids population output from the HPSDRM.

### 3.2.3. HPSDRM model setup

For the HPSDRM model application in the Nashville case study, weak informative hyperpriors were provided and summarized in Table 1. Specifically, penalized complexity (PC) priors were used for the range and scale of the Matérn kernel to regularize the model towards a flatter field with a smaller magnitude (Fuglstad et al., 2019). These PC priors are determined to help avoid the problem of overfitting, simplifying the interpretation of the posterior results (Fuglstad et al., 2019). A negative joint log-likelihood objective function was implemented, based on the theories discussed above. Then, the model was fitted using Bayesian Integrated Nested Laplace Approximation (INLA), using observed tract populations and selected land cover covariates as input data. The SPDE evaluation was performed on a triangular mesh, which is described in the supplemental material. The fitted model yields approximate joint posterior probability distributions for populations in each of the fine grid cells. These posteriors are Gaussian, with the means corresponding to the estimated marginal maximum a posteriori (MMAP) predicted values for the parameters and the joint variance-covariance matrix calculated from the TMB. For all model parameters except the two Matérn kernel hyperparameters, random samples were drawn from the joint posteriors to compare with the priors.

Because the HPSDRM model is intrinsically a Bayesian inference model, it provides full joint-probability distributions for all the parameters, and thus allows easy uncertainty analysis for its predictions. Since each sample from the model's posterior maps to a prediction realization, uncertainties of the model prediction were evaluated by sampling from the posteriors and then calculating the confidence interval at the 0.975 and 0.025 levels from the sampled prediction realizations. Detailed

**Table 1**  
Summary of the (hyper)priors used in the HPSDRM model application in the Nashville case study.

Parameter	Family	Prior parameters
Intercept	Gaussian	mean = 0, sd* = 2
Land Cover (LC) (Developed open space (DOS), Developed Low Intensity (DLI), Developed Medium Intensity (DMI))	Gaussian	mean = 0, sd* = 1
Range ( $r$ )	PC	min* = 1.5, prob* = 0.01
Scale ( $\sigma_\eta$ )	PC	max* = 0.25, prob* = 0.01
Logit (lambda) (logit ( $\lambda$ ))	Gaussian	mean = 0, sd* = 15
Precision ( $\frac{1}{\sigma_\varphi^2}$ )	Gamma	shape = 1, scale = 2

sd\* = standard deviation, min\* = minimum, prob\* = probability, max\* = maximum.

information regarding the HPSDRM model fitted results of the Nashville disaggregated grids population are presented in the Results section.

A series of sensitivity analyses regarding the (hyper)priors for the HPSDRM were conducted, and detailed information is provided in the supplemental material.

### 3.3. Comparative analysis

For the comparative analysis, the disaggregated grids population with a resolution of 150m \* 150m per grid was spatially redistributed from the 2020 Decennial census tract by the four different spatial interpolation models, including Intelligent Dasymetric Mapping (IDM), Areal Weighting Mapping (AWM), Ordinary Least Squares linear regression (OLS). These models are reviewed in the supplementary materials section 4 - 7. Then, the spatially disaggregated grids' population was aggregated into each census block to compare with the true 2020 Decennial census blocks population data. The performance of the models' disaggregation prediction accuracy was evaluated by computing three interpolation accuracy performance metrics: (1)  $R^2$ ; (2) Root Mean Squared Error (RMSE); (3) Mean Absolute Error (MAE). We also calculated the Pearson, Spearman, and Log-Pearson correlation coefficients based on the true block data and the predicted block interpolations for the four dasymetric mapping models. These calculations were implemented in R.

The R code for Moran's I analysis; cross-validated regularized analysis and its results; the simulation study and analysis; and the C++ template code for the TMB, as used in the Nashville case study, are all available on GitHub and in the online supplemental materials.

## 4. Results

The Results section is organized as follows. First, we present the Nashville case study findings. The results of spatial autocorrelation analysis, cross-validated regularized regression analysis, and simulation study analysis are discussed in the [Supplementary Material section 11 - 14](#). Finally, we finish by demonstrating the comparative analysis results.

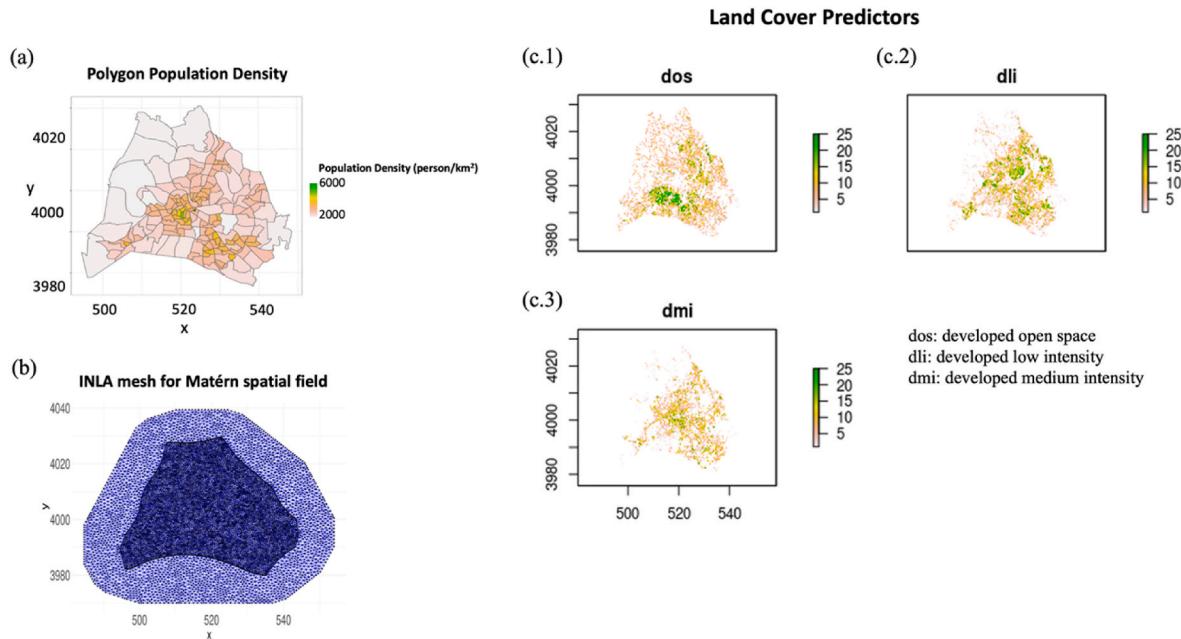
### 4.1. Davidson County, Nashville case study

