The sociospatial factors of death: Analyzing effects of geospatially-distributed variables in a Bayesian mortality model for Hong Kong

Thayer Alshaabi,[1, 2, *] David Rushing Dewhurst,[1, 2, 3] James P. Bagrow,[1, 4]
Peter Sheridan Dodds,[1, 2, 4] and Christopher M. Danforth[1, 2, 4, †]

[1]*Vermont Complex Systems Center, The University of Vermont, Burlington, VT 05405.*
[2]*Computational Story Lab, The University of Vermont, Burlington, VT 05405*
[3]*MassMutual Data Science, Boston, MA 02110*
[4]*Department of Mathematics & Statistics, The University of Vermont, Burlington, VT 05405*

(Dated: June 16, 2020)

Human mortality is in part a function of multiple socioeconomic factors that differ both spatially and temporally. Adjusting for other covariates, the human lifespan is positively associated with household wealth. However, the extent to which mortality in a geographical region is a function of socioeconomic factors in both that region and its neighbors is unclear. There is also little information on the temporal components of this relationship. Using the districts of Hong Kong over multiple census years as a case study, we demonstrate that there are differences in how wealth indicator variables are associated with longevity in (a) areas that are affluent but neighbored by socially deprived districts versus (b) wealthy areas surrounded by similarly wealthy districts. We also show that the inclusion of spatially-distributed variables reduces uncertainty in mortality rate predictions in each census year when compared with a baseline model. Our results suggest that geographic mortality models should incorporate nonlocal information (e.g., spatial neighbors) to lower the variance of their mortality estimates, and point to a more in-depth analysis of sociospatial spillover effects on mortality rates.

I. INTRODUCTION

A. Mortality risks and social deprivation

Temporal dynamics of mortality risks with regards to nation-wide epidemics [1, 2], pollution [3, 4], and life quality and expectancy [5] have been widely studied throughout the last decade in Hong Kong and around the world. Researchers have hypothesized and identified many connections between longevity, social deprivation, and socioeconomic inequality. Studies have put forward evidence of disparities in mortality risks all around the globe [6–8].

Notably, there are many different definitions of social deprivation. Messer *et al.*'s study [9] provides a well-written overview of the literature regarding neighborhood-level socioeconomic deprivation. The study not only discusses ways in which deprivation has been defined throughout the last decade of research, but it also proposes a new method to calculate and hopefully standardize what they call a "neighborhood deprivation index" (NDI). The authors highlight the limitations of definitions found in the literature. Using principal components analysis (PCA) on census data from 1995 to 2001, they demonstrate the effectiveness of their proposed measurement at capturing socioeconomically deprived areas across the US.

In the present study, we use a set of social socioeconomic attributes including income, unemployment, access to life insurance, and mobility, to define and capture the magnitude of social deprivation compared to an average baseline for Hong Kong.

The notion of disparity in health and mortality risks is often studied using population-scale inputs such as air pollution, as well as sensitive individual variables such as age, race, and gender. Ou *et al.* [10] infer socioeconomic status by type of housing, education, and occupation. They find that regions with lower socioeconomic status have a higher ratio of air pollution when they look at the spatial correlation of groups living in public rental housing compared to areas with a higher density of private housing. They also report that areas with a higher density of blue-collar workers have a higher air pollution-associated mortality than others.

Chung *et al.* [11] show evidence of disparities in mortality risks when considering age as a control variable. The authors investigate the effect of socioeconomic status concerning the rapid economic development of Hong Kong from 1976 to 2010. They find a decline in socioeconomic disparity in mortality risks across the country throughout that period. Mortality rates are generally higher for lower socioeconomic status regions regardless of gender. They also show that various health benefits brought by economic growth are greater for regions with higher socioeconomic status. The market share of health benefits is unequally distributed among groups of varying status: Individuals with higher socioeconomic status have access to greater benefits than those of lower socioeconomic status.

———

[*] thayer.alshaabi@uvm.edu

[†] chris.danforth@uvm.edu

Typeset by REVTₑX

B. Spatial association of mortality risks

The spatial associations between income inequality and health risks are widely understood both internationally [12–16], and for individual cities and states [17, 18]. Although many studies have investigated income inequality [19–22] and compared that across countries [23], most studies use a spatially localized approach in their investigations. Nonetheless, studies have shown the importance of spatial associations in identifying relations between socioeconomic deprivation and longevity.

Geographically weighted regression (GWR) [24, 25] is a commonly used approach designed to account for spatial associations. Local attributes can play a strong role in the model dynamics given the assumption that socioeconomic factors vary geographically, and GWR was introduced to overcome the challenge of uncertainty in a global (e.g., average) spatial regression model. The authors argue that socioeconomic attributes are intrinsically intra-connected over space due to the mechanisms by which communities evolve. Their study makes an empirical comparison of their method to other stationary regression models to investigate the spatial distribution of long-term illness on a data set published by the UK Census of Population. They show that GWR can account for the spatial nonstationarity between different variables in the data.

Researchers have looked into the spatial association between air pollution and mortality in Hong Kong [26], Czech [27], Rome [28], and France [29]. Cossman et al. [30] examine the spatial distribution of mortality rates over 35 years starting from 1968 to 2002 across all counties in the US. Unlike previous studies, the Cossman et al. study highlights a nonrandom pattern of clustering in mortality rates in the US where high mortality rates are primarily driven by economic decline. However, counties in the middle-upper plain of the US show an interestingly unique pattern. Although these counties have followed a similar economic decline and an increasing rate of out-migration, they still have the lowest mortality rates across the country. By contrast, the southeastern counties have the highest mortality rates despite having similar rates of economic decline and migration.

To assess geospatial associations between pollution and mortality in Hong Kong, Thach et al. [31] examine the spatial interactions of tertiary planning units (TPUs) [32]—similar to census-blocks in the US. The authors indicate a positive spatial correlation between mortality rates and seasonal thermal changes across the region. They argue that the variation between TPUs is a key factor for cause-specific mortality rates. Their results show that socioeconomically deprived regions have a higher rate of mortality, especially during the winter season.

C. Sociospatial factors of death

Studying the relative spatial interactions of social and economic indicators is certainly not a new field. Investigations of such relations can be noted throughout the literature measuring nonlocal and/or interdependent interactions of inequality in health care [33], peer pressure in education [34, 35], and decision-making [36, 37].

Yang et al. [38] argue that mortality rates of counties in the US are highly associated with social and economical features found in neighboring counties. Their findings suggest that mortality rates in a given county are highly driven by social indicators from bordering counties due to the spillover of socioeconomic wealth or social deprivation across neighborhoods.

Another recent study by Holtz et al. [39] highlights the significant influence of nonlocal interactions such as peer effect and spillovers on regional policies regarding the global outbreak of COVID–19.

To identify and examine broader dimensions of inequality from a spatial point of view, various studies have either investigated the geographical effect of neighborhoods [40], or proposed new methods [41] similar to Moran's I [42] and spatial auto-regression. Using a network-based approach to explore the dynamics of communities and their impact on mortality risks, we present here a small-case study using a collection of datasets from Hong Kong.

Though Hong Kong is a small island territory, it exhibits significant variation in occupations, income, foreign inhabitant density, and residence status of workers. We demonstrate heterogeneity of such exogenous features and investigate nonlocal behavioral interactions of wealth and social deprivation across neighborhoods. We present a statistical analysis comparing local and nonlocal models to show the importance of spatial associations for mortality modeling. In particular, we use a spatial network approach to study a socioeconomic peer-pressure effect among communities. For instance, we investigate how the magnitude of a socially deprived area can consequently have a nonlocal effect on its neighbors' mortality risks. Similarly, we examine how the spatial spread of wealth of an affluent area can spillover to its surrounding areas, and thus impact their longevity. Our work not only shows the deep influence of these spatial interactions of neighborhoods on predicting mortality rates, but also provides a method for investigating systematic inference errors of mortality models.

II. METHODS

We introduce and analyze the datasets in Sec. II A. Detailed descriptive analysis of our data set variables can be found in Appendix A. For our statistical analysis, we employ a set of simple Bayesian generalized additive models to predict mortality rates across districts in Hong Kong. We use three different models to illustrate the role

of spatial associations by comparing models with spatial features to a baseline model without spatial features. First, we present our local model that does not use any spatial information in Sec. II B 1. We compare our Baseline model to two nonlocal spatial models in Sec. II B 2. Our first nonlocal model uses spatial features from the nearest neighbours, while the second uses features from all neighbours weighted by their distance to the target area. We will refer to the nonlocal models as SP, and WSP respectively.

A. Data sources

Mortality data: We use the official counts of known and registered deaths provided by the Census and Statistics Department of Hong Kong [46]. The data set contains 892,055 death records between 1995 and 2017. Every record includes a wide range of information such as age, gender, and place of death (TPU) [32]—a geospatial reference system used to report population census statistics.

Life insurance data: We have obtained data from MassMutual Asia (MMA)—the current branch of MassMutual Inc. in Hong Kong. The data set has a total of 472,922 life insurance policies issued between 1995 and 2018. According to the Hong Kong Insurance Authority [47], MMA had roughly 2.5% market share of all non-linked individual life insurance policies issued in Hong Kong in 2016. For each policy, we have a few attributes including age, gender, and a TPU code. These records will be aggregated at the district level, then normalized by the population size at each time snapshot to report the proportion of individuals insured by each district.

Census data: We collected socioeconomic variables curated by the Census and Statistics Department of Hong Kong [48]. We have three snapshots at 5-year intervals, 2006, 2011, and 2016. For each year, the data set includes the total population density by district, median income, median rent to income ratio, median monthly household income, unemployment rates, unemployment rates across households, unemployment rates among minorities, the proportion of homeless people, the proportion of homeless mobile residents, the proportion of single parents, the proportion of households with kids in school, the proportion of households with children (aged under 15), and the proportion of households with elderly (aged 65 or above).

Geospatial unit: Initially, we planned to use TPU as the main geospatial unit to combine our data sets spatially. Unfortunately, most death records are found in a only small fraction of TPUs likely due to data privacy concerns—records in small TPUs may reveal sensitive information about specific individuals there. Given that the mortality data set has a large fraction of missing TPUs, and to avoid any risk of identifying individuals in the data set, we use districts as our main spatial unit of analysis [43]. This choice is consistent with prior work, where most studies have either filtered out small TPUs

in their analyses [11, 26], or aggregated their records at the district level [31] to overcome this challenge.

Categorization: We organize our features into three different categories.

1. **Base**: This set has most of the socioeconomic features in our data sets such as population density, unemployment rates, the proportion of homeless people, mobile residents, and single-parent households. However, we do not include wealth-, age-, or race-related features here.

2. **Wealth**: In addition to the base features described above, we include median income, median rent-to-income ratio, median monthly household income, and life insurance coverage by the district.

3. **All**: This set includes all features in our data set including sensitive variables—from a sociopolitical perspective—such as the proportion of minorities, and unemployment rates among minorities. This set also includes age-related features such as the proportion of young and elderly households.

B. Statistical methods

1. Local model (Baseline)

We treat the design tensors $\boldsymbol{X}$ as exogenous variables and do not model their evolution across time. A "design tensor" is a rank 3 tensor given by $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_T)$ where $\boldsymbol{X}_t$ is the design matrix for time period t. Each design matrix $\boldsymbol{X}_t$ is of dimension $N \times (p+1)$, where N is the number of observations and p is the number of predictors. The $+1$ in the dimensionality of the design matrix is because we add a constant to the linear model. The dynamics of the local models are described by a system of linear equations,

$$\boldsymbol{y}_t = \boldsymbol{X}_t\boldsymbol{\beta}_t + \sigma_t\boldsymbol{u}_t, \quad (1)$$
$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{v}_t, \quad (2)$$
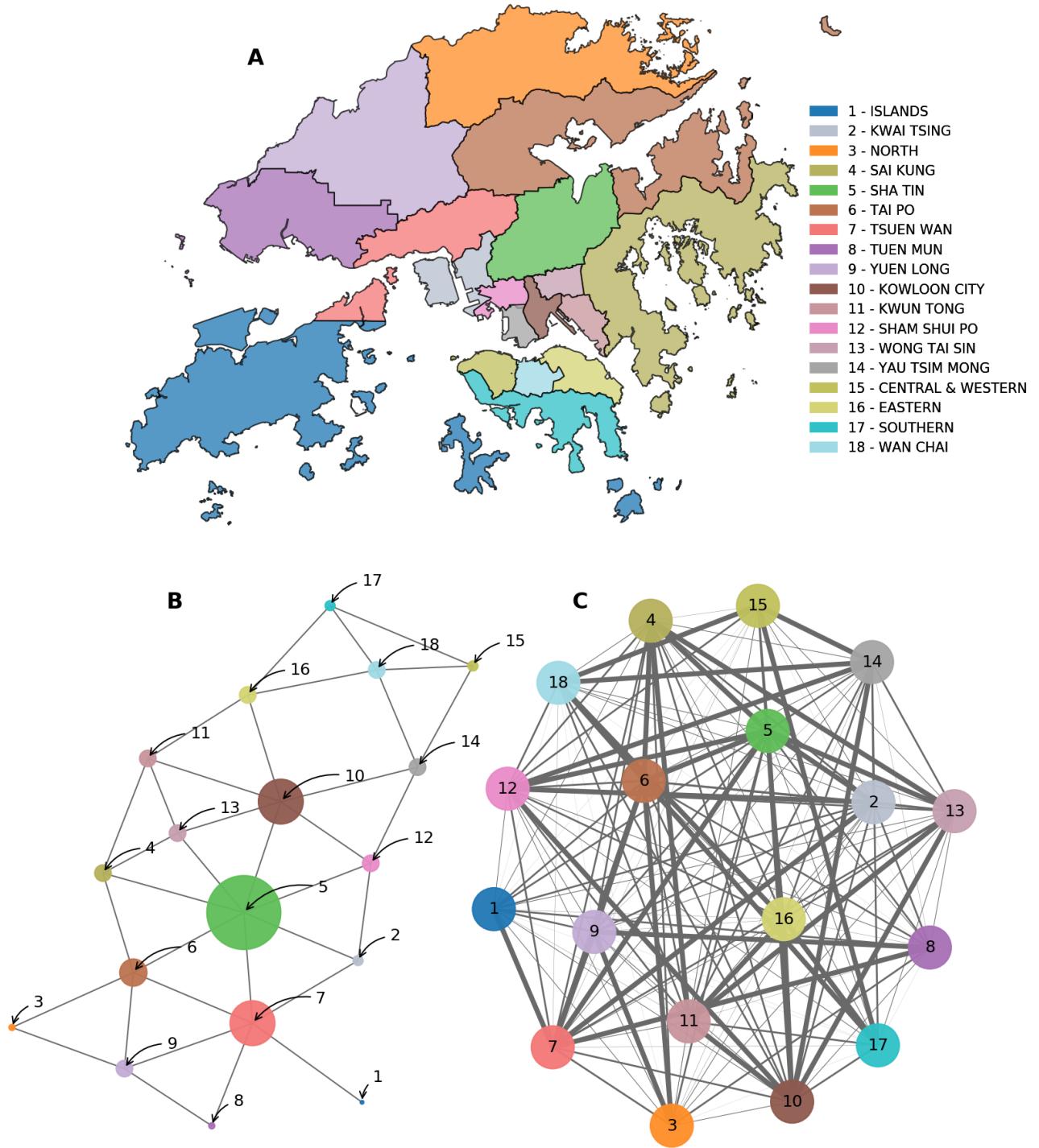$$\log\sigma_t = \log\sigma_{t-1} + \mu + \ell w_t, \quad (3)$$

for $t = 1, ..., T$.

