Diving into the consumer nutrition environment: a Bayesian spatial factor analysis of neighborhood restaurant environment

Hui Luan, Jane Law, Martin Lysy

PII: S1877-5845(16)30080-6
DOI: 10.1016/j.sste.2017.12.001
Reference: SSTE 232

To appear in: Spatial and Spatio-temporal Epidemiology

Received date: 11 October 2016
Revised date: 3 December 2017
Accepted date: 9 December 2017

Please cite this article as: Hui Luan, Jane Law, Martin Lysy, Diving into the consumer nutrition environment: a Bayesian spatial factor analysis of neighborhood restaurant environment, Spatial and Spatio-temporal Epidemiology (2017), doi: 10.1016/j.sste.2017.12.001

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Highlights

- This paper illustrates a Bayesian Spatial Factor Analysis of neighborhood restaurant environment;
- Uncertainties in the healthfulness index of neighborhood restaurant environment are quantified;
- Facilitator/barrier to healthy eating is most relevant to restaurant environment healthfulness;
- The healthfulness of neighborhoods without accessible restaurants is estimated.
**Diving into the consumer nutrition environment**: a Bayesian spatial factor analysis of neighborhood restaurant environment

**Abstract**

Neighborhood restaurant environment (NRE) plays a vital role in shaping residents’ eating behaviors. While NRE ‘healthfulness’ is a multi-facet concept, most studies evaluate it based only on restaurant type, thus largely ignoring variations of in-restaurant features. In the few studies that do account for such features, healthfulness scores are simply averaged over accessible restaurants, thereby concealing any uncertainty that attributed to neighborhoods’ size or spatial correlation. To address these limitations, this paper presents a Bayesian Spatial Factor Analysis for assessing NRE healthfulness in the city of Kitchener, Canada. Several in-restaurant characteristics are included. By treating NRE healthfulness as a spatially correlated latent variable, the adopted modeling approach can: (i) identify specific indicators most relevant to NRE healthfulness, (ii) provide healthfulness estimates for neighborhoods without accessible restaurants, and (iii) readily quantify uncertainties in the healthfulness index. Implications of the analysis for intervention program development and community food planning are discussed.

**Keywords**: neighborhood restaurant environment assessment; consumer nutrition environment; Bayesian inference; spatial modeling; factor analysis; community food planning.

---

1 The consumer nutrition environment represents in-store characteristics that consumers encounter when they reach a food retailer (Glanz et al., 2005)
Introduction

Neighborhood restaurant environment (NRE)\(^2\) is the place where residents can eat away from home or buy take-out foods. It has become an indispensable component in residents’ daily life. For example, in North America, Canadians and Americans spend over 25% and 50%, respectively, of their food expenditures on foods away from home (Statistics Canada, 2014; United States Development of Agriculture Economic Research Service, 2016). According to the report on Canada’s Restaurant Industry, over 35% Canadians rank eating out in a restaurant as their top preferred activity with friends and families, and over 60% Canadians eat out in restaurants at least once per week (Canadian Restaurant and Foodservices Association, 2010). In this context, NRE is playing a vital role in shaping residents’ eating behaviors, resulting in the development of numerous measures for assessing NRE healthfulness from researchers in multiple fields including public health, geography, and urban planning.

1.1. Evaluating neighborhood restaurant environment

Absolute restaurant density in a neighborhood, represented as total numbers of accessible restaurants (Jeffery et al., 2006; Polsky et al., 2016) or restaurant density per population or per area (Maddock, 2004; Mehta and Chang, 2008; Moore et al., 2009), is the most common measure for evaluating NRE. This measure has been extensively applied in public health studies exploring, for example, whether absolute densities of fast-food restaurants contribute to unhealthy eating and excess weights. Mixed findings, however, have been identified (Jeffery et al., 2006; Maddock, 2004; Mehta and Chang, 2008; Polsky et al., 2016), partly attributable to the application of absolute density measures that assess a single dimension of the multi-faceted NRE. While composite measures such as the ratio between unhealthy (e.g., fast-food) and healthy (e.g., full-service) restaurants have been used for NRE assessment (Mehta and Chang, 2008; Mercille et al., 2013; Polsky et al., 2016), such measures ignore restaurants that cannot be simply classified as unhealthy or healthy, which is predominantly the case for restaurants that are independently owned (as opposed to franchised or part of a chain). Furthermore, measures focusing on the community nutrition environment (e.g., restaurant types and numbers) fail to acknowledge differences between in-restaurant features such as availability of healthy eating options between restaurants of the same type in different neighborhoods. Additionally, in-restaurant features other than availability also have a role in defining NRE healthfulness. The presence of healthy eating options in restaurants does not necessarily guarantee a healthy NRE, given that higher prices of healthy eating options and barriers to healthy eating (e.g., overeating encouraged on the menu) could potentially prohibit consumers from making healthy consumption decisions (Hammond