#### 4.1.1. Hierarchical Poisson spatial disaggregation regression model (HPSDRM)

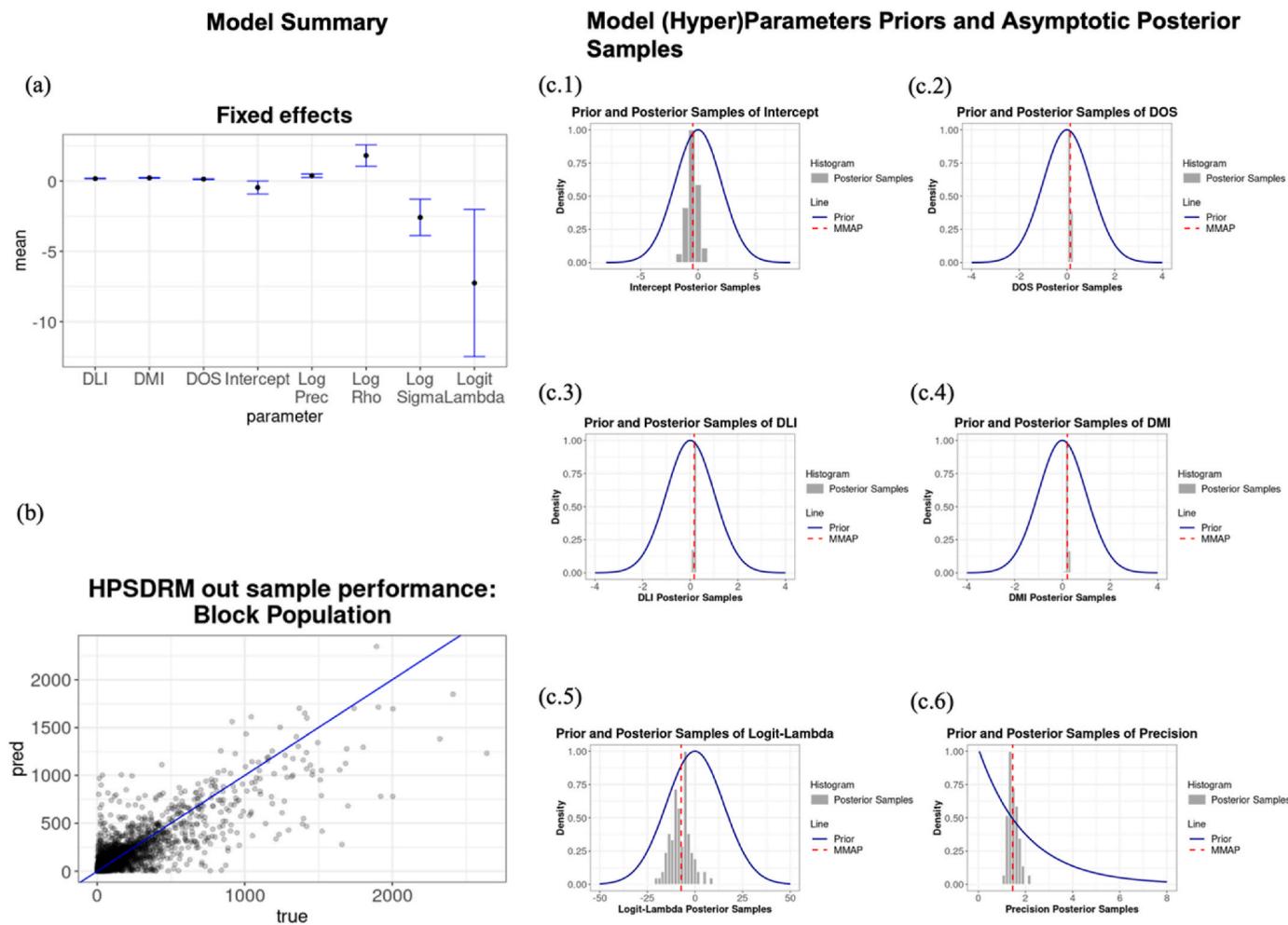
The HPSDRM model input information is displayed in [Fig. 4](#). The polygon response data was obtained from the US decennial census 2020 tract population, and the land cover predictors were retrieved by aggregating the 30m × 30m NLCD 2019 land cover raster onto a 150m × 150m grid ([Fig. 4\(c\)](#)). [Fig. 4\(b\)](#) shows the INLA finite-element mesh used for evaluating the SPDE associated with the Matérn random field. [Fig. 4](#) reveals the land cover covariates numbers that are count number of each land cover grids in the 150m target area.

The summary of the model fitted results are presented in [Fig. 5](#) and [Table 2](#). The MMAP estimates and standard error of the model's fixed effects, including the intercept, three land cover predictors, two hyperparameters for the Matérn field, and two hyperparameters for the CAR field are revealed in [Fig. 5\(a\)](#). [Fig. 5\(a\)](#) suggests that MMAP estimates for the hyperparameters of the random effects generally have greater uncertainties than the MMAP estimates for the coefficients for land cover and the intercepts. [Fig. 5\(c.1\) – \(c.6\)](#) exhibit the priors and posterior samples drawn from the asymptotic normality posterior distribution for all the (hyper)parameters except the range and scale for the Matérn field. The red dotted vertical line is the MMAP estimate for the parameter ([Fig. 5\(c.1\) – \(c.6\)](#)). All parameters' posterior samples were restricted well within the prior, indicating that the prior is non-informative enough without biasing the parameters' inferencing process. The model's out sample performance was evaluated by comparing the true block population and the predicted block population shown in [Fig. 5\(b\)](#). The blue line is the  $y = x$  and the black dots are the true and predicted population value for each block ([Fig. 5\(b\)](#)). A clear positive correlation between the true and predictions suggests that the model successfully interpolated grids population in each tract with good accuracy.

[Table 2](#) summarizes the model fitted results. Specifically, the negative log-likelihood (NLL) is the proxy where TMB adopts to solve the maximum likelihood estimation problem. More detailed information regarding the maximum likelihood estimation approach can be found in the [supplementary material section 9](#) Template Model Builder and section 10 Bayesian inference of Matérn kernel covariance parameters



**Fig. 4.** HPSDRM model input: (a) polygon response data of population density (person/km<sup>2</sup>); (b) INLA mesh for solving SPDE of Matérn spatial field; (c.1) - (c.3) land cover predictors data: (c.1) DOS: developed open space; (c.2) DLI: developed low intensity; (c.3) DMI: developed medium intensity.



**Fig. 5.** Summary of model fitting posteriors: (a) fixed effects parameters; (b) model out sample performance assessed by block population; (c) (hyper)parameters priors and asymptotic normality posterior samples.