Eq. 1 is an ordinary linear model for the response vector $\boldsymbol{y}_t$ as a function of the design matrix $\boldsymbol{X}_t$ and coefficients $\boldsymbol{\beta}_t$. We presently define the quantities that compose Eqs. 2 and 3. We set

$$\boldsymbol{u}_t, \boldsymbol{v}_t \sim \text{MultivariateNormal}(\boldsymbol{0}, \boldsymbol{I}) \quad (4)$$

in Eqs. 1 and 2, while $w_t \sim \text{Normal}(0, 1)$. By $\boldsymbol{I}$ we mean the $(p+1) \times (p+1)$ identity matrix. Hence the model likelihood is

$$p(\boldsymbol{y}|\boldsymbol{\beta}, \sigma) = \prod_{t=1}^{T}\prod_{n=1}^{N} p(y_{tn}|\boldsymbol{X}_{tn}\boldsymbol{\beta}_t, \sigma_t)$$
$$= \prod_{t=1}^{T}\prod_{n=1}^{N} \text{Normal}(\boldsymbol{X}_{tn}\boldsymbol{\beta}_t, \sigma_t^2). \quad (5)$$

FIG. 1. **Spatial networks of Hong Kong's districts**. (**A**) We display a map of the 18 districts of Hong Kong as outlined in Ref. [43]. We cross-reference the roads and bridges connecting these districts to build a spatial network [44, 45]. (**B**) We demonstrate the first undirected network layout of Hong Kong's districts. Districts (nodes) are linked if they border each other or share a direct road/bridge in a binary fashion. (**C**) We show a fully connected version of the network. For the fully connected network, edges to neighboring districts are weighted exponentially. Different weighting schemes can be applied here, however, for our application we use the spatial distance measured by the length of the shortest path connecting any two districts on the network.

FIG. 2. A graphical model representing the likelihood function given in Eq. 5. Latent $\log \sigma_t$ and β_t evolve as biased random walks, while y_{tn} and $\boldsymbol{X}_{tn}$ are treated as observable random variables and exogenous parameters respectively. The entire temporal model is plated across districts $N = 18$.

We display a graphical model corresponding to Eq. 5 in Fig. 2.

We assume a prior on $\boldsymbol{\beta}_t$ that evolves as a biased random walk with drift given by $\boldsymbol{\mu}$ and correlation matrix $\boldsymbol{\Sigma}$ with Cholesky decomposition $\boldsymbol{L}$. We make this choice because we *a priori* believe that $\boldsymbol{\beta}_t$ does not remain constant throughout the time under study, though we are unsure of exactly how it changes over this time. Likewise, we suppose that $\log \sigma_t$ evolves according to a univariate random walk with drift given by μ and standard deviation ℓ. We make this assumption for the same reason: We do not believe it is likely that σ_t remains constant over the entire time period of study. The random walk priors for $\boldsymbol{\beta}_t$ and σ_t are each centered about zero because we impose a zero mean prior on $\boldsymbol{\mu}$ and μ. We initialize these random walks with zero-centered multivariate normal initial conditions,

$$\boldsymbol{\beta}_0 \sim \text{MultivariateNormal}(\boldsymbol{0}, \boldsymbol{\Sigma}), \quad (6)$$

and

$$\log \sigma_0 \sim \text{Normal}(0, \ell^2). \quad (7)$$

The distribution of $\boldsymbol{\beta}_{1:T} \equiv (\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_T)$ is thus given by

$$p(\boldsymbol{\beta}_{1:T}|\boldsymbol{\mu}, \boldsymbol{\Sigma} = \prod_{t=1}^{T} p(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad (8)$$

$$= \prod_{t=1}^{T} \text{MultivariateNormal}(\boldsymbol{\beta}_{t-1} + \boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad (9)$$

with an analogous (univariate) distribution holding for $\log \sigma$. We set

$$\boldsymbol{\mu} \sim \text{MultivariateNormal}(\boldsymbol{0}, \boldsymbol{I}), \quad (10)$$

$$\mu \sim \text{Normal}(0, 1), \quad (11)$$

so that the prior distribution over the paths of the regression coefficients and log standard deviation are centered about zero—the null hypothesis—for all t.

We place a uniform prior (LKJ(1)) over the correlation component of $\boldsymbol{\Sigma}$. The mean of this prior is at the identity matrix. The vector of standard deviations of $\boldsymbol{\Sigma}$, $\boldsymbol{s}$, is hypothesized to follow an isotropic multivariate log normal distribution as $\boldsymbol{s} \sim \text{LogNormal}(0, 1)$. We also place a univariate $\text{LogNormal}(0, 1)$ prior on ℓ, the standard deviation of the increments of $\log \sigma_t$. We make this choice because we do not possess prior information about the appropriate noise scale of $\boldsymbol{\beta}_t$ or $\log \sigma_t$ and the log normal distribution is a weakly-informative prior that does not encode much prior information about their noise scales.

We did not perform exact inference of this model but rather fit parameters of a surrogate variational posterior ("guide") distribution. We made this choice because variational inference replaces the process of sampling with the much faster problem of optimization [49]. Denoting the vector of all latent random variables by

$$\boldsymbol{z} \equiv (\mu^{(\sigma)}, \ell, \boldsymbol{\mu}^{(\beta)}, \Sigma^{(\beta)}, \log \sigma_{0:T}, \boldsymbol{\beta}_{0:T}), \quad (12)$$

we fit the parameters ψ of an approximate posterior distribution $q_\psi(\boldsymbol{z})$ to maximize the variational lower bound, defined as the expectation under $q_\psi(\boldsymbol{z})$ of the difference between the log joint probability and $q_\psi(\boldsymbol{z})$ [50]. We chose a low-rank multivariate normal guide with rank equal to approximately $\sqrt{\dim \boldsymbol{z}}$. This low rank approximation allows for modeling of correlations in the posterior distribution of $\boldsymbol{z}$ with a lower number of parameters than, for example, a full-rank multivariate normal guide distribution. All bounded latent random variables are reparameterized to lie in an unconstrained space so that we could approximate them with the multivariate normal guide.
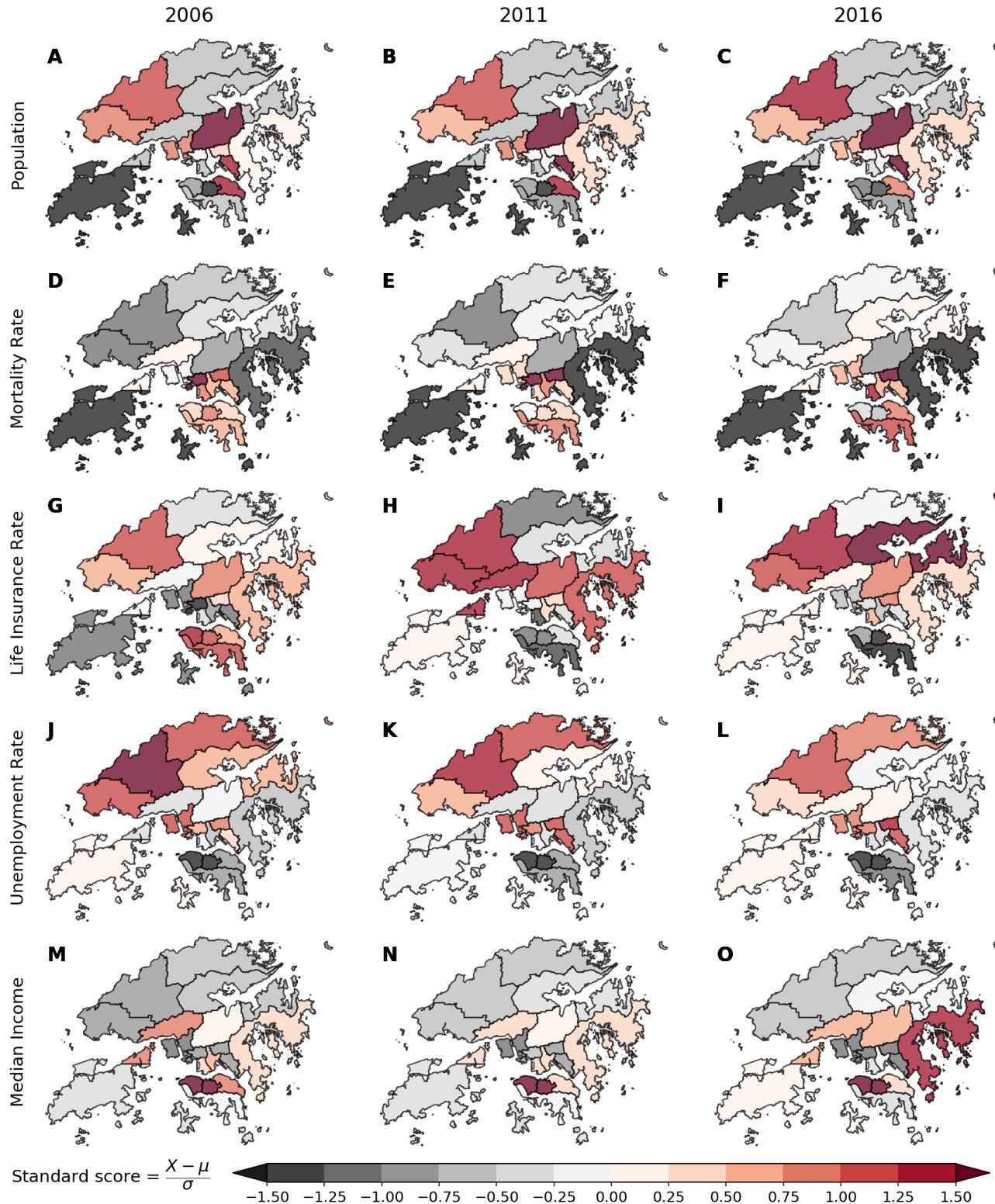
2. Nonlocal models (SP and WSP)

We use the road networks in Hong Kong [44, 45] to build a spatial network of the 18 districts of Hong Kong. Each node in the network represents a single district. Nodes are linked if they share a direct road or bridge. In Fig. 1A, we show a map of Hong Kong's districts. We display an undirected network of districts in Fig. 1B. By contrast, we show a fully connected version of the network in Fig. 1C. Edges are weighted by their spatial distance measured by the length of the shortest path d_{ij} to reach from district i to district j

$$w_{ij} = \exp\{-(d_{ij} - 1)\}, \quad (13)$$

and weights decays exponentially as the length of the shortest path increases between any two districts.

Similarly, we treat the adjacency matrix $\boldsymbol{A}$ as an exogenous variable. We fit two nonlocal models that take account of the design matrix associated with each district's neighbors; a spatial model (SP) that uses the binary adjacency matrix and a weighted spatial model (WSP) that uses the weighted adjacency matrix. The equations

FIG. 3. **Temporal dynamics of spatial socioeconomic characteristics of Hong Kong**. We show the spatial distribution of five features in our datasets for three different census years. Here, heatmaps are normalized by the mean and standard deviation. Darker shades of red show areas above the mean for each of these variables while shades of grey show areas below the mean. (**A**–**C**) We display the spatial growth of population over time. (**D**–**F**) We demonstrate the variation of mortality rates, and life insurance converge (**G**–**I**). We see some segregation of unemployment rates in (**J**–**L**), and median income in (**M**–**L**).

describing the time evolution of this data generating process are

$$\boldsymbol{y}_t = \boldsymbol{X}_t\boldsymbol{\beta}_t + \boldsymbol{f}\left(\boldsymbol{X}_t \otimes \boldsymbol{B}\right)\boldsymbol{\gamma}_t + \sigma_t\boldsymbol{u}_t, \qquad (14)$$

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{v}_t, \qquad (15)$$

$$\boldsymbol{\gamma}_t = \boldsymbol{\gamma}_{t-1} + \boldsymbol{m} + \boldsymbol{\Lambda}\boldsymbol{q}_t, \qquad (16)$$

$$\log\sigma_t = \log\sigma_{t-1} + \mu + \ell w_t, \qquad (17)$$

for $t = 1, ..., T$. The rank three tensor $\boldsymbol{X}_t \otimes \boldsymbol{B}$ is the outer product of $\boldsymbol{B} \equiv \boldsymbol{A} - \boldsymbol{I}$ with the design matrix $\boldsymbol{X}_t$. The function $\boldsymbol{f}$ is a reduction function that lowers the rank of the tensor by one by collapsing the first dimension. Here we take $\boldsymbol{f}$ to be the mean across the first dimension. In other words, $\boldsymbol{f}\left(\boldsymbol{X}_t \otimes \boldsymbol{B}\right)$ is a design matrix where $\boldsymbol{f}\left(\boldsymbol{X}_t \otimes \boldsymbol{B}\right)_{ij}$ is the average of the values of predictor j over all of the neighbors of district i in the network.