\(^2\) Neighborhood is area-based rather than point-based (e.g., individual, household, or public institution) in this paper.
et al., 2013; Haws and Liu, 2016; Nordström and Thunström, 2015). These limitations are problematic either in studies exploring geographical disparities of NRE healthfulness, or in studies examining the association between NRE and diet-related outcomes in that a measure evaluating the partial rather than complete NRE is used.

Recently, in-store audit tools have been developed to assess restaurants. For example, the Nutrition Environment Measure Survey – Restaurant (NEMS-R) (Saëns et al., 2007) assesses in-restaurant features including availability, affordability, and facilitator/barrier of healthy eating, providing a composite measure of overall restaurant healthfulness. This tool allows to account for all restaurants and in-restaurant characteristics for assessing NRE healthfulness. However, the mean NEMS-R score per neighborhood is typically used for subsequent analyses (Duran et al., 2013; Wang et al., 2016), for example, exploring its association with neighborhood distress level. Although the mean NEMS-R score provides a simple and intuitive measure for assessing NRE, it suffers from a number of limitations. First, it masks the total number of accessible restaurants to a neighborhood as well as variations of in-restaurant features. Second, using the mean NEMS-R score to evaluate NRE healthfulness of a neighborhood ignores information of NRE in adjacent neighborhoods. In reality, people could travel beyond their own neighborhoods to procure foods, making it necessary to account for information of adjacent NRE to strengthen and stabilize the estimation (Luan et al., 2015). Third, the mean score does not reflect which in-restaurant feature contributes the most to (or are most relevant to) NRE healthfulness. Ignoring the difference of importance between in-restaurant features restricts the potential to inform food planning and interventions for promoting population-wide healthy eating.

1.2. Bayesian spatial factor analysis

To address the limitations associated with the mean NEMS-R score, we propose a Bayesian spatial factor analysis (BSFA) for assessing NRE healthfulness. Originating in psychometrics, factor analysis is a statistical approach that describes the relation between multiple co-dependent observable indicators with a small number of latent factors, i.e., which cannot be directly observed or measured (e.g., Brown, 2015). Conceptually, NRE healthfulness is abstract and unobservable, but manifests in the form of a number of NRE indicators (i.e., availability, affordability, facilitator/barrier, etc.). In this sense, factor analysis is well-suited to assess NRE healthfulness. For example, recognizing the correlation in terms of food provision and quality between different food outlet types, Michimi and Wimberly (2015) applied factor analysis to construct two factors representing the healthy and unhealthy dimensions of the food environment, respectively. In particular, Factor 1 consists of food outlets providing healthy options including supermarkets, snack/coffee shops, and full-service restaurants, while Factor 2 represents unhealthy food outlets including convenience stores and fast-food restaurants.
Traditional factor models applied in the spatial context make the highly unrealistic assumption of normally distributed and independent indicators in adjacent areas. To overcome such limitations, spatial factor analysis requires sophisticated latent variable models that typically preclude closed-form likelihoods. In such cases, Bayesian inference via Markov chain Monte Carlo (MCMC) sampling of the posterior parameter distribution is often the most viable estimation approach (Mari-Dell’Olmo et al., 2011; Wang and Wall, 2003). Additionally, it has been noted that for many hierarchical models, Bayesian approaches more readily account for parameter uncertainties than their Frequentist counterparts (Morris and Lysy, 2012; Rubin, 1981). BSFA has been applied in various fields in addition to psychology (Stakhovych et al., 2012), especially in estimating deprivation (Abellan et al., 2007; Congdon, 2016; Hogan and Tchernis, 2004; Mari-Dell’Olmo et al., 2011) and spatial and spatio-temporal common risk factors of mortalities and morbidities (Courtemanche et al., 2015; Lawson, 2013; Mezzetti, 2012; Tzala and Best, 2006; Wang and Wall, 2003). These studies demonstrate that BSFA is capable of quantifying uncertainties, tackling spatial autocorrelation, and assessing neighborhoods without observations. A recent study from Congdon (2016) exemplifies the only application in the literature that applies BSFA to assess the food environment. This study constructs a healthy food access index at the Metropolitan county level in the U.S., and then uses the index as a predictor for explaining geographical variations of obesity. The ratio between convenience stores and grocery stores was identified as the central indicator in defining healthy food access. Nevertheless, Congdon’s study does not take into account variances of food outlets’ in-store features, but focuses on outlet types only. Further, the proposed index was created at a relatively large-area (i.e., county) level; therefore, heterogeneity of the food environment is largely dissimulated.