**Table 2**  
Summary of model fitted results.

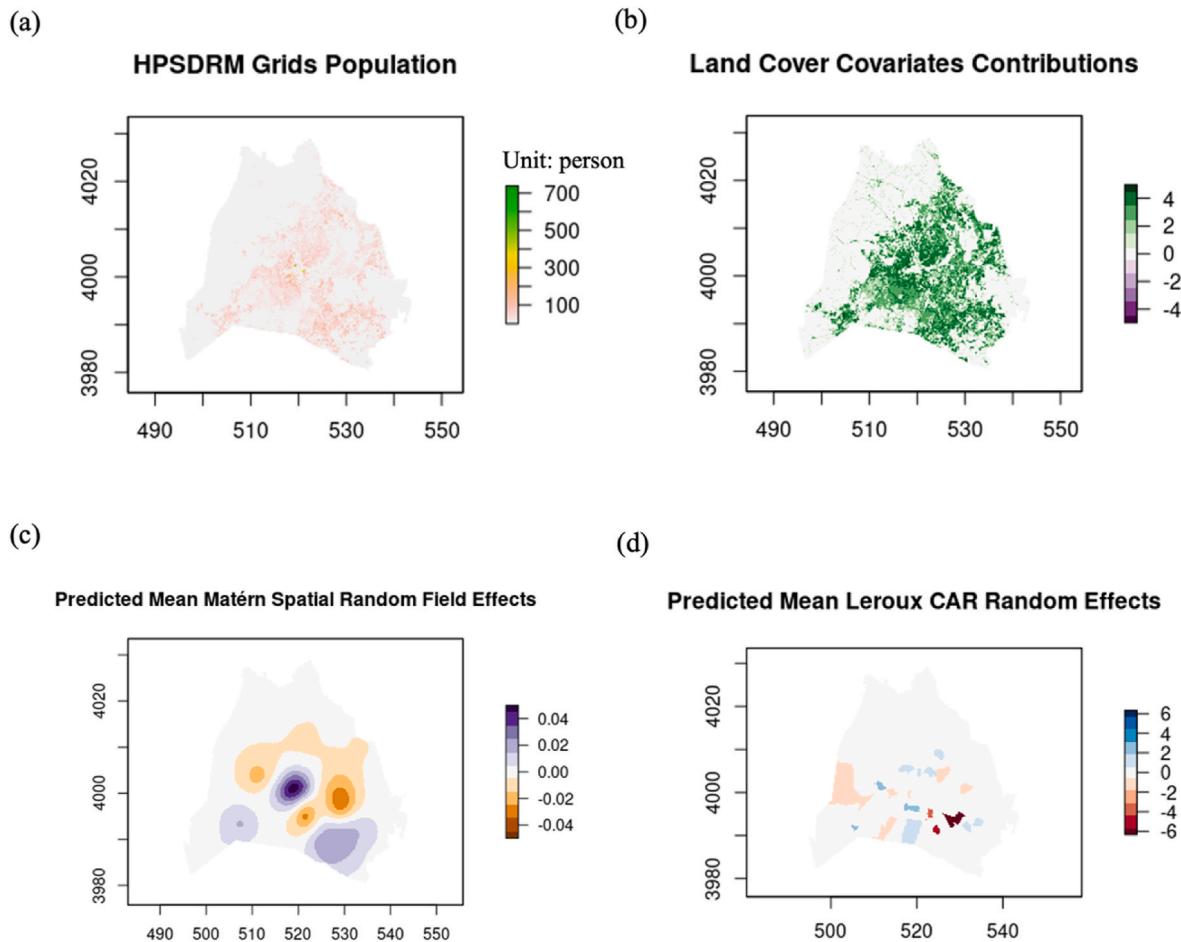
Model Parameters	Mean	Std. Error
Fixed Effects		
Intercept	-0.46	0.47
DOS	0.13	0.03
DLI	0.17	0.02
DMI	0.22	0.03
Log-Sigma	-2.59	1.29
Log-Rho	1.80	0.76
Logit-Lambda	-7.25	5.23
Log-Precision	0.37	0.12
Objective Function-Negative Log Likelihood (NLL)	-11089.45	NA
Priors-Negative Log Likelihood (NLL Priors)	-11967.34	NA
Out Sample (block) Performance	Value	
R <sup>2</sup>	0.70	
RMSE	99.26	
MAE	42.39	
Pearson	0.84	
Spearman	0.75	
Log-Pearson	0.75	

using the Spatial Partial Differential Equation (SPDE) approach.

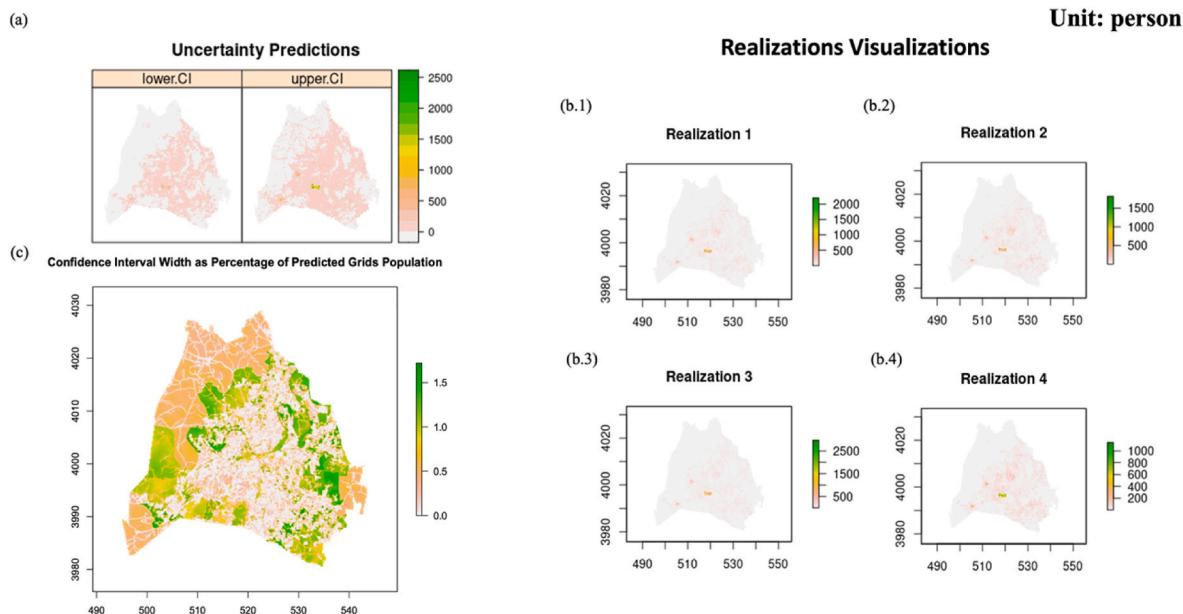
Fig. 6 presents the model prediction results and the decomposition of the latent surface contributions by the land cover covariates, the Matérn random field, and the CAR random field. Fig. 6(a) suggests significant heterogeneities in population distribution within each tract, which are

not apparent when examining the tract data shown in Fig. 4(a). Significant contributions of land cover covariates to the grid predictions can be easily identified in Fig. 6(b). Meanwhile, the contribution of the Matérn random field, which represents spatial autocorrelation at the grid level is small, but clearly visible in Fig. 6(c). This effect accounts for the clustering of population within the downtown and southeastern areas. Fig. 6(c) shows that the population tends to be similar and clustered in the southeastern part of Davidson County compared with Fig. 4(a). Since Fig. 4(a) represent absolute population count, the downtown tracts have a higher population density because of their small tracts area that can be partially explained by the dark blue area in Fig. 6(c). The CAR random effects correspond to spatially discrete effects corresponding to tracts and their neighbors, as opposed to the spatial effects represented by the continuous Matérn field. These effects make larger contributions to tracts surrounding the downtown area and in the south part of the county. In conclusion, the areal level CAR spatial dependence and the grids level spatial autocorrelation are crucial in determining the finer grid population distribution characteristics.

Finally, uncertainties associated with the predictions were computed from total of 1000 model realizations drawn from the predicted distributions, as shown in Fig. 7. Each realization consists of joint samples on the fine-scaled grid cell, where the joint sampling ensures consistency (e.g., pycnophylactic property is preserved). Fig. 7(a) shows the (0.025, 0.975) confidence interval for each grid cell, and Fig. 7(b.1-4) show four examples of prediction realizations. By comparing the two confidence levels predictions in Fig. 7(a), areas with more prediction confidence



**Fig. 6.** HPSDRM model mean prediction decompositions: (a) HPSDRM fine-scale gridded population estimates; (b) land cover covariates contributions; (c) Matérn random field mean contributions; (d) CAR random field mean contributions.