The prior distributions for $\boldsymbol{\gamma}_t$, $\boldsymbol{m}$, $\boldsymbol{\Lambda}$, and $\boldsymbol{q}_t$ are identical to those for $\boldsymbol{\beta}_t$, $\boldsymbol{\mu}$, $\boldsymbol{L}$, and $\boldsymbol{u}_t$ except their dimensionality is lowered from $p+1$ to p since we do not include an intercept term in $\boldsymbol{f}\left(\boldsymbol{X}_t \otimes \boldsymbol{B}\right)$.

We use Pyro [51], a probabilistic programming language that operates on top of Pytorch [52], a dynamic graph differentiable programming library, to implement our models. Our source code is available on Gitlab [53].

III. RESULTS AND DISCUSSION

In Fig. 3, we show the spatial distribution of socioeconomic characteristics for 2006, 2011, and 2016. Each heatmap is normalized by the mean and standard deviation for each year. Darker shades of red show areas above the mean for each of these variables while shades of grey show areas below the mean. We show normalized population density in Figs. 3A through C. We see dense clusters both at the center of the country and on the northwestern side. We note high mortality rates (see Figs. 3D–F) are evident in the southern islands while life insurance policies decline there. Life insurance coverage varies across Hong Kong but is still higher in densely urbanized areas shown in Figs. 3G through I. The northwestern territories have higher rates of unemployment compared to the southeastern side of Hong Kong as we see in Figs. 3J–L. In Figs. 3M–O, we observe that the east and center districts have higher normalized median income when adjusted for inflation. We display additional statistics regarding households rather than individuals in Fig. S2.

Our models provide evidence to suggest that there are significant relationships between socioeconomic variables, such as household unemployment, percentage of single parents, and mortality rate. Many of these relations are significant in each of the census periods under study (2006, 2011, and 2016) while other relations are significant for some of the census periods but not others. We display the distributions of our models' coefficients and associated significance assessments in Figs. S6–S9.

To take a closer look at this variation across models, we compute the difference between the mean of the prediction distributions and the ground truth mortality rates for each district. We provide empirical pdfs of the error distributions in Figs. 4 and S13–S14. If the models accurately associate features with observed mortality rate, these distributions should be centered around zero. Conversely, if the models display systematic over- or under-estimation of mortality rate, these distribution will not be centered around zero. In Fig. 4, we examine distribution of errors for each model and for each district (observation) for 2016 which is possible since our models are fully Bayesian and generate a distribution of possible outcomes.

We assess significance of model coefficients using centered $Q\%$ credible intervals (CI). A centered $Q\%$ credible interval of a probability density function (pdf) $p(x)$ is an interval (a, b) defined such that $\frac{1}{2}(1 - \frac{Q}{100}) = \int_{-\infty}^{a} dx\ p(x) = \int_{b}^{\infty} dx\ p(x)$. We measure the significance of systematic errors in each model by computing the 80% CI, whereby systematic overestimation is colored in orange (CI > 0), and systematic underestimation is colored in blue (CI < 0). We note that our spatial models, especially the weighted spatial model, are effective at reducing systematic over or underestimations. For example, the spatial models reduce the margin of errors in panels B, H, L, and R of Fig. 4. By contrast, all three models either overshoot or undershoot mortality rates drastically in a few districts (see panels D, E, M, and Q in Fig. 4).

Furthermore, the districts of Sai Kung, Sha Tin, Wong Tai Sin, and Southern are poorly fit by our models in 2016. Failures are, perhaps, the most interesting to investigate. In Fig. 5, we show inference error of these four districts. The first three rows demonstrate how all three models overestimate mortality rates for the Sai Kung and Sha Tin districts (Fig. 5A–B), while underestimating the other two (Fig. 5C–D). For each district, we plot the standardized score of some socioeconomic features. To examine the effect of neighborhoods on these areas, we also plot the corresponding average value of these features from the neighboring districts derived from our network and indicated by a star (*) at the end of the variable name.

The first two districts (Fig. 5A–B) have lower mortality rates than average mortality of Hong Kong and are often significantly overestimated by our models while their neighboring districts have around average mortality rate. By contrast, the other two districts (Fig. 5C–D) have higher mortality rates and are surrounded by districts with lower mortality rates. This provides additional evidence of sociospatial and economic factors in mortality risks. Relatively of wealth, on the other hand, can be seen in Figs. 5A–B with a higher median income in the district being associated with a higher median income in its neighborhood. Though districts in Figs. 5C–D are located in a wealthy neighborhood they still have lower median income. The Wong Tai Sin district has a lower
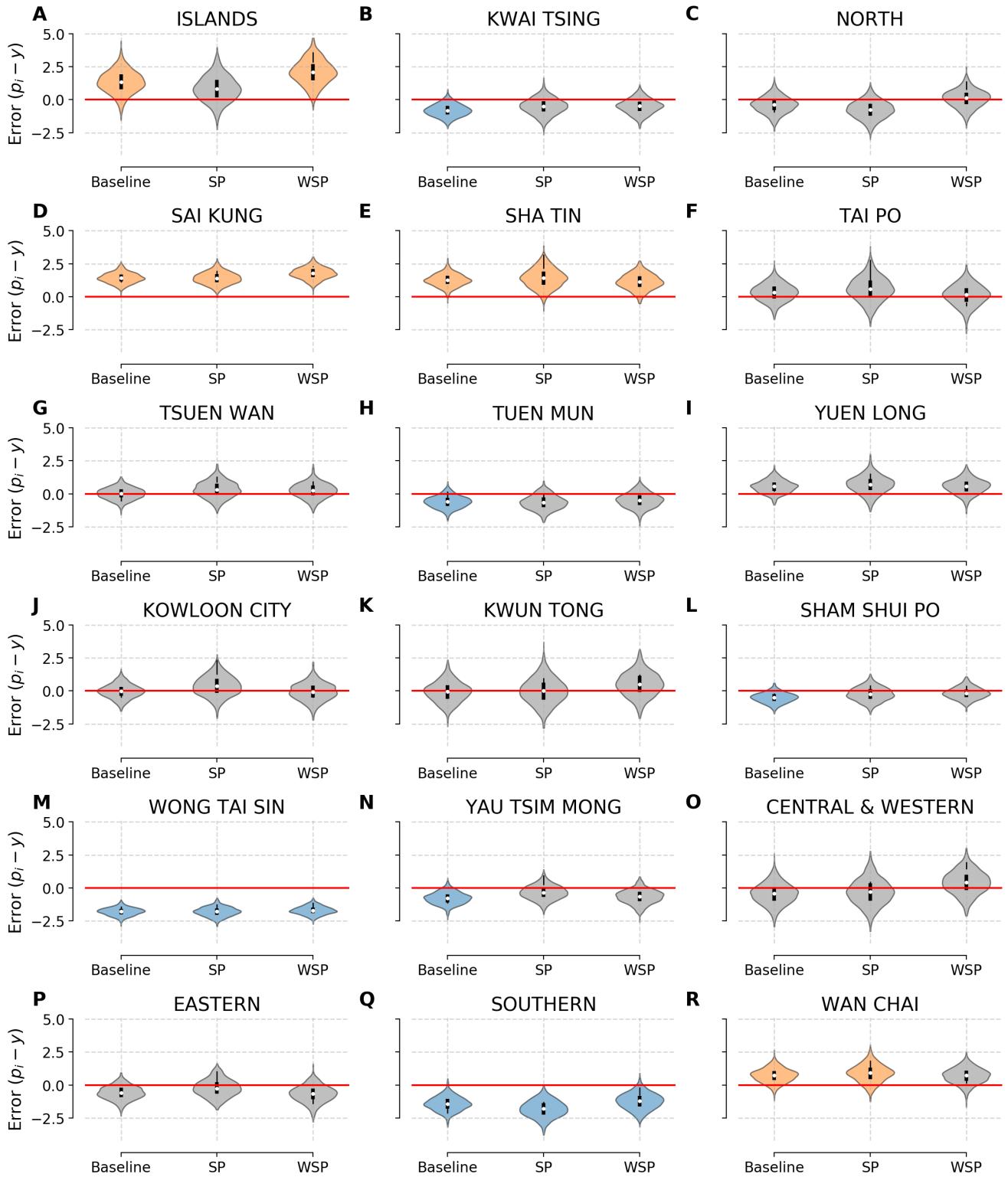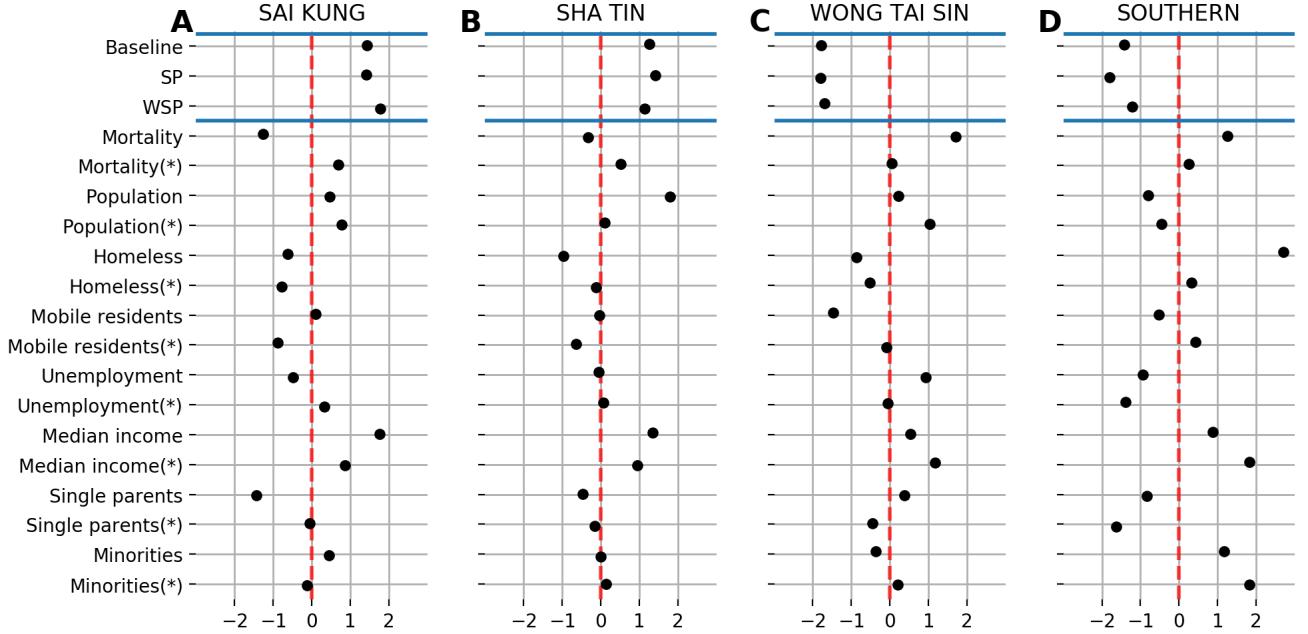
FIG. 4. **Systematic error distributions for models trained on the default set of features in 2016**. We examine signed errors by computing the difference between the mean of prediction distributions p_i and the ground truth mortality rates y for each district. A perfect model would have a narrow error distribution centred on 0 (solid red line going across). Positive error values demonstrate overestimation whereas negative values show underestimation of mortality rates for each district. Models with significance systematic overestimation are colored in orange while models with significance systematic underestimation are colored in blue as measured by the 80% CI. We show the same distributions for models trained on wealth-related, and all features in Fig. S13 and Fig. S14, respectively.

FIG. 5. **Impact of sociospatial factors on mortality risks..** The first three rows show normalized inference error for four districts that are poorly fit by our models. (**A–B**) We show districts with systematic overestimation of mortality rates, while (**C–D**) show districts where mortality rates are systematically underestimated. For each district, we show the normalized value of some features of interest along with the average value of these features in the neighboring districts indicated by (∗) symbol. The red dashed line indicate the average value for each of these normalized features centered at zero. Evidence of sociospatial factors of longevity can be seen in all four districts. Particularly, we note a spillover of wealth measured by median income. Districts in (**A**) and (**B**) maintain lower mortality rates while surrounded by districts with around average mortality rates. By contrast, districts in (**C**) and (**D**) have a socioeconomic pull driving the whole neighborhood to have higher mortality rates.

number of mobile residents than average (see Fig. 5C). In Fig. 5D, we see an extraordinary higher number of homeless people compared to the neighboring population. Our systematic errors can also be driven by a higher percentage of minorities in this district, which hints at disparities in mortality risks in the area. Further analyses with finer geospatial resolution would be needed to explain this behavior.

We also analyze our mortality models' error distributions through a local ego network approach. An ego network of a node is the network consisting that node and its nearest neighbors. Here each node is a district and its neighbors are that district's neighboring districts. The joint distribution of mortality at time t and district i, y_i^t, and model mortality prediction error $e_i^t \equiv p_i^t - y_i^t$ conditioned on location at node i, can be concisely represented in the form of an ego network for each district. We display an example of this representation in Fig. 6. We color nodes by standardized mortality and edges by standardized prediction error. Though this is an exploratory methodology that deserves greater expansion and attention in future work, we note qualitatively that the local view of predicted mortality versus true mortality varies substantially as a function of district. For example, the district of Wan Chai is connected to four other districts,

three of which have substantially-higher mortality than average and in particular higher mortality than Wan Chai itself, which has lower mortality than average. The model predictions for these neighbors are lower than their true mortality. An observer in Wan Chai who has access only to the model predictions and not the true mortality data would rationally assume that mortality in these districts is much lower than it truly is and could subsequently make further inferences or decisions based on faulty-but-rational assumption.