1.3. Research questions

This study aims to answer two research questions:

(1) Which neighborhoods have the least healthy NRE (simultaneously suffer from deprived availability, affordability, and facilitator/barrier of healthy eating)?

BSFA is used to create a composite NRE index at the neighborhood level. Being a combination of weighted restaurant assessment indicators, this index reflects the underlying NRE ‘healthfulness’. Neighborhoods with an index value in the lowest quintile are identified as neighborhoods with least healthy NRE. Two metrics are applied for quantifying uncertainties associated with the composite index thus NRE healthfulness: one, the 95% credible interval (CrI)\(^3\) of the index; and two, the posterior probability of the index falling into the lowest quintile.

\(^3\) There is 95% probability that the posterior mean of the index occurs within the 95% credible interval.
(2) What is the indicator (availability, affordability, or facilitator/barrier) that contributes the most to (or most relevant to) NRE ‘healthfulness’?

Statistically, the indicator is the one with the highest factor loading on the composite NRE index. Its variance is also best explained by the NRE index. Overall, this specific indicator should be prioritized for interventions to improve NRE in the study region.

2. Study area and data

2.1 Study area

The analysis was conducted for the city of Kitchener at the dissemination area (DA) level. Kitchener is composed of 299 DAs and is located at the center of the Region of Waterloo, a municipality seated approximately one-hour west of Toronto. DA is the smallest census unit that covers the entirety of Canada (Statistics Canada, 2012). The population size of a DA generally ranges from 400 to 700. Considering the inconsistency of neighborhood definitions, we used DAs to represent neighborhoods, an approach that benefits policy implementation and planning because local governments have jurisdiction over administrative areas (Health Canada, 2012). Figure 1 displays the DA boundaries of Kitchener city and spatial distributions of restaurants (351 in total) in 2010. Generally, restaurants concentrate at downtown Kitchener along the arterial road (i.e., King Street). NEMS-R scores of restaurants accessible to Kitchener are presented with proportional dots.
Fig. 1: DA boundaries of Kitchener city and distributions of restaurants, 2010

2.2. Restaurant assessment indicators

Three correlated restaurant assessment indicators were used for constructing the composite index: availability of healthy eating options, affordability of healthy eating, and facilitators or barriers to healthy eating (hereafter called availability, affordability, and facilitator/barrier, respectively). These indicators were collected in 2010 based on adapted NEMS-R for Canadian food environment studies. More details are provided elsewhere (Minaker et al., 2013, 2014). Specifically, availability of a restaurant is assessed with a score that measures the availability of healthy food items such as main-dish salads with low calorie; affordability indicates the comparative pricing between ‘healthy’ and ‘unhealthy’ foods; and a score of facilitator/barrier reflects whether a restaurant includes measures for facilitating (e.g., providing nutrition information on the menu) or prohibiting (e.g., encouraging larger portions on the menu) healthy eating. For all three indicators, a restaurant with a higher score is deemed healthier. Scores of availability, affordability, and facilitator/barrier range from -1 to 21, -3 to 3, and -9 to 24, respectively (Table 1). We performed correlation analyses for the indicators using Spearman’s rho. Results indicate that availability, affordability, and facilitator/barrier
are significantly correlated. In particular, availability is positively associated with facilitator/barrier while affordability is negatively associated with availability and facilitator/barrier, suggesting that a restaurant with higher scores of availability and facilitator/barrier usually have a lower score of affordability (i.e., higher prices of healthy eating). In this sense, assuming that high availability is a positive contributor to NRE healthfulness, the three indicators used to construct the index represent high availability, low affordability, and high/low facilitator/barrier, respectively.