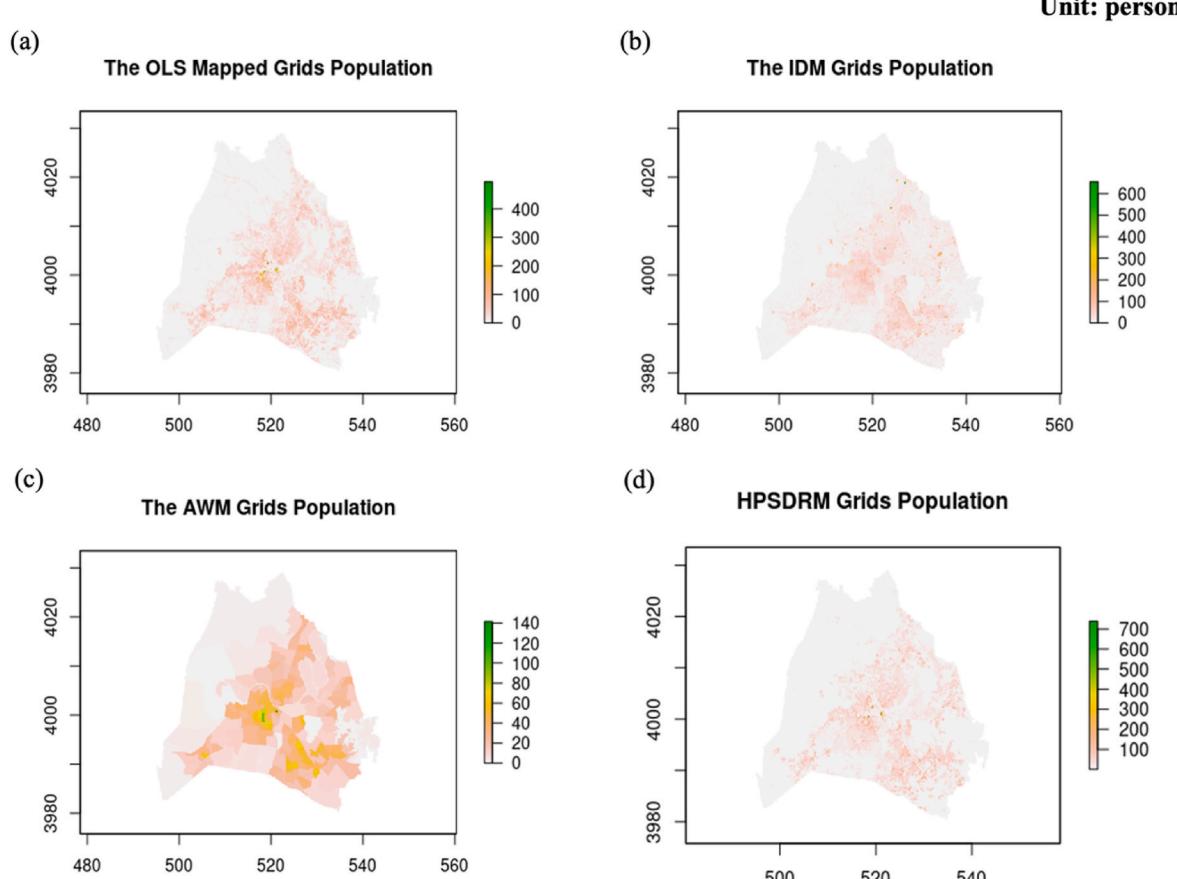


**Fig. 7.** HPSDRM model uncertainty predictions: (a) confidence interval of model predictions; (b) visualization of 4 sampled prediction realizations; (c) Confidence Interval Width as Percentage of Predicted Grids Population.

and uncertainties can be easily identified. In this case, most areas have relatively small disaggregation prediction uncertainties (Fig. 7(a)). Specifically, we computed the confidence interval width as percentage of predicted grids population to show the spatial distribution of interpolation uncertainties shown in Fig. 7(c). Comparing with Figs. 4(a) and 7(c), we see that prediction uncertainties generally are higher in those tracts with smaller population density, which aligns with our interest since model needs to be more accurate in interpolating those tracts with higher population density.

#### 4.2. Comparative analysis

The disaggregated grid population interpolated by the four dasymetric mapping models is displayed in Fig. 8. Comparing the four versions of the high-resolution grids' population distributions, we find that the population interpolated by the OLS linear regression, the IDM, and the HPSDRM all reveal some extents of heterogeneities within the tracts. In contrast, the AWM grid population shows the same pattern as the aggregated tract population shown in Fig. 4(a). This finding illustrates that the AWM model fails to uncover the variations masked by the aggregated tract data, which is the core of the dasymetric mapping task. For the other three interpolated grids population, the hotspots of the distribution differ across the space. Compared to the OLS and the IDM mapped grids population, the HPSDRM mapped grids population exhibits more hot spots with higher values around the downtown areas (Fig. 9), which is contributed by the grids Matérn and areal CAR random effects in the latent field. These hot spots, or spatial clustering pattern can be easily lost in the traditional OLS linear regression model and the IDM sampling process, leading to the underestimation of the population clustering within small areas.



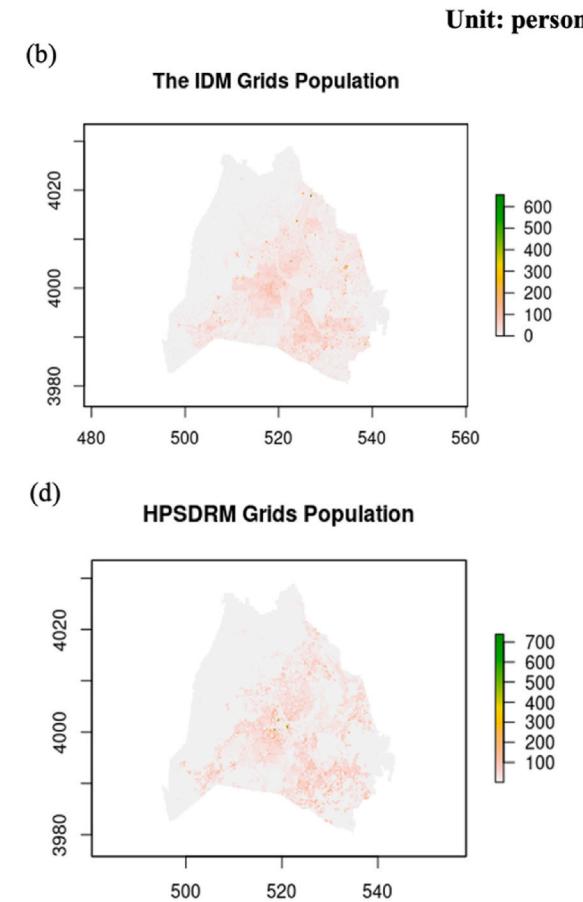
**Fig. 8.** The grids population interpolated by the 4 dasymetric mapping models: (a) OLS mapped grids population; (b) IDM mapped grids population; (c) AWM mapped grids population; (d) HPSDRM mapped grids population.