Notably, socioeconomic diversity or heterogeneity of the neighborhood can change the local perception of mortality models across neighboring communities. We observe that high divergence from the neighborhood is associated with higher rates of uncertainty/error in the models. One can propose using prediction error as a proxy of what the models have learned (or failed to learn) about socioeconomic status of neighborhood relative to the target district. We envision future work incorporating this sort of information to fine tune mortality models.
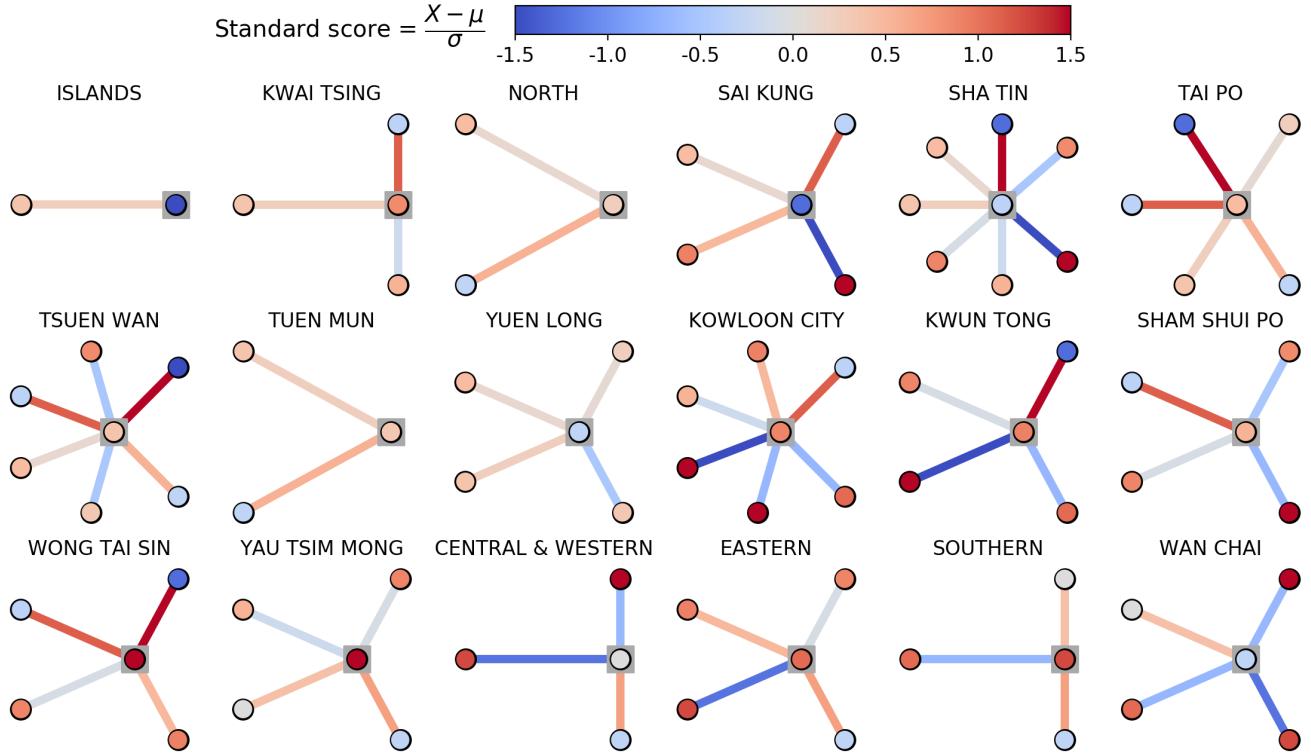
FIG. 6. **Ego networks of each district demonstrating sociospatial factors of mortality for the 2016 weighted spatial model**. We display ego networks of each district in Hong Kong and its nearest neighbors in the road and bridge network. The central node (highlighted with a grey box) of each network corresponds to the labelled district. Neighbors are not arranged around the ego district geographically. Node color corresponds to normalized mortality rate and edge color corresponds to signed prediction error for the 2016 WSP model. These ego networks encode a qualitative measure of the sociospatial factors in mortality modeling. We display the equivalent networks for the Baseline and SP models in Fig. S15 and Fig. S16 respectively.

IV. CONCLUDING REMARKS

Our results support our hypothesis that spatial associations of wealth or social deprivation among neighborhoods has a direct and sometimes substantial impact on mortality risks. Our experiments reveal that localized models—which do not account for sociospatial factors—can systematically over- or under-estimate mortality rate—while spatial models manage to reduce the error of predicting mortality rate. We find varying degrees of inference error as a factor of the marginal function of the design tensor. Different features can introduce different biases to the models indicating different levels of disparities in predicting mortality risks. We recommend using a non-local approach to account for all of these factors.

Our findings are limited for four reasons. First and foremost, we only have access to census data for three individual years spanning a decade and a half. A better explanation of the nonlocalized effect of neighborhoods could be achieved by testing our hypotheses on additional years, along with more socioeconomic features to enrich our design tensor.

Second, our geospatial resolution is unfortunately not high enough to identify some of the dynamics of connected communities. Nonetheless, our methodology can be implemented in a similar manner regardless of the spatial unit used for the experiment. Our spatial network is mainly based on the road network of Hong Kong, which could be extended to account for roads/bridges and public transport across any desired spatial unit.

Third, our models are rather simple but explainable to demonstrate the role of relative spatial associations om longevity. More sophisticated geospatial techniques could help us understand the impact of socioeconomic factors on mortality risks.

Finally, we have only explored a distance-based weighting scheme for the connections across districts in the network. Population density could be included to enrich the socioeconomic effect of neighboring regions on a given node within the network (for example, theory of intervening opportunities [54]). Other attributes such as geographic information associated with community health services could help us assess their value and reallocate these centers to more optimized locations.

[1] I. O. Wong, C. Schooling, B. J. Cowling, and G. M. Leung. Breast cancer incidence and mortality in a transitioning Chinese population: Current and future trends. *British Journal of Cancer*, 112(1):167–170, 2015.

[2] P. Wu, A. M. Presanis, H. S. Bond, E. H. Lau, V. J. Fang, and B. J. Cowling. A joint analysis of influenza-associated hospitalizations and mortality in Hong Kong, 1998–2013. *Scientific Reports*, 7(1):929, 2017.

[3] C.-M. Wong, S. Ma, A. J. Hedley, and T.-H. Lam. Effect of air pollution on daily mortality in Hong Kong. *Environmental Health Perspectives*, 109(4):335–340, 2001.

[4] H. Qiu, L. Tian, K.-f. Ho, V. C. Pun, X. Wang, and T. Ignatius. Air pollution and mortality: Effect modification by personal characteristics and specific cause of death in a case-only study. *Environmental Pollution*, 199:192–197, 2015.

[5] T.-H. Lam, S.-Y. Ho, A. J. Hedley, K.-H. Mak, and G. M. Leung. Leisure time physical activity and mortality in Hong Kong: Case-control study of all adult deaths in 1998. *Annals of Epidemiology*, 14(6):391–398, 2004.

[6] M. D. Hayward, A. M. Pienta, and D. K. McLaughlin. Inequality in men's mortality: The socioeconomic status gradient and geographic context. *Journal of Health and Social Behavior*, pages 313–330, 1997.

[7] A. Deaton and D. Lubotsky. Mortality, inequality and race in American cities and states. *Social Science & Medicine*, 56(6):1139–1153, 2003.

[8] G. Rey, E. Jougla, A. Fouillet, and D. Hémon. Ecological association between a deprivation index and mortality in France over the period 1997–2001: Variations with spatial scale, degree of urbanicity, age, gender and cause of death. *BMC Public Health*, 9(1):33, 2009.

[9] L. C. Messer, B. A. Laraia, J. S. Kaufman, J. Eyster, C. Holzman, J. Culhane, I. Elo, J. G. Burke, and P. Ocampo. The development of a standardized neighborhood deprivation index. *Journal of Urban Health*, 83(6):1041–1062, 2006.

[10] C.-Q. Ou, A. J. Hedley, R. Y. Chung, T.-Q. Thach, Y.-K. Chau, K.-P. Chan, L. Yang, S.-Y. Ho, C.-M. Wong, and T.-H. Lam. Socioeconomic disparities in air pollution-associated mortality. *Environmental Research*, 107(2):237–244, 2008.

[11] R. Y. Chung, F. T. Lai, G. K. Chung, B. H. Yip, S. Y. Wong, and E. K. Yeoh. Socioeconomic disparity in mortality risks widened across generations during rapid economic development in Hong Kong: An age-period-cohort analysis from 1976 to 2010. *Annals of Epidemiology*, 28(11):743–752, 2018.

[12] K. Buckingham and P. R. Freeman. Sociodemographic and morbidity indicators of need in relation to the use of community health services: Observational study. *BMJ*, 315(7114):994–996, 1997.

[13] D. Kim, I. Kawachi, S. Vander Hoorn, and M. Ezzati. Is inequality at the heart of it? Cross-country associations of income inequality with cardiovascular diseases and risk factors. *Social Science & Medicine*, 66(8):1719–1732, 2008.

[14] Z. Chen and C. A. G. Crawford. The role of geographic scale in testing the income inequality hypothesis as an explanation of health disparities. *Social Science & Medicine*, 75(6):1022–1031, 2012.

[15] J. X. Fan, M. Wen, and L. Kowaleski-Jones. Tract-and county-level income inequality and individual risk of obesity in the United States. *Social Science Research*, 55:75–82, 2016.

[16] T.-C. Yang, S. A. Matthews, F. Sun, and M. Armendariz. Modeling the importance of within-and between-county effects in an ecological study of the association between social capital and mental distress. *Preventing Chronic Disease*, 16:E75–E75, 2019.

[17] E. E. Bjornstrom. An examination of the relationship between neighborhood income inequality, social resources, and obesity in Los Angeles county. *American Journal of Health Promotion*, 26(2):109–115, 2011.

[18] D. Kim, F. Wang, and C. Arcan. Peer reviewed: Geographic association between income inequality and obesity among adults in New York state. *Preventing Chronic Disease*, 15, 2018.

[19] I. Kawachi and B. P. Kennedy. The relationship of income inequality to mortality: Does the choice of indicator matter? *Social Science & Medicine*, 45(7):1121–1127, 1997.

[20] J. M. Major, C. A. Doubeni, N. D. Freedman, Y. Park, M. Lian, A. R. Hollenbeck, A. Schatzkin, B. I. Graubard, and R. Sinha. Neighborhood socioeconomic deprivation and mortality: NIH-AARP diet and health study. *PLOS ONE*, 5(11), 2010.

[21] S. F. Reardon and K. Bischoff. Income inequality and income segregation. *American Journal of Sociology*, 116(4):1092–1153, 2011.

[22] T.-C. Yang and L. Jensen. Exploring the inequality-mortality relationship in the US with Bayesian spatial modeling. *Population Research and Policy Review*, 34(3):437–460, 2015.

[23] N. A. Ross, M. C. Wolfson, J. R. Dunn, J.-M. Berthelot, G. A. Kaplan, and J. W. Lynch. Relation between income inequality and mortality in Canada and in the United States: Cross sectional assessment using census data and vital statistics. *BMJ*, 320(7239):898–902, 2000.

[24] A. S. Fotheringham, M. E. Charlton, and C. Brunsdon. Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. *Environment and Planning A*, 30(11):1905–1927, 1998.

[25] A. S. Fotheringham, C. Brunsdon, and M. Charlton. *Geographically weighted regression: The analysis of spatially varying relationships*. John Wiley & Sons, 2003.

[26] C.-M. Wong, C.-Q. Ou, K.-P. Chan, Y.-K. Chau, T.-Q. Thach, L. Yang, R. Y.-N. Chung, G. N. Thomas, J. S. M. Peiris, T.-W. Wong, et al. The effects of air pollution on mortality in socially deprived urban areas in Hong Kong, China. *Environmental Health Perspectives*, 116(9):1189–1194, 2008.

[27] M. Branis and M. Linhartova. Association between unemployment, income, education level, population size and air pollution in Czech cities: Evidence for environmental inequality? A pilot national scale analysis. *Health & Place*, 18(5):1110–1114, 2012.

[28] F. Forastiere, M. Stafoggia, C. Tasco, S. Picciotto, N. Agabiti, G. Cesaroni, and C. A. Perucci. Socioeconomic status, particulate air pollution, and daily mortality: Differential exposure or differential susceptibility. *American Journal of Industrial Medicine*, 50(3):208–216, 2007.

[29] C. M. Padilla, W. Kihal-Talantikite, V. M. Vieira, P. Rossello, G. Le Nir, D. Zmirou-Navier, and S. Deguen. Air quality and social deprivation in four French metropolitan areas—a localized spatio-temporal environmental inequality analysis. *Environmental Research*, 134:315–324, 2014.

[30] J. S. Cossman, R. E. Cossman, W. L. James, C. R. Campbell, T. C. Blanchard, and A. G. Cosby. Persistent clusters of mortality in the United States. *American Journal of Public Health*, 97(12):2148–2150, 2007.

[31] T.-Q. Thach, Q. Zheng, P.-C. Lai, P. P.-Y. Wong, P. Y.-K. Chau, H. J. Jahn, D. Plass, L. Katzschner, A. Kraemer, and C.-M. Wong. Assessing spatial associations between thermal stress and mortality in Hong Kong: A small-area ecological study. *Science of the Total Environment*, 502:666–672, 2015.

[32] Tertiary Planning Units. https://www.bycensus2016.gov.hk/en/bc-dp-tpu.html, 2016. Data retrieved from Census and Statistics Department of Hong Kong.

[33] G. Erreygers and T. Van Ourti. Measuring socioeconomic inequality in health, health care and health financing by means of rank-dependent indices: A recipe for good practice. *Journal of Health Economics*, 30(4):685–694, 2011.

[34] R. W. Zimmer and E. F. Toma. Peer effects in private and public schools across countries. *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management*, 19(1):75–92, 2000.

[35] B. Sacerdote. Peer effects in education: How might they work, how big are they and how much do we know thus far? In *Handbook of the Economics of Education*, volume 3, pages 249–277. Elsevier, 2011.

[36] E. Bodine-Baron, S. Nowak, R. Varadavas, and N. Sood. Conforming and non-conforming peer effects in vaccination decisions. Technical report, National Bureau of Economic Research, 2013.

[37] D. Albert, J. Chein, and L. Steinberg. The teenage brain: Peer influences on adolescent decision making. *Current Directions in Psychological Science*, 22(2):114–120, 2013.

[38] T.-C. Yang, A. J. Noah, and C. Shoff. Exploring geographic variation in US mortality rates using a spatial Durbin approach. *Population, Space and Place*, 21(1):18–37, 2015.