Table 1: Descriptive statistics of restaurant indicators for all restaurants accessible to Kitchener

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Availability</td>
<td>8.15</td>
<td>-1</td>
<td>21</td>
<td>4.64</td>
</tr>
<tr>
<td>Affordability</td>
<td>-1.48</td>
<td>-3</td>
<td>3</td>
<td>1.78</td>
</tr>
<tr>
<td>Facilitator/barrier</td>
<td>4.5</td>
<td>-9</td>
<td>24</td>
<td>6.46</td>
</tr>
</tbody>
</table>

A 1km road network buffer was created around each restaurant using ArcGIS 10.2 to identify neighborhoods that can access to this specific restaurant (i.e., the centroid of the neighborhood falls inside the buffering zone of the restaurant). The distance demarcation, 1km, represents a 10~15 mins walking distance, which has been widely used in Canadian food environment studies (Apparicio et al., 2007; Black et al., 2011; Larsen and Gilliland, 2008; Luan et al., 2016; Smoyer-Tomic et al., 2006). Another reason for choosing a walkable distance for NRE assessment is that active transportation including walking is essential for creating healthy communities and combating obesity epidemics (Ontario Professional Planners Institute, 2014). Our study aims to inform health planners to establish intervention programs for promoting healthy eating with active transportation modes especially walking. Therefore, we did not choose other (larger) buffering distances such as 4km and 8km (Luan et al., 2016). The number of neighborhoods that can access a specific restaurant ranges from 1 to 19, and the number of restaurants accessible to a neighborhood within 1km ranges from 0 to 87.

3. Statistical modeling

With BSFA, the unobservable concept, NRE healthfulness, can be inferred by multiple observable restaurant assessment indicators at the consumer nutrition environment level. In Model I, \( Y_{jk} \), the \( j^{th} \) restaurant indicator (normalized scores of availability, affordability, or facilitator/barrier) of the \( k^{th} \) restaurant, is assumed to follow a Normal distribution with mean \( \frac{1}{n_k} \sum_{i \in N_k} \mu_{ij} \) and variance \( \sigma_j^2 \) (Level 1, Table 2), where \( n_k \) is the number of neighborhoods whose centroids fall inside into the 1km buffering zone of the \( k^{th} \) restaurant; \( N_k \) is the ID set of the \( n_k \) neighborhoods; and \( \mu_{ij} \) the latent value of indicator \( j \).
at neighborhood \( i \). Specifically, \( Y_{jk} \) is connected to the neighborhood index \( i \) via: 1) the estimated \( \mu_{ij} \), which represents the expected mean of indicator \( j \) at neighborhood \( i \); and 2) \( N_k \), the IDs of the neighborhoods that can access restaurant \( k \). If neighborhood \( i \) can access the \( k \)th restaurant, its ID is included in \( N_k \). \( n_k \) averages the latent \( \mu_{kj} \) given that different restaurants can be accessed by different numbers of DAs.