The out-sample performance was evaluated by comparing the true block population and the aggregated block population from the grids' predictions by the four models (Fig. 10). In Fig. 10, the blue line is the line of  $y = x$  and the colored dots represent the true and aggregated predicted block population values. Meanwhile, the Pearson, Spearman, and Log-Pearson correlation coefficients were calculated for each model's predictions to assess the correlation between the true and predictions (Fig. 10). The HPSDRM results show the best correlation coefficients of the block population out sample assessment, suggesting that it interpolated the grid population with the highest accuracy (Fig. 10(d)).

The three performance metrics show that HPSDRM has considerably better performance than AWM and IDM, and slightly better performance than OLS. Other performance metrics show the advantages of HPSDRM clearly: HPSDRM has a larger  $R^2$  and smaller RMSE and MAE than the other models (Fig. 11, Table 3). This finding confirms our hypothesis that spatial autocorrelation is vital in redistributing the coarser area data to high-resolution grids. The ignorance of the spatial autocorrelation pattern in the model would cause the missing interpretation of the small-area clustering, resulting in lower prediction accuracy. The summary of the performance metrics comparison is shown in Fig. 11 and Table 3.

#### 5. Discussion

IDM modeling has been widely applied for spatial disaggregation. However, we found several ways in which HPSDRM can improve on IDM. First, HPSDRM avoids problems of unbalanced and biased sampling that arise from the arbitrary threshold used in the IDM "percent cover" sampling strategy. Adjusting the threshold produces a tradeoff between two kinds of error: When the threshold is set high, problems can



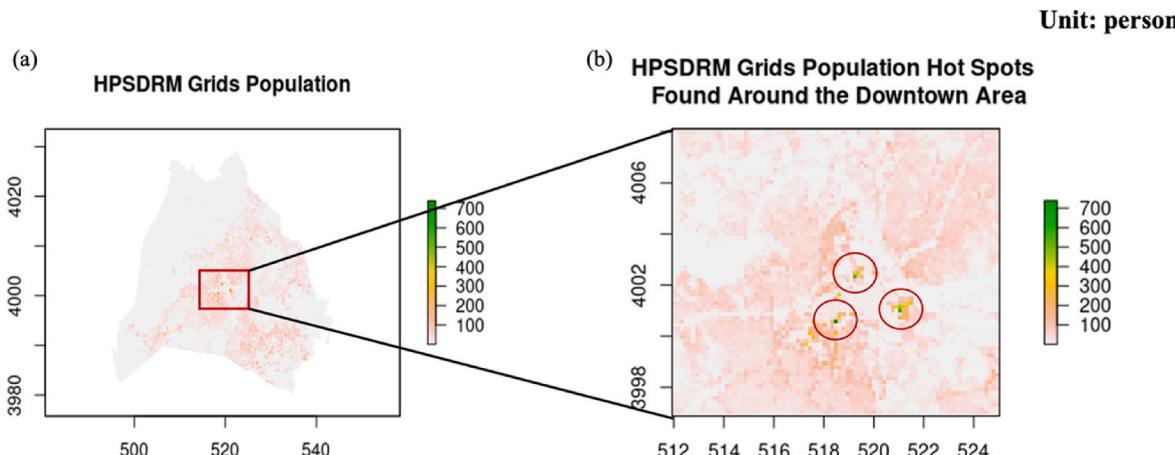


Fig. 9. The hot spots of the HPSDRM grids population identified around the downtown area.