[39] D. Holtz, M. Zhao, S. G. Benzell, C. Y. Cao, M. A. Rahimian, J. Yang, J. N. L. Allen, A. Collis, A. V. Moehring, T. Sowrirajan, et al. Interdependence and the cost of uncoordinated responses to COVID–19. 2020.

[40] T.-C. Yang, L. Jensen, and M. Haran. Social capital and human mortality: Explaining the rural paradox with county-level mortality data. *Rural Sociology*, 76(3):347–374, 2011.

[41] H. Li, C. A. Calder, and N. Cressie. Beyond Moran's I: Testing for spatial dependence based on the spatial autoregressive model. *Geographical Analysis*, 39(4):357–375, 2007.

[42] P. A. Moran. Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2):17–23, 1950.

[43] District and Constituency Area. https://www.bycensus2016.gov.hk/en/bc-dp.html, 2016. Data retrieved from Census and Statistics Department of Hong Kong.

[44] Q. Zhou and Z. Li. Empirical determination of geometric parameters for selective omission in a road network. *International Journal of Geographical Information Science*, 30(2):263–299, 2016.

[45] B. Y. Chen, W. H. Lam, A. Sumalee, Q. Li, and Z.-C. Li. Vulnerability analysis for large-scale and congested road networks with demand uncertainty. *Transportation Research Part A: Policy and Practice*, 46(3):501–516, 2012.

[46] Micro-data set of known and registered deaths in Hong Kong. https://www.censtatd.gov.hk/service_desk/list/microdata/index.jsp, 2017.

[47] Annual long term business statistics. https://www.ia.org.hk/en/infocenter/statistics/annual_long_term_business_statistics.html, 2016.

[48] Population census data for Hong Kong. https://www.bycensus2016.gov.hk/en/index.html, 2016.

[49] R. Ranganath, S. Gerrish, and D. M. Blei. Black box variational inference. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, 2014.

[50] M. D. Hoffman, D. M. Blei, C. Wang, and J. Paisley. Stochastic variational inference. *The Journal of Machine Learning Research*, 14(1):1303–1347, 2013.

[51] E. Bingham, J. P. Chen, M. Jankowiak, F. Obermeyer, N. Pradhan, T. Karaletsos, R. Singh, P. Szerlip, P. Horsfall, and N. D. Goodman. Pyro: Deep universal probabilistic programming. *Journal of Machine Learning Research*, 2018.

[52] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[53] https://gitlab.com/compstorylab/asis.

[54] S. A. Stouffer. Intervening opportunities: A theory relating mobility and distance. *American Sociological Review*, 5(6):845–867, 1940.
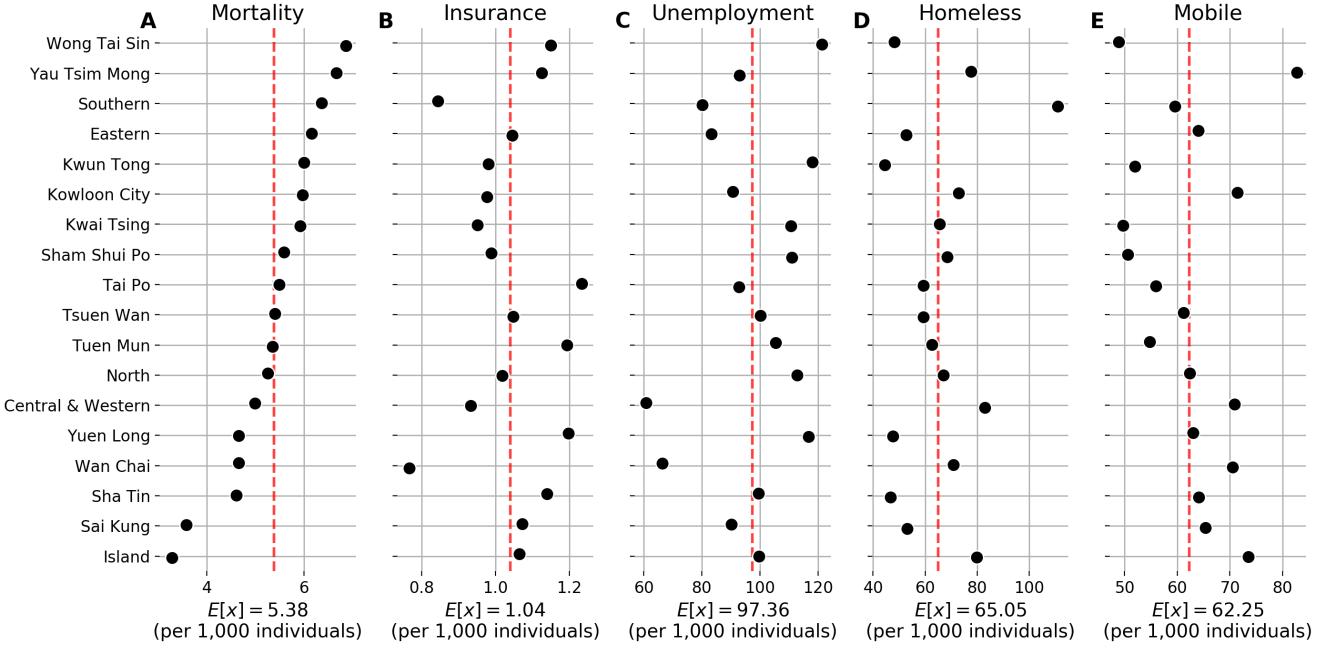
FIG. S1. **Cross-sectional attributes for calendar year 2016 by district** including **(A)** mortality rates, **(B)** number of life insurance policies, **(C)** unemployment rates, **(D)** number of homeless individuals, and **(E)** number of mobile residents. The dashed red lines indicate the average value for each of these variables in Hong Kong. The four attributes presented here are not strongly correlated with mortality.

- **Mortality rates**: Normalized death rates by districts derived from the registered deaths provided by the Census and Statistics Department of Hong Kong [46].

- **MMA life insurance coverage**: Annual proportion of people with life insurance policies in each district issued by MassMutual Asia.

- **Median income**: Estimation of median income by the Census and Statistics Department of Hong Kong for each district.

- **Median rent to income ratio**: The percentage of monthly household income spent on monthly household rent in inflation-adjusted terms.

- **Median monthly household income**: Similar to median income, but this is computed for households rather than individuals.

- **Unemployment rates**: The proportion of people without an active job normalized by population size. This is a proxy variable derived from the gap between the available labour force and economically active individuals—aged over 15 and has been working for at least a week in the reference year [48].

- **Unemployment rates across households**: Similar to unemployment rates but derived at the households level rather than individuals.

- **Unemployment rates among minorities**: Similar to unemployment rates but computed for minorities only.

- **Proportion of homeless people**: Another proxy variable to estimate the proportion of people not living in registered domestic households.

- **Proportion of homeless mobile residents**: Hong Kong's permanent residents who had stayed in a given district for "at least 1 month but less than 3 months" during the the 6 months before or after the reference moment [48].
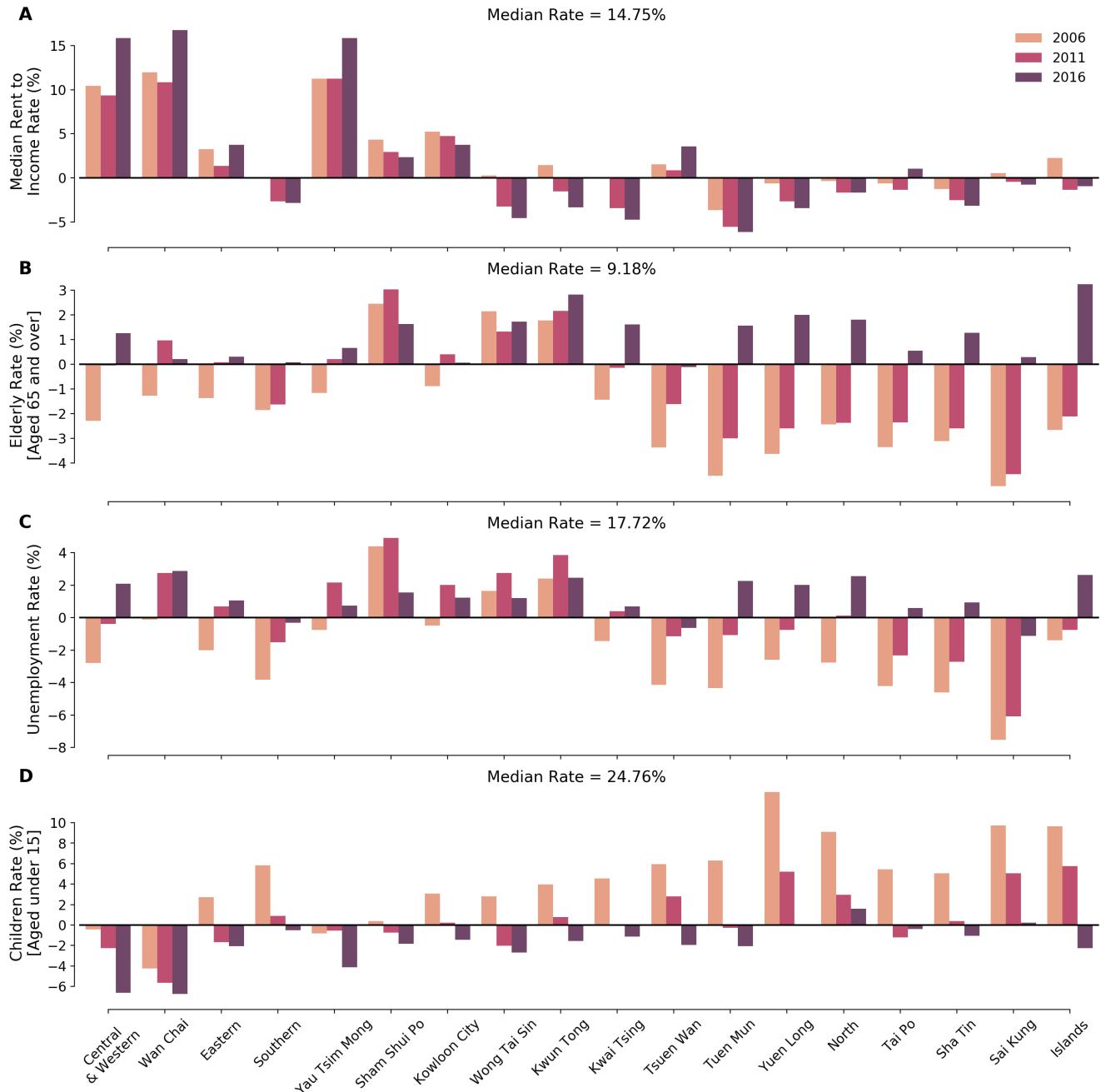
FIG. S2. **Households statistics by year.**

- **Proportion of single parents**: "Mothers or fathers who are never married, widowed, divorced or separated, with child(ren) aged under 18 living with them in the same household" [48].

- **Proportion of households with kids in school**: The proportion of households with young adults attending full-time courses in educational institutions in Hong Kong.

- **Proportion of households with children**: The proportion of households with young kids aged under 15.

- **Proportion of households with elderly**: The proportion of households with seniors aged 65 or above.

Our initial analysis shows variations in mortality rates across districts. In Fig. S1, we show a cross-sectional analysis of 4 different socioeconomic indices with respect to our response variable (mortality rate) for 2016. We normalize

FIG. S3. **Records by TPU. (A)** shows the fraction of records reported by each TPU in the mortality dataset. **(B)** shows an exponential declines in the number of records by TPU whereby the horizontal axis shows TPUs ranked by the fraction of death records reported on the vertical axis. Similarly, **(C)** shows the annual average of deaths for each decade by TPU.

these values to show the number of people in each bin per 1,000 individuals for each variable. The dashed red lines indicate the average value. Districts with a lower number of mobile residents have higher rates of morality such as Wong Tai Sin, Kwun Tong, Kwai Tsing, and Sham Shui Po. We note high mortality rates in the Yau Tsim Mong and Southern districts, while having a higher number of homeless people. By contrast, many districts with a lower number of homeless people have higher rates of mortality such as Wong Tai Sin, Eastern, and Kwun Tong, and Kowloon City.