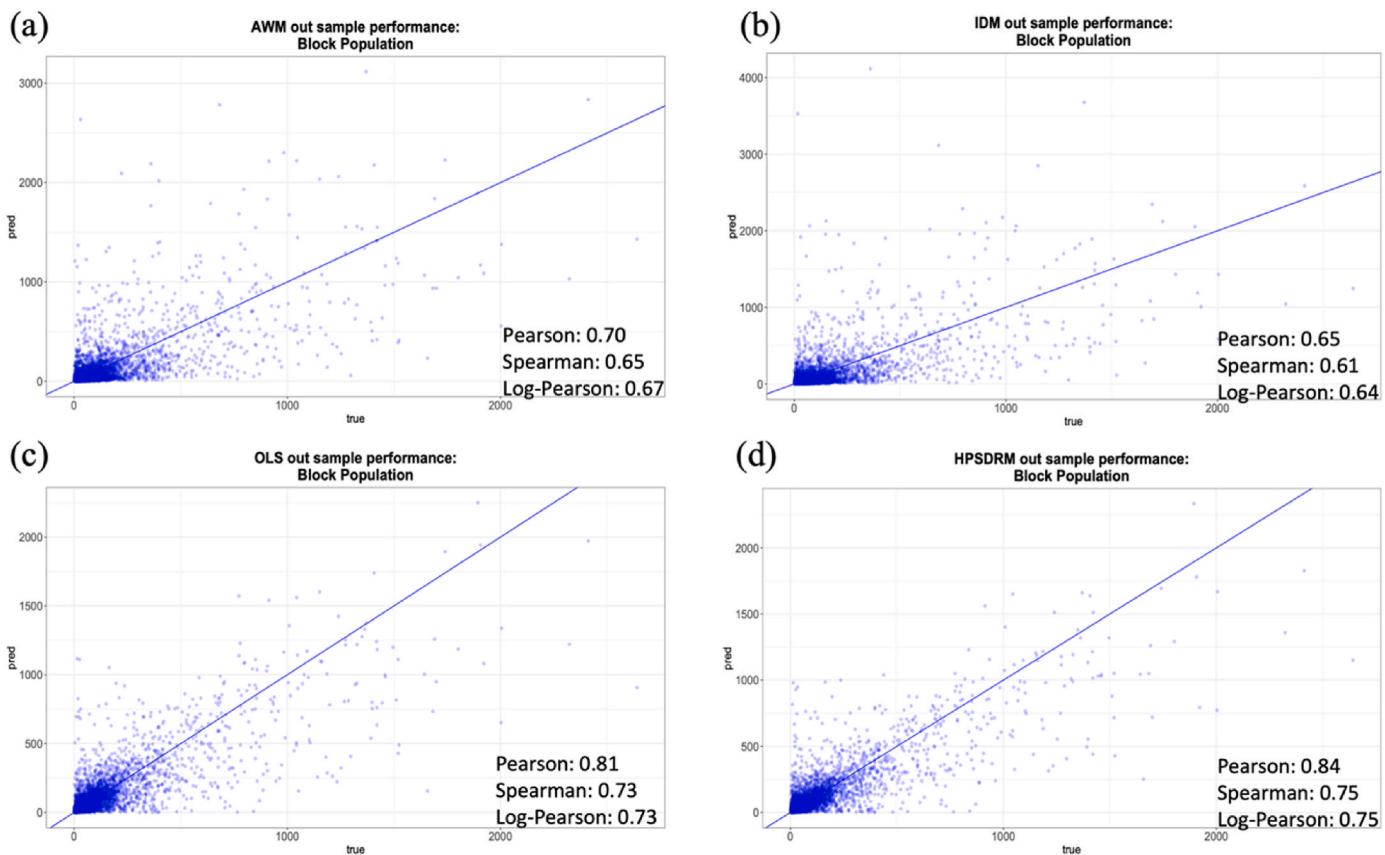


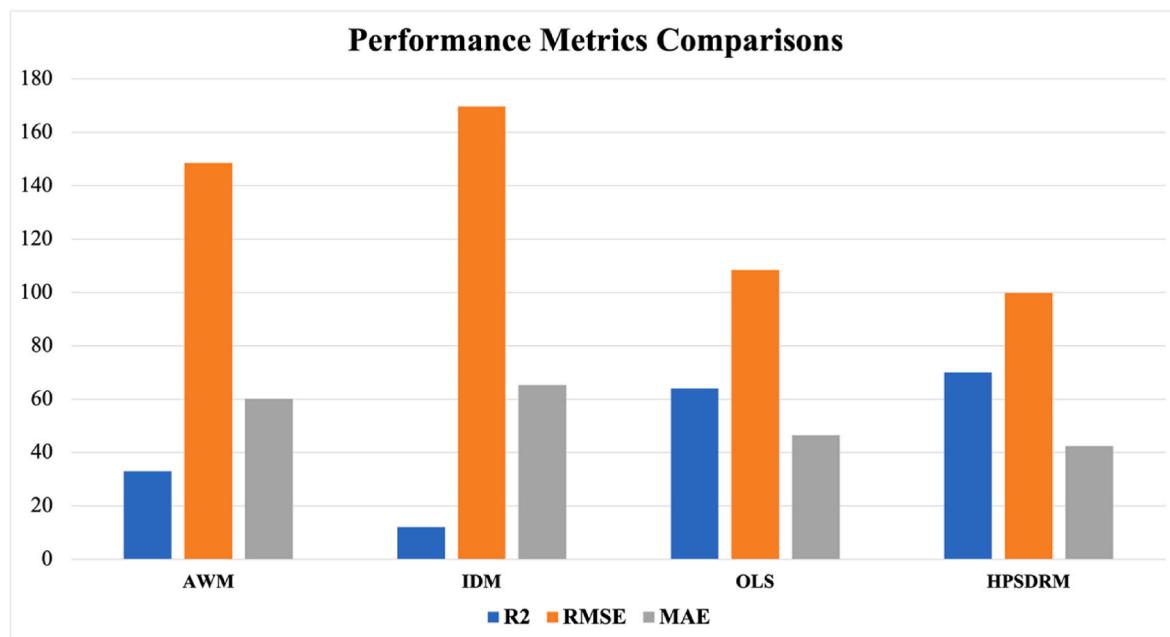
Fig. 10. Models' out sample performance evaluations using the block population data: (a) linear regression model (OLS); (b) IDM; (c) AWM; (d) HPSDRM.

arise from under-sampling ancillary land cover classes because there may not be source zones where the coverage of certain ancillary classes exceeds the threshold; but when the threshold is set low, so more source zones are exceed the threshold coverage for those ancillary land cover classes, the quality of those sampled source zones can become problematic because such a small fraction of the zone has the land cover class in question, so the sampled zones may not truly reflect the density of the ancillary classes (Baynes et al., 2022). As a result, for some places, there may be no value for the threshold that adequately samples many of the ancillary land cover classes, and in these cases, the IDM degenerates into something close to an areal-weighting model.

Another issue is that the IDM process relies on a 'black-box' sampling

design, and this sampling design can play a critical role in determining the accuracy of the model. In other words, a chicken-and-egg problem arises: identifying the best sampling design requires being able to ground-truth the model's predictions, but the value of such a model is interpolating when that fine-scaled ground-truth data is not available. Thus, designing the IDM's best or appropriate sampling strategy is always challenging. Our proposed HPSDRM avoids these issues, because the land cover coefficients are estimated by Bayesian regression, without needing to conduct the kind of sampling that IDM uses.

In comparison to OLS models for population disaggregation, the fundamental problem is that the non-stationarity may exist between model observations and predictions. Specifically, in the spatial



**Fig. 11.** Predictive performance metrics comparisons of the 4 dasymetric mapping models.

**Table 3**

Summary of the predictive performance metrics for the 4 dasymetric mapping models.

Model/Metrics	AWM	IDM	OLS	HPSDRM
R <sup>2</sup> (%)	33	12	64	70
RMSE	148.58	169.73	108.41	99.26
MAE	60.13	65.34	46.58	42.39

disaggregation regression modeling process, the model calibration based on coarser observations may not be able to accurately reveal the relationship between covariates and the target variable in a finer spatial scale, leading to misinterpretations in a finer spatial scale. Thus, the non-stationarity of model coefficients between coarser and finer spatial scales cannot be ignored, which is a vulnerability of linear regression model in spatial disaggregation regression application scenarios.

We address spatially stratified heterogeneity through the intrinsic architecture of our model rather than relying on stratified sampling techniques. In particular, our Hierarchical Poisson Spatial Disaggregation Regression Model (HPSDRM) is designed with a multilevel or hierarchical structure that includes both a Matérn kernel and Conditional Autoregressive Models (CAR). This setup is specifically engineered to handle spatial variability and dependence across various spatial layers, such as grids or census tracts. By integrating these components, our model can capture distinct spatial dependencies in regression analyses, effectively adjusting for spatial heterogeneity by modeling the diverse relationships between neighboring areas. Meanwhile, we employed the *q*-statistic measurement to assess the spatial stratified heterogeneity (SSH) of our spatially disaggregated population product (Wang et al., 2016). Based on the realization by the census tract, which is the major determinant of the observations of population, we found the *q*-statistic of the dasymetric grid population of the census tract to be 0.68, suggesting a strong spatial stratification of heterogeneity was detected from our dasymetric grid population product. The detailed information regarding the *q*-statistic measurement is included in supplementary material section 18: Spatial Stratification of Heterogeneity (SSH) *q*-statistic.

However, there are limitations in this work regarding the understanding of the relationship between the model's interpolation

performance and its application schemes, including disaggregation ratio, response observation data structure, and the population characteristics of the disaggregation task area. Specifically, to assess the generalizability of the HPSDRM, it is crucial to examine how well the findings might apply beyond the specific conditions and data of the case study employed here. Here, we consider three critical factors that include population characteristics, sample conditions, and inference methods.

We believe the study population in Davidson County, Nashville, is sufficiently diverse to reflect the broader target demographic. We intentionally did not exclude any subgroups to maintain a diverse and comprehensive representation, avoiding over-specificity to any particular region or demographic. Additionally, our study encompasses a variety of environmental, social, and economic conditions, as it includes the entire Nashville metropolitan area. This broad coverage supports our confidence that our methods are adaptable to various population characteristics. Meanwhile, we conducted a generalizability analysis to apply the HPSDRM on other three areas that include NYC, NY, Atlanta, Georgia, and Denver, Colorado to evaluate the advancements of our proposed approach in terms of predictive capabilities. The results of generalizability analysis suggest that our HPSDRM approach exceeds the other traditional spatial disaggregation interpolation models not only in the Nashville metropolitan area, but in various application scheme with different extents of population autocorrelation mechanisms.

In our methodology, we did not employ stratified or random sampling, ensuring that all subgroups within the population are represented. This enhances our ability to generalize our findings across a wider population spectrum. Furthermore, we implemented cross-validation techniques to verify the relationship between land cover, land use, and population distributions, ensuring the model's effectiveness across different data subsets within the area.

Our overarching goal is to evaluate the model's robustness by applying it to different datasets and observing whether it yields consistent results across various geographic areas with differing land cover and land use patterns as well as application schemes. Based on these considerations, we conducted a generalizability analysis by comparing the proposed HPSDRM model's predictive capability with the other three traditional dasymetric mapping models on three other areas with various population distribution characteristics that include NYC, NY, Atlanta, Georgia, and Denver, Colorado. The generalizability

analysis results also echo our findings in this study that HPSDRM surpass the other three traditional dasymetric mapping approaches on all considered areas in terms of spatial disaggregation interpolation capabilities. The generalizability analysis results are included in the supplementary material generalizability analysis section.

Additionally, we envisage future efforts in several directions. Here, we focused on modeling the Poisson count responses. Nonetheless, the model can be easily extensible to other distributions like binomial responses for modeling other health and socioeconomic indicators. Additionally, we only examined a single disaggregation ratio when applying the model to implement the dasymetric mapping task for the simulation study and the case study area. More disaggregation schemes with various downscaling ratios could be tested under varying numbers of data points, sizes of aggregated regions, and levels of model complexity. Meanwhile, we only applied this method to the Metropolitan Nashville-Davidson County area. Extensions of the model applied to other sites with various population distribution patterns also need to be considered. For example, we expect a greater improvement regarding the prediction accuracy of grids' population surface when applying the HPSDRM model to more urban areas like New York City since stronger spatial autocorrelation signals can be detected and explained. We have verified this assumption by conducting the generalizability analysis, and the results are included in the supplementary material generalizability analysis results section. Further, we identified the significant land-cover classes to use in the HPSDRM model by applying cross-validated regression to finer-scale data at the block level. If the finer-scaled data are not available for a different region of interest, it may not be possible to identify the relevant land-cover classes for that region. However, that does not mean the HPSDRM developed in this study cannot be applied in this scenario since the most significant land cover types can be identified based on the most information available, and the HPSDRM can interpolate finer-resolution values based on much coarser observations while incorporating multiple level of spatial autocorrelation.

Meanwhile, the accuracy of predictions made by the HPSDRM is contingent upon various factors, including the inherent characteristics of the population under study, the conditions under which sampling is conducted, and the availability and quality of ancillary information resources utilized. Given these considerations, it becomes imperative to conduct comprehensive evaluations of the HPSDRM's generalizability and to compare its performance with that of other dasymetric mapping models. Therefore, it is recommended that future research endeavors dedicate significant efforts towards thoroughly examining the advantages and limitations of the HPSDRM within its established framework. Based on these considerations, we propose several methods to improve the model's generalizability that include conducting external validations, sensitivity analysis, and cross-validation approaches. Since the sensitivity analysis and cross-validation approach were employed in this study, external validations will be the focus of future research.

In summary, we incorporate spatial autocorrelation in the spatial disaggregation regression modeling to better understand finer grids' population distribution within tracts in this study. The proposed HPSDRM in this study arguably increases the accuracy of spatially disaggregated grid population distribution and provides a step forward towards better identifying the hot spots and cold spots variations within aggregated tract observations. More importantly, the application scenario of such model can certainly extend beyond population disaggregation interpolation. The potential uses of this model in practical applications include emergency management planning, healthcare resource allocation estimations, urban planning in deciding the locations of public services, and interpolation of disease cases or rates where the data are suppressed due to privacy concern, etc. The more accurate estimates of the hot spots and clusters of the values in small areas make these applications more efficient. For instance, a better estimation of the vulnerable population counts in the downtown area facilitates allocating more resources to that area. Meanwhile, the HPSDRM only requires the land cover information as ancillary data. The U.S.-wide coverage of the

NLCD product provides the necessary framework to expand the HPSDRM outlined in this paper to future applications at a national scale.

## 6. Conclusions

This study introduces a Hierarchical Poisson Spatial Disaggregation Regression Model (HPSDRM) that integrates a Matérn kernel with conditional autoregression (CAR) to handle multiple layers of spatial autocorrelation, based on Bayesian inference principles. A thorough sensitivity analysis was conducted to optimize the HPSDRM's hyperparameters, enhancing its effectiveness and robustness. We tested the model via simulations on a square test area, showcasing its ability to accurately disaggregate coarse spatial data into finer grids. Results indicated high R<sup>2</sup> values between 0.93 and 0.99, confirming the model's consistent performance in explaining variance. The HPSDRM was also applied to real-world data in Davidson County, outperforming traditional dasymetric mapping methods by at least 10% in prediction accuracy, thus proving its potential in spatial data disaggregation and interpolation. The model efficiently handles spatial autocorrelation using fewer predictors and relies minimally on external data, facilitating its broad implementation using easily accessible national land cover data.

## Funding

This work was supported by the U.S. Department of Transportation Maritime Transportation Research and Education Center (MarTREC).

## R code repository

Github: <https://github.com/bowenhesteven/HPSDRM>.

## CRediT authorship contribution statement

**Bowen He:** Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jonathan M. Gilligan:** Writing – review & editing, Visualization, Validation, Supervision, Project administration, Methodology, Funding acquisition. **Janey V. Camp:** Writing – review & editing, Project administration, Funding acquisition.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.apgeog.2024.103333>.

## References

- Abramowitz, M., & Stegun, I. (1972). *Handbook of mathematical functions*. New York: Courier Dover Publications.
- Alahmadi, M., Atkinson, P. M., & Martin, D. (2014). A comparison of small-area population estimation techniques using built-area and height data, Riyadh, Saudi Arabia. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(5), 1959–1969.
- Anderson, J. R., Hardy, E. E., Roach, J. T., & Witmer, R. E. (1976). A land use and land cover classification system for use with remote sensor data. *USGS Professional Paper 964. A revision of the land use classification system as presented in the USGS Circular 671*.
- Bakka, H., Rue, H., Fuglstad, G. A., Riebler, A., Bolin, D., Krainski, E., Simpson, D., & Lindgren, F. (2018). Spatial modelling with R-INLA: A review. *Invited extended review, arxiv:1802.06350*.
- Baynes, J., Neale, A., & Hultgren, T. (2022). Improving intelligent dasymetric mapping population density estimates at 30 m resolution for the conterminous United States by excluding uninhabited areas. *Earth System Science Data*, 14(6), 2833–2849.