Appendix B: Tertiary Planning Unit (TPU)

A Tertiary Planning Unit (TPU) [32]—similar to a census-block in the US—is a geospatial reference system used by the Hong Kong's Census and Statistics Department to report population census statistics. To take a closer look at the geospatial tags we have in our dataset, we show the fraction of death records found in our data set by each TPU (Fig. S3) and by districts (Fig. S4). Unfortunately, most death records were found in a small fraction of TPUs, which is understandable due to known data privacy concerns about the dataset as records in small TPUs may reveal sensitive information about specific individuals there. In Fig. S5, we show the annual number of reported deaths

by TPU. We note that death records are not reported frequently at the TPU level in the dataset. Missing records are flagged with XXX as indicated in the data set dictionary provided by the Hong Kong's Census and Statistics Department.
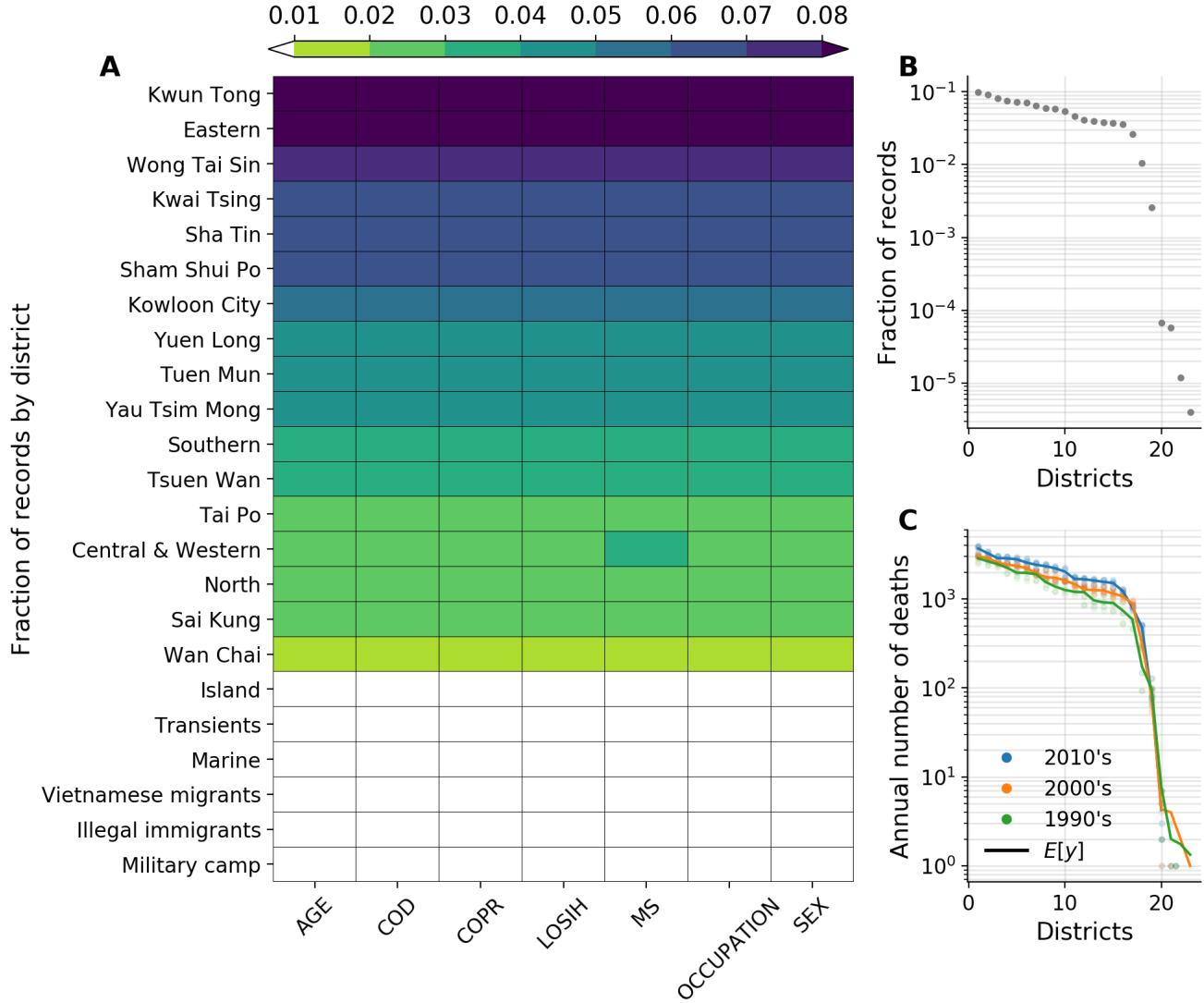

FIG. S4. **Records by district.** **(A)** shows the fraction of records reported by each district in the mortality dataset. **(B)** shows an exponential declines in the number of records by district whereby the horizontal axis shows districts ranked by the fraction of death records reported on the vertical axis. Similarly, **(C)** shows the annual average of deaths for each decade by district.

To avoid any risk of identifying individuals in the dataset, we use districts as our main spatial unit trying our best to stay away from the pitfalls of data privacy in census data. In fact, most studies have either filtered out small TPUs in their analyses [11, 26], or aggregated their records at the district level [31] to overcome this challenge in the dataset. In Fig. S4, the main 18 districts of Hong Kong show very consistent data distributions for all variables in the data set throughout the last few decades. We note that the last 5 entries in the heatmap are not necessarily districts per se, but they are only used by the census department to report deaths outside the borders of the main 18 districts. However, given access to a more granular geospatial resolution (*i.e.,* TPU) our methodology can be implemented in a similar manner to identify and explore some of these sociotechnical dynamics.

FIG. S5. **Annual number of deaths by TPU.** Death records are not reported frequently at the TPU level in our mortality dataset. We note that the code (XXX) is used to indicate missing labels.

Appendix C: Model parameter estimates

We show the distributions of β's in Fig. S6—the parameter for the baseline model. For each panel, we show the kernel density estimation of β as a function of each variable in the design tensor. Distributions that are significantly above 0 are colored in orange while distributions significantly below 0 are colored in blue as measured by the 80% CI. Similar demonstrations of spatial and weighted spatial models can be found in Fig. S7, and Fig. S8 respectively. In addition to the distributions of β's, we also show the kernel density estimation of γ's—the hyperparameter used for the spatial competent in each model in Fig. S9.

Appendix D: Kernel density estimation of normalized mortality rates

In Fig. S10, we show the kernel density estimation of normalized mortality rates by a model trained on the default set of base features for the calendar year 2016. The red dashed line indicates the true normalized mortality rate for each district in 2016. The KDE of each model is plotted by district: Baseline model (grey), Spatial model (blue), and Weighted Spatial model (orange). For additional insights into the models, Fig. S11 and Fig. S12 show different variants of these models trained on wealth-related and all features respectively.