- Briggs, D. J., Gulliver, J., Fecht, D., & Vienneau, D. M. (2007). Dasymetric modelling of small-area population distribution using land cover and light emissions data. *Remote Sensing of Environment*, 108(4), 451–466.
- Camp, J., Gilligan, J., & He, B. (2023). The unintended consequences of flood mitigation along inland waterways—a look at resilience and social vulnerabilities through a case study analysis. *Final report submitted to MarTREC US DOT university transportation center*.
- Cockx, K., & Canters, F. (2015). Incorporating spatial non-stationarity to improve dasymetric mapping of population. *Applied Geography*, 63, 220–230.
- Cromley, R. G., Hanink, D. M., & Bentley, G. C. (2012). A quantile regression approach to areal interpolation. *Annals of the Association of American Geographers*, 102(4), 763–777.
- Dewitz, J., & Geological Survey, U. S. (2021). National land cover database (NLCD) 2019 products (ver. 2.0, june 2021. U.S. Geological Survey data release. <https://doi.org/10.5066/P9KZCM54>
- Fuglstad, G. A., Simpson, D., Lindgren, F., & Rue, H. (2019). Constructing priors that penalize the complexity of Gaussian random fields. *Journal of the American Statistical Association*, 114(525), 445–452.
- Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.
- Giordano, A., & Cheever, L. (2010). Using dasymetric mapping to identify communities at risk from hazardous waste generation in San Antonio, Texas. *Urban Geography*, 31(5), 623–647.
- He, B. (2023). *Efficient computational evaluation tools to accelerate the planning of vulnerability, resilience, and sustainability of the social-environmental systems in the city of Nashville* (doctoral dissertation).
- He, B., & Ding, K. J. (2021). Localize the impact of global greenhouse gases emissions under an uncertain future: A case study in western cape, South Africa. *Earth*, 2(1), 111–123.
- He, B., & Ding, K. J. (2023). Global greenhouse gases emissions effect on extreme events under an uncertain future: A case study in western cape, South Africa. *PLOS Climate*, 2(1), Article e0000107.
- He, B., Gilligan, J. M., & Camp, J. V. (2023a). An index of social fabric for assessing community vulnerability to natural hazards: Model development and analysis of uncertainty and sensitivity. *International Journal of Disaster Risk Reduction*, 96, Article 103913.
- He, B., & Guan, Q. (2021a). A mathematical approach to improving the representation of surface water-groundwater exchange in the hyporheic zone. *Journal of Water and Climate Change*, 12(5), 1788–1801.
- He, B., & Guan, Q. (2021b). A risk and decision analysis framework to evaluate future PM2. 5 risk: A case study in los angeles-long beach metro area. *International Journal of Environmental Research and Public Health*, 18(9), 4905.
- He, B., & Guan, Q. (2022a). Analysis and prediction of the correlation between environmental ecology and future global climate change. *Journal of HFUT: Nature and Science*, 6(45), 818–824.
- He, B., & Guan, Q. (2022b). The statistical analysis and prediction associated with nuclear meltdown accidents risk evaluation. *International Journal of Nuclear Safety and Security*, 1(2), 104–123.
- He, B., & Guan, Q. (2024a). Investigating the effects of spatial scales on social vulnerability index: A hybrid uncertainty and sensitivity analysis approach combined with remote sensing land cover data. *Risk Analysis*.
- He, B., Zheng, H., & Guan, Q. (2023b). Evaluation of future-integrated urban water management using a risk and decision analysis framework: A case study in denver-Colorado metro area (dcma). *Water*, 15(22), 4020.
- He, B., Zheng, H., & Guan, Q. (2024b). Toward revolutionizing water-energy-food nexus composite index model: From availability, accessibility, and governance. *Frontiers in Water*, 6, Article 1338534.
- Jia, P., & Gaughan, A. E. (2016). Dasymetric modeling: A hybrid approach using land cover and tax parcel data for mapping population in alachua county, Florida. *Applied Geography*, 66, 100–108.
- Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. M. (2016). Tmb: Automatic differentiation and Laplace approximation. *Journal of Statistical Software*, 70(5), 1–21. <https://doi.org/10.18637/jss.v070.i05>
- Langford, M. (2007). Rapid facilitation of dasymetric-based population interpolation by means of raster pixel maps. *Computers, Environment and Urban Systems*, 31(1), 19–32.
- Lee, D. (2011). A comparison of conditional autoregressive models used in Bayesian disease mapping. *Spatial and spatio-temporal epidemiology*, 2(2), 79–89.
- Leroux, B. G., Lei, X., & Breslow, N. (2000). Estimation of disease rates in small areas: A new mixed model for spatial dependence. In *Statistical models in epidemiology, the environment, and clinical trials* (pp. 179–191). New York, NY: Springer.
- Li, T., & Corcoran, J. (2011). Testing dasymetric techniques to spatially disaggregate the regional population forecasts for Southeast Queensland. *Journal of Spatial Science*, 56(2), 203–221.
- Lin, J., Cromley, R., & Zhang, C. (2011). Using geographically weighted regression to solve the areal interpolation problem. *Annals of GIS*, 17(1), 1–14.
- Lindgren, F., Rue, H., & Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B*, 73(4), 423–498.
- Maantay, J. A., Maroko, A. R., & Herrmann, C. (2007). Mapping population distribution in the urban environment: The cadastral-based expert dasymetric system (CEDS). *Cartography and Geographic Information Science*, 34(2), 77–102.
- Matérn, B. (1986). *Spatial variation*. Berlin: Springer.
- Mennis, J. (2003). Generating surface models of population using dasymetric mapping. *The Professional Geographer*, 55(1), 31–42.
- Mennis, J. (2009). Dasymetric mapping for small area population estimation. *Geography Compass*, 3, 727–745.
- Mennis, J., & Hultgren, T. (2006). Intelligent dasymetric mapping and its application to areal interpolation. *Cartography and Geographic Information Science*, 33(3), 179–194.
- Nelson, K. S., Abkowitz, M. D., & Camp, J. V. (2015). A method for creating high resolution maps of social vulnerability in the context of environmental hazards. *Applied Geography*, 63, 89–100.
- Openshaw, S. (1983). The modifiable areal unit problem. In *Concepts and techniques in modern geography*, 38. Norwich: Geobooks.
- Petrov, A. (2012). One hundred years of dasymetric mapping: Back to the origin. *The Cartographic Journal*, 49(3), 256–264.
- Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). Disaggregating census data for population mapping using random forests with remotely sensed and ancillary data. *PLoS One*, 10(2), Article e0107042.
- Su, M. D., Lin, M. C., Hsieh, H. I., Tsai, B. W., & Lin, C. H. (2010). Multi-layer multi-class dasymetric mapping to estimate population distribution. *Science of the Total Environment*, 408(20), 4807–4816.
- United States Census Bureau. (2021). *Population estimates, july 1, 2021 (V2021) – Davidson county, TN*. QuickFacts. Retrieved from <https://www.census.gov/quickfacts/fact/table/davidsoncountytennesseee>
- Utazi, C. E., Thorley, J., Alegana, V. A., Ferrari, M. J., Nilsen, K., Takahashi, S., Metcalf, C. J. E., Lessler, J., & Tatem, A. J. (2019). A spatial regression model for the disaggregation of areal unit based data to high-resolution grids with application to vaccination coverage mapping. *Statistical Methods in Medical Research*, 28(10–11), 3226–3241.
- Wang, J. F., Zhang, T. L., & Fu, B. J. (2016). A measure of spatial stratified heterogeneity. *Ecological Indicators*, 67, 250–256.
- Woltman, H., Feldstain, A., MacKay, J. C., & Rocchi, M. (2012). An introduction to hierarchical linear modeling. *Tutorials in quantitative methods for psychology*, 8(1), 52–69.
- Wu, C., & Murray, A. T. (2005). A cokriging method for estimating population density in urban areas. *Computers, Environment and Urban Systems*, 29, 558–579.
- Zandbergen, P. A. (2011). Dasymetric mapping using high resolution address point datasets. *Transactions in GIS*, 15, 5–27.