Appendix E: Supplementary Figures

FIG. S6.   **Distributions of β for the baseline model.** We show the kernel density estimation of β coefficient for each feature in the design tensor. Distributions that are significantly above the mean are colored in orange while distributions significantly below the mean are colored in blue as measured by the 80% CI.
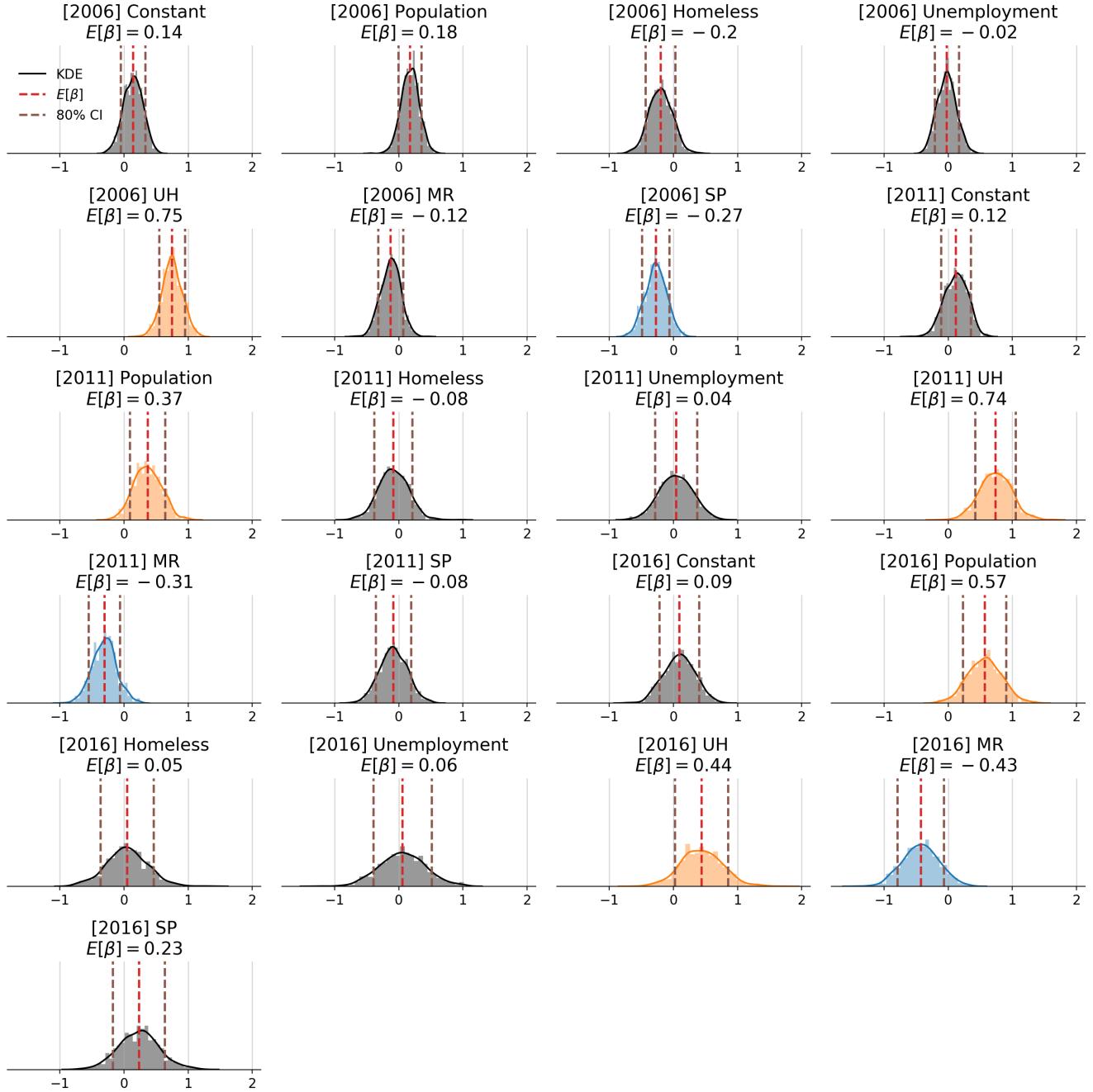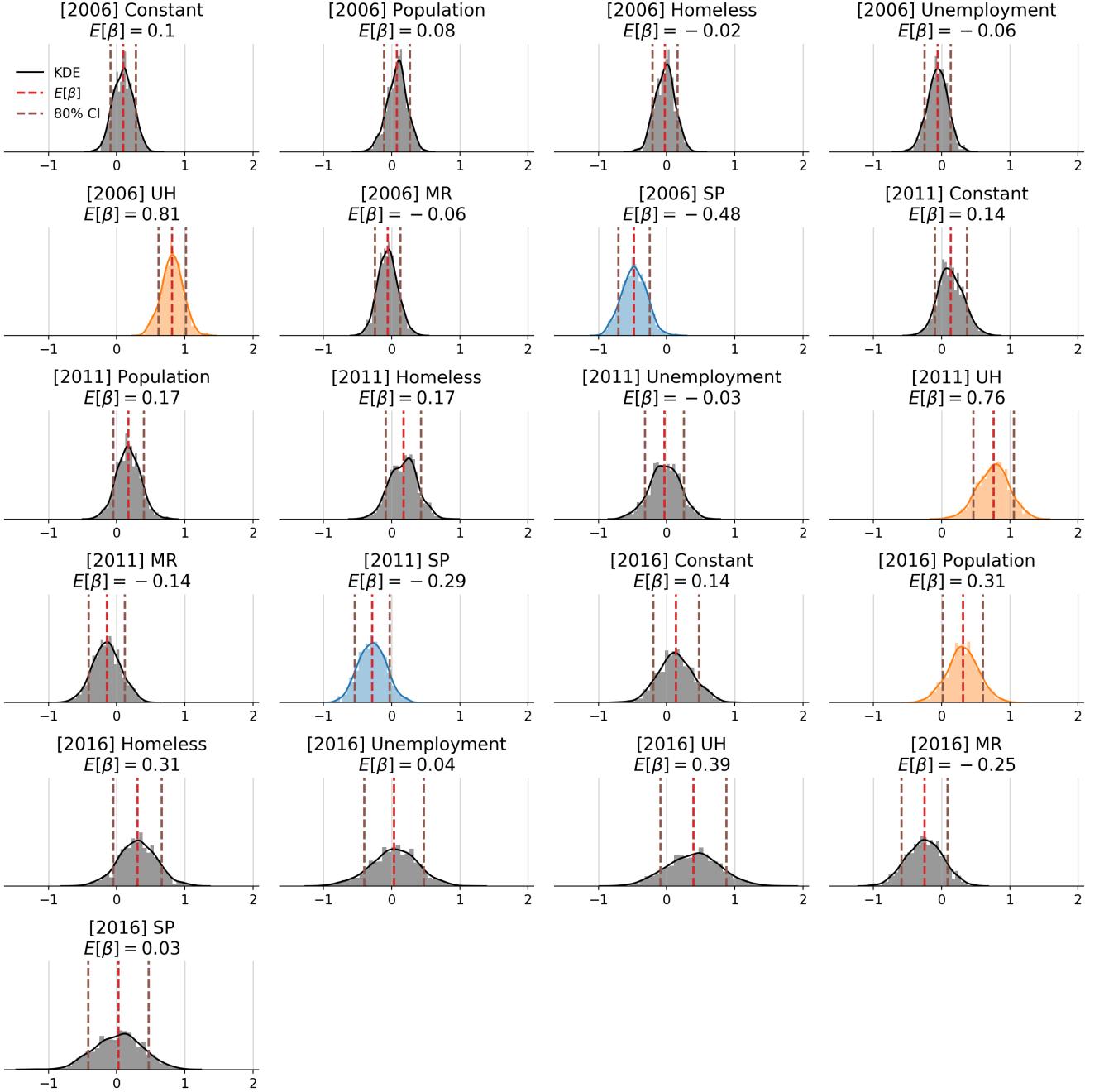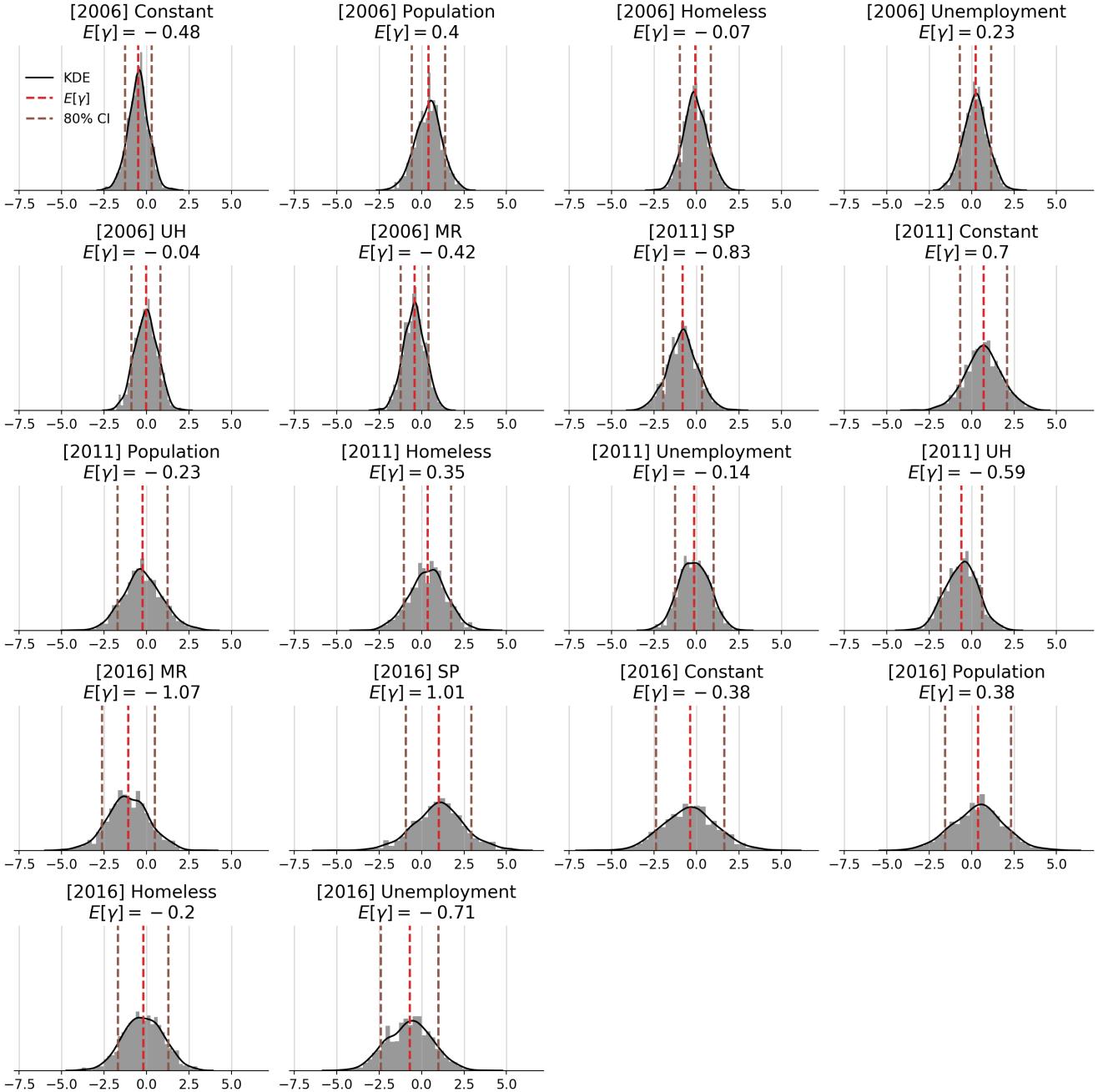
FIG. S7. **Distributions of β for the spatial model.** We show the kernel density estimation of β coefficient for each feature in the design tensor. Distributions that are significantly above the mean are colored in orange while distributions significantly below the mean are colored in blue as measured by the 80% CI.
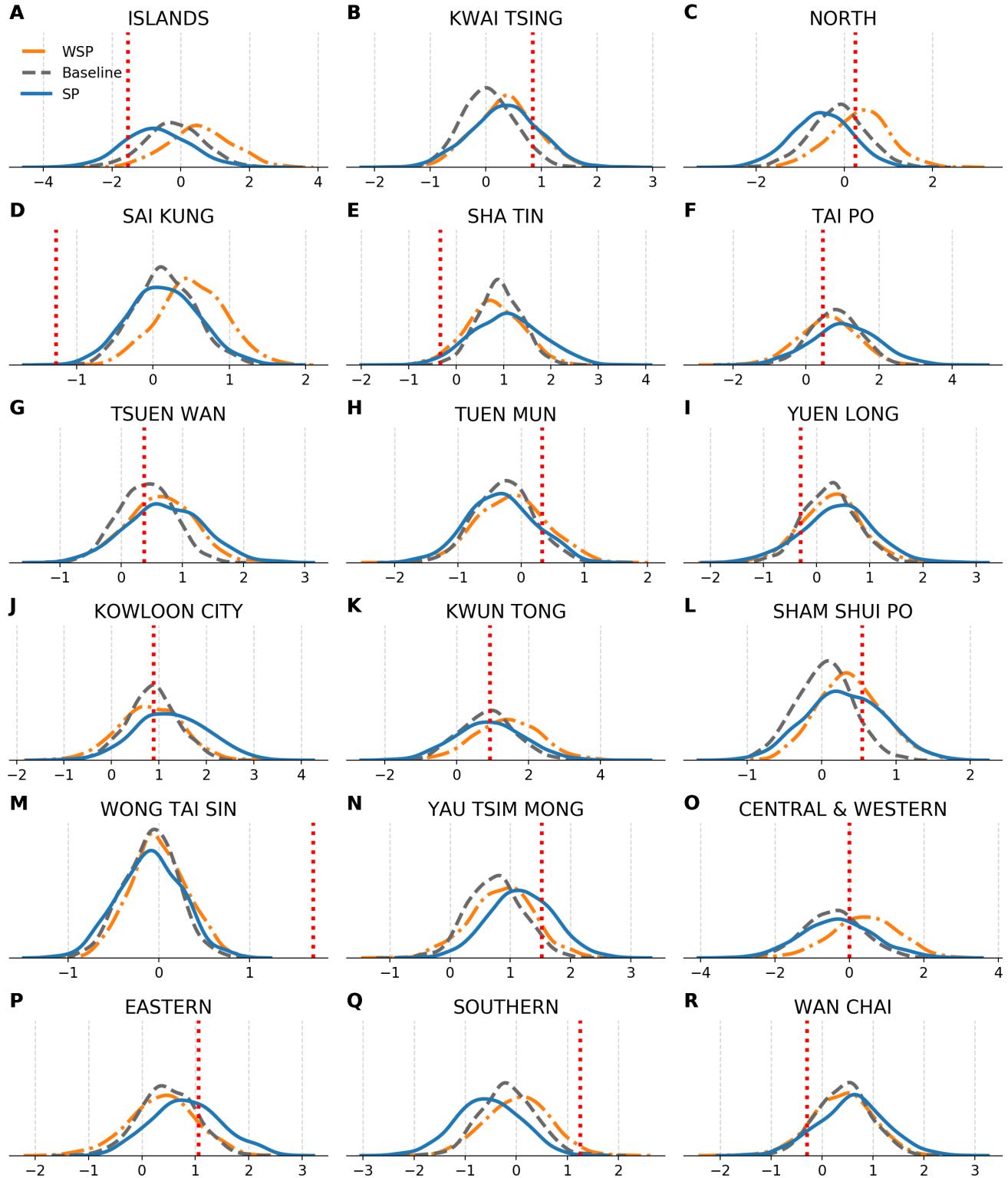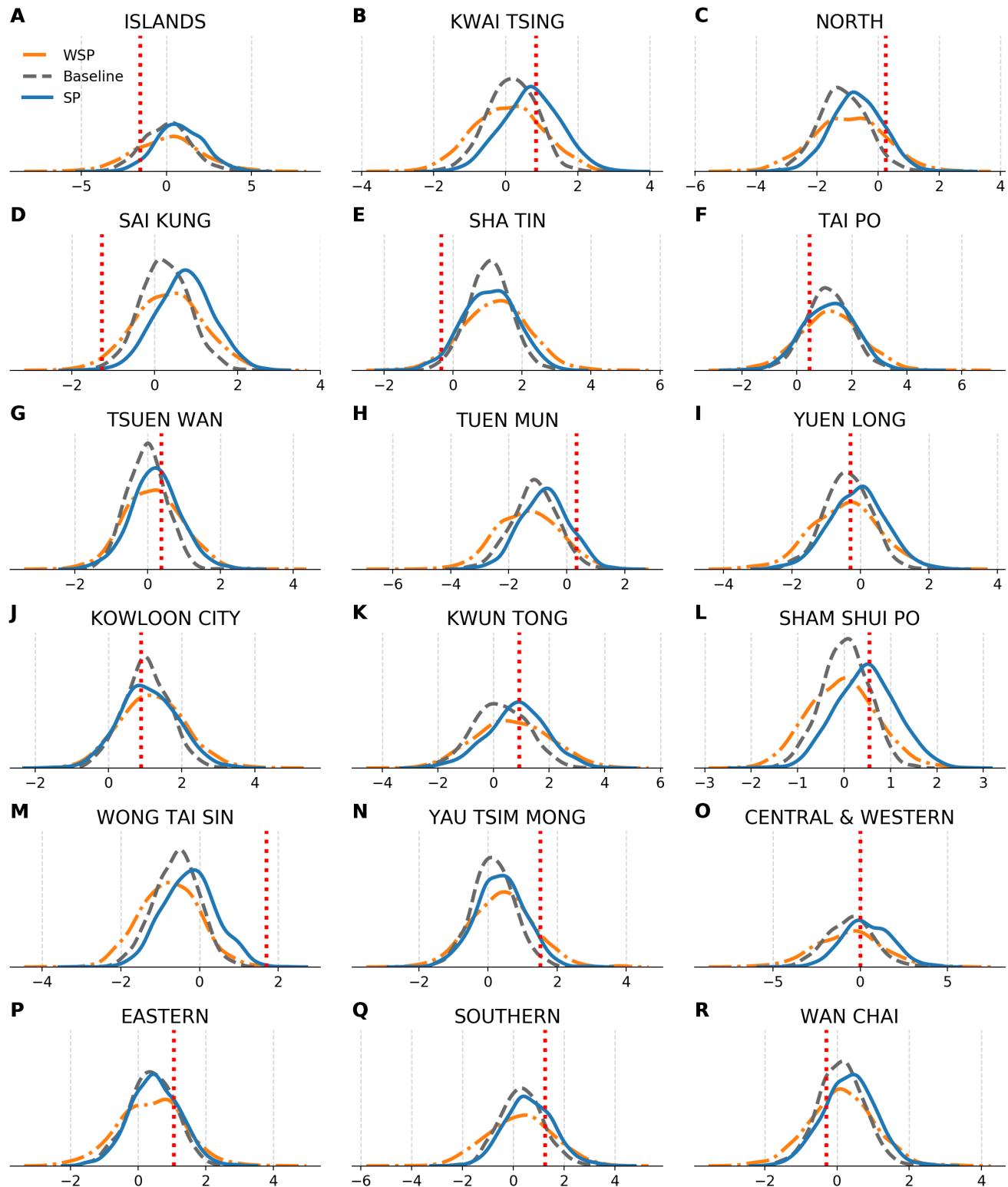
FIG. S8. **Distributions of β for the weighted spatial model.** We show the kernel density estimation of β coefficient for each feature in the design tensor. Distributions that are significantly above the mean are colored in orange while distributions significantly below the mean are colored in blue as measured by the 80% CI.

FIG. S9. **Distributions of γ for the weighted spatial model.** We show the kernel density estimation of γ coefficient for each feature in the design tensor. Distributions that are significantly above the mean are colored in orange while distributions significantly below the mean are colored in blue as measured by the 80% CI.

FIG. S10. **KDE of normalized mortality rates trained on the default base features for calendar year 2016**. We show the KDE the Baseline model (grey), Spatial model (blue), and Weighted Spatial model (orange). The red dashed line indicates the true normalized mortality rate for each district in 2016.
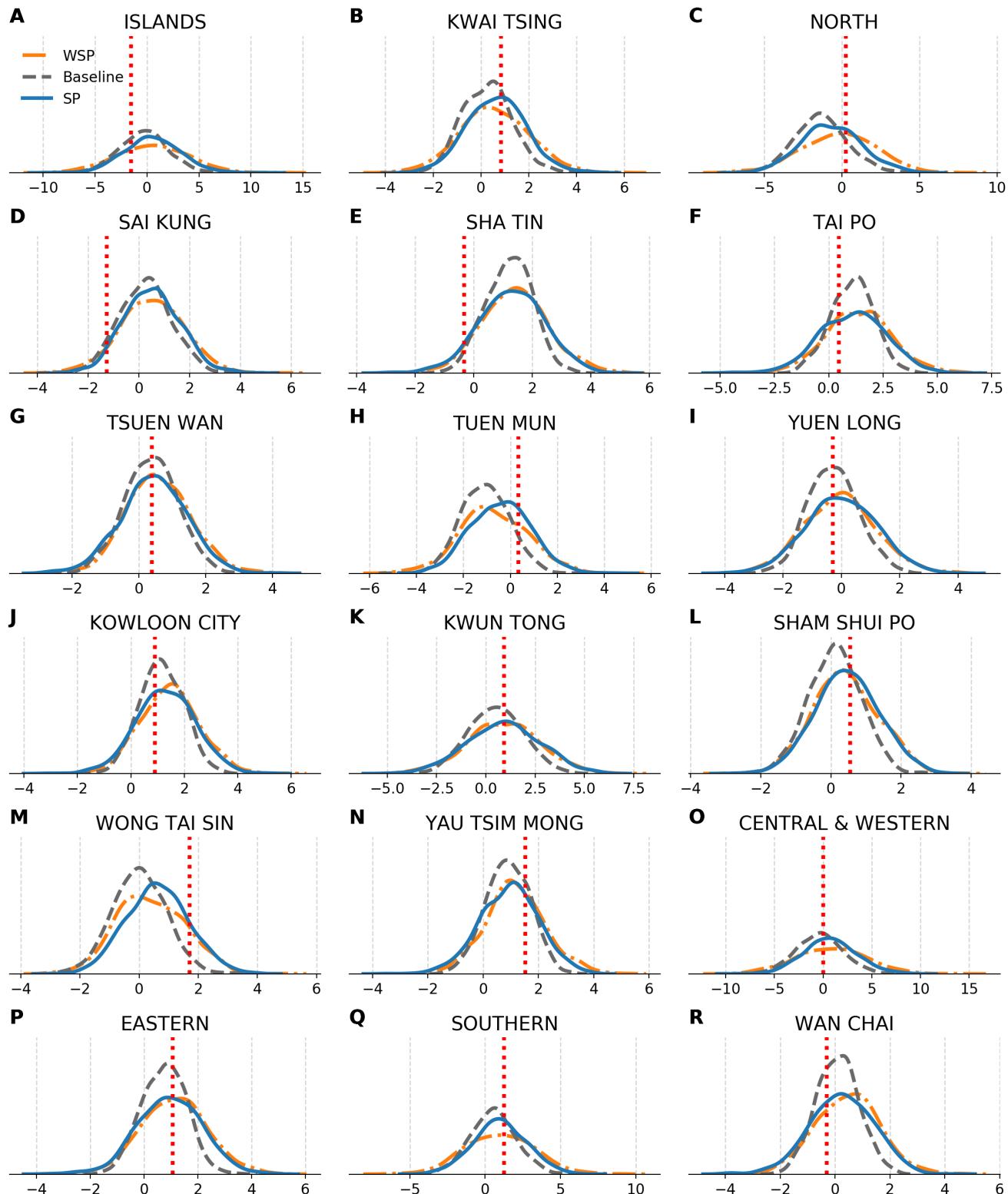
FIG. S11. **KDE of normalized mortality rates by model trained on wealth-related in addition to the set of base features for calendar year 2016**. We show the KDE the Baseline model (grey), Spatial model (blue), and Weighted Spatial model (orange). The red dashed line indicates the true normalized mortality rate for each district in 2016.
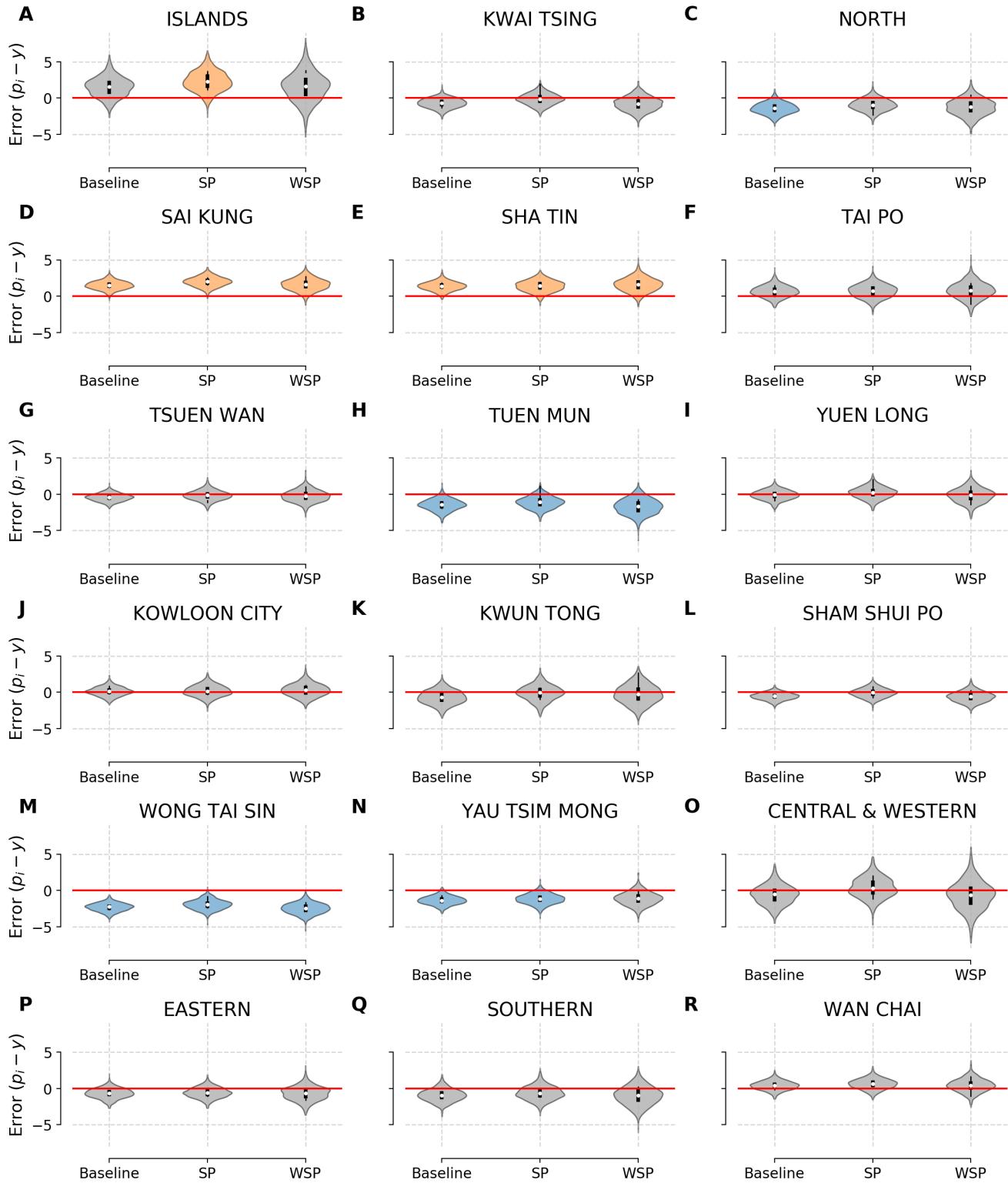
FIG. S12. **KDE of normalized mortality rates by model trained on all features for calendar year 2016**. We show the KDE the Baseline model (grey), Spatial model (blue), and Weighted Spatial model (orange). The red dashed line indicates the true normalized mortality rate for each district in 2016.

FIG. S13. **Systematic error distributions for models trained on the wealth-related set of features in 2016**. Models with significance systematic overestimation are colored in orange while models with significance systematic underestimation are colored in blue as measured by the 80% CI.
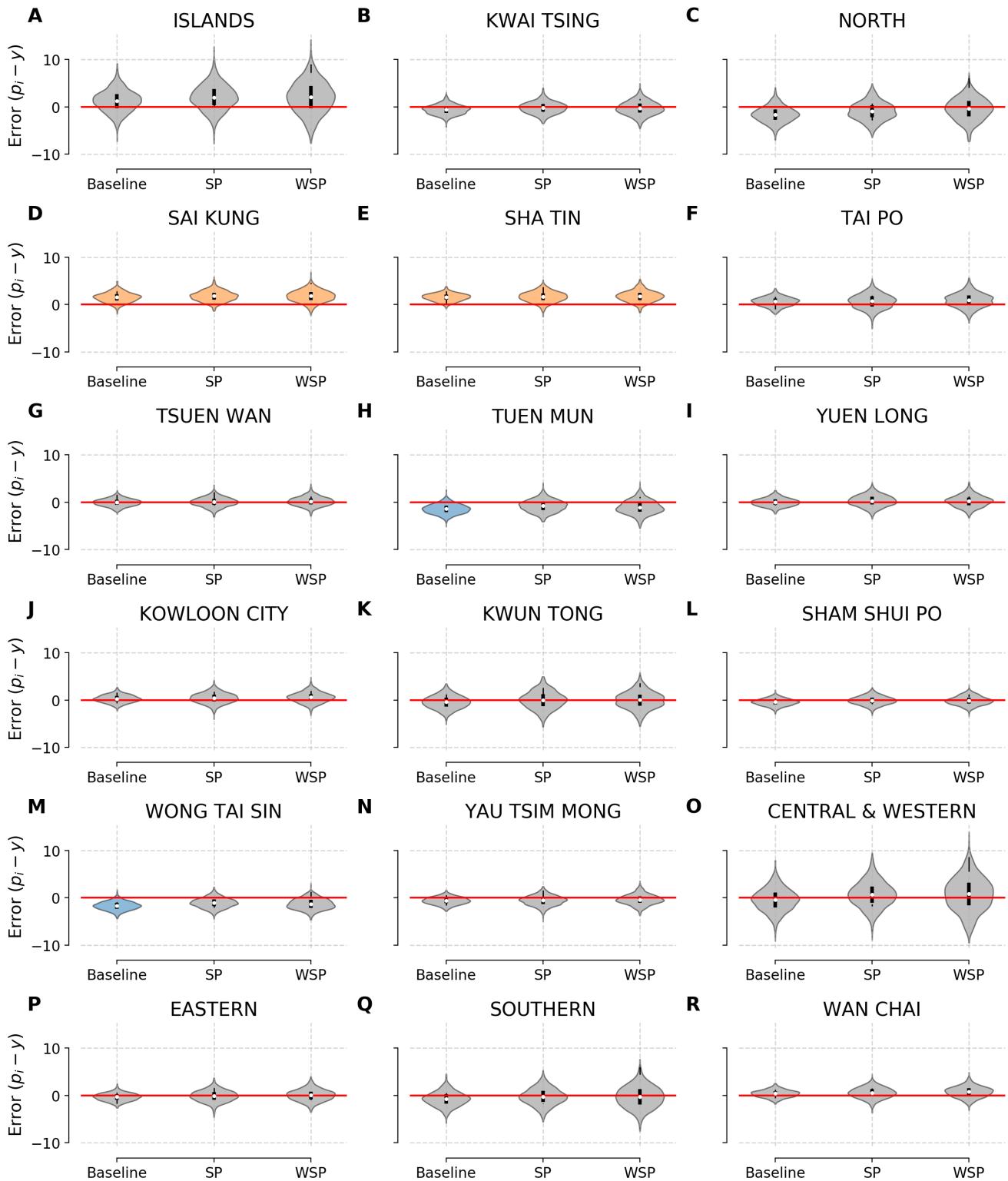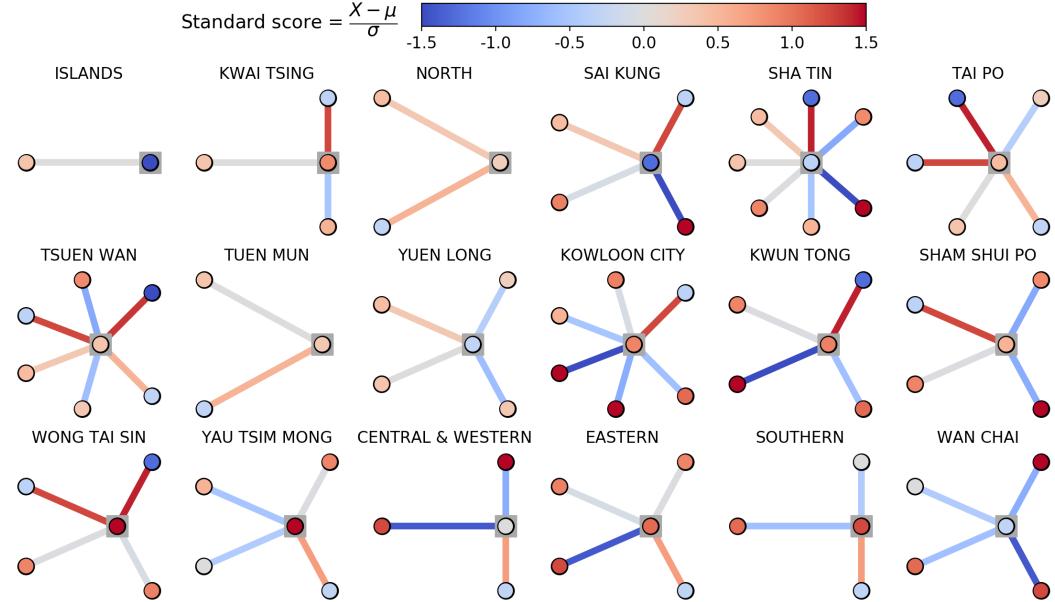
FIG. S14. **Systematic error distributions for models trained on all features in 2016**. Models with significance systematic overestimation are colored in orange while models with significance systematic underestimation are colored in blue as measured by the 80% CI.
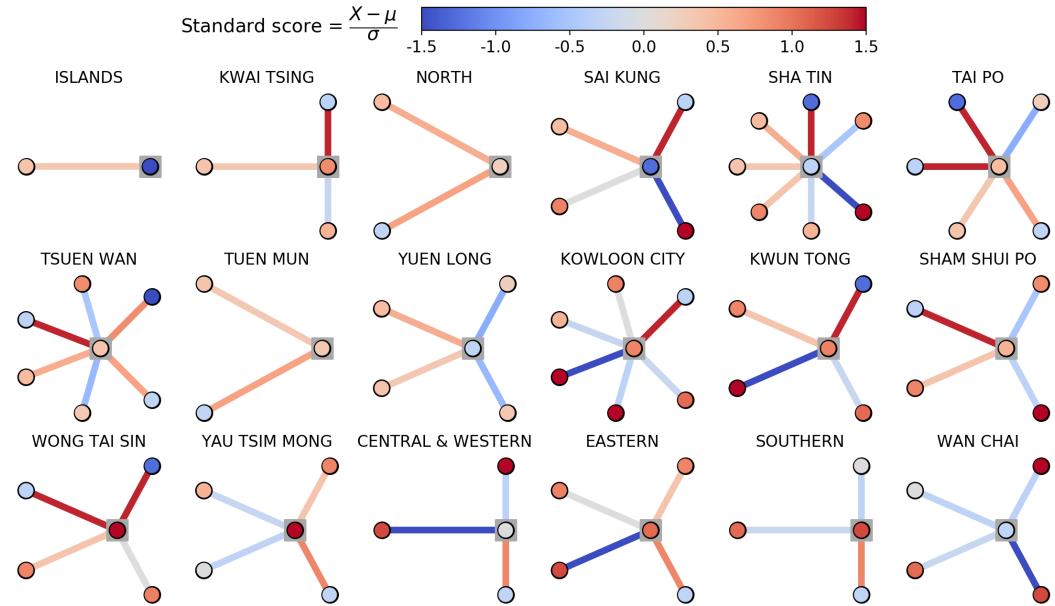
FIG. S15. **Ego networks of each district demonstrating sociospatial factors of mortality for the 2016 baseline model**. We display ego networks of each district in Hong Kong and its nearest neighbors in the road and bridge network. The central node (highlighted with a grey box) of each network corresponds to the labelled district. Neighbors are not arranged around the ego district geographically. Node color corresponds to normalized mortality rate and edge color corresponds to signed prediction error for the 2016 WSP model. These ego networks encode a qualitative measure of the sociospatial factors in mortality modeling.

FIG. S16. **Ego networks of each district demonstrating sociospatial factors of mortality for the 2016 spatial model**. We display ego networks of each district in Hong Kong and its nearest neighbors in the road and bridge network. The central node (highlighted with a grey box) of each network corresponds to the labelled district. Neighbors are not arranged around the ego district geographically. Node color corresponds to normalized mortality rate and edge color corresponds to signed prediction error for the 2016 WSP model. These ego networks encode a qualitative measure of the sociospatial factors in mortality modeling.