In the food environment literature, the buffering and container approaches\(^4\) are most commonly used for characterizing food access (Charreire et al., 2010; Health Canada, 2012). The former one is preferred since it could also account for outlets that are just outside the neighborhood boundary thus reduces the “edge effect” and improves the accuracy of food access estimation (Sadler and Gilliland, 2011). While the container approach links a food outlet to a unique DA so the convolution-based model (Calder and Cressie, 2007) or the method used by Mugglin et al. (2000) can be applied, buffering from a DA’s centroid and counting the number of outlets within the buffering zone however, is problematic for subsequent statistical modeling from a data-generating perspective. An outlet (and associated assessment indicators, \( Y_{jk} \)) accessible to multiple DAs is used for multiple times to estimate the latent \( \mu_{kj} \) at different neighborhoods \( i \), resulting in a unique value of \( Y_{jk} \) corresponding to multiple different models. To avoid this statistical issue, we buffer from each restaurant instead (a.k.a. the coverage or service area approach, which has been applied in Canadian studies, for example, Larsen and Gilliland, 2008). The assessment indicator of a specific restaurant, \( Y_{jk} \), is associated with the latent values of the neighborhoods it serves in a unique model (Level 1, Table 2).

We further decompose \( \mu_{ij} \) into an intercept \( \alpha_j \) (the average of indicator \( j \) over the study region), a product of factor loading \( \delta_j \) (the loading of indicator \( j \) on the index) and index \( \theta_i \) (restaurant environment index at neighborhood \( i \)), and indicator-specific random noise \( \epsilon_{ij} \) (Level 2, Table 2). Notably, several neighborhoods do not have direct access to any restaurant within a walkable distance such that corresponding \( \mu_{kj} \) are not connected directly to the data at Level 2. Their composite index \( \theta_i \) however, can be imputed by specifying an intrinsic Conditional Autoregressive (ICAR) distribution (Besag et al., 1991) to \( \theta_i \). Specifically, \( \theta_i \) follows a normal distribution with conditional mean that equals to the average of neighboring \( \theta_m \)’s and conditional variance that is inversely proportional to the number of neighbors, \( n_i \). For reference, two areas are defined as neighbors if they share at least one common vertex, a common approach used in spatial statistical studies (Law et al., 2013). Note that \( w_{im} = 1 \) if DA \( i \) and DA \( m \) are neighbors; otherwise, \( w_{im} = 0 \). Under the ICAR distribution, NRE healthfulness of a neighborhood

\(^4\) The buffering approach counts the number of food outlets within a neighborhood’s buffering zone or the number of neighborhoods intersected with a food outlet’s buffering zone, while the container approach calculates the number of food outlets within the administrative boundary of a neighborhood.
without accessible restaurants is estimated by ‘borrowing information’ from adjacent neighborhoods.

Table 2: Model specification

<table>
<thead>
<tr>
<th>Level 1</th>
<th>( Y_{jk}^{ind} \sim \text{Normal}\left(\frac{1}{n_k} \sum_{i \in N_k} \mu_{ij}, \sigma_j^2\right) )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \mu_{ij} = \alpha_j + \delta_j \theta_i + \epsilon_{ij} + \varphi_{ij} ) where ( \epsilon_{ij}^{ind} \sim \text{Normal}(0, \sigma_{\epsilon_j}^2) )</td>
</tr>
<tr>
<td>Level 2</td>
<td>( \theta_i</td>
</tr>
<tr>
<td></td>
<td>Model I: ( \varphi_j \equiv 0 )</td>
</tr>
<tr>
<td></td>
<td>Model II: ( \varphi_j</td>
</tr>
<tr>
<td></td>
<td>( \alpha_j^{id} \sim \text{Lebesgue}(-\infty, \infty) )</td>
</tr>
<tr>
<td></td>
<td>( \log(\delta_j) \sim \text{Normal}(0, 100) ) ( \delta_j^{id} \sim \text{Normal}(0, 1000), j = 2, 3 )</td>
</tr>
<tr>
<td>Prior</td>
<td>( \sigma_{\theta_j}^2 = 1 )</td>
</tr>
<tr>
<td></td>
<td>( \frac{1}{\sigma_j^2}, \frac{1}{\sigma_{\epsilon_j}^2}, \frac{1}{\sigma_{\varphi_j}^2} \text{ (Model II only) } \sim \Gamma(0.5, 0.0005) )</td>
</tr>
</tbody>
</table>

To estimate the parameters of Model I, we employed a Bayesian MCMC sampling approach, which begins by specifying priors on all the model parameters. Priors specified to the unknown parameters are provided in Table 2. In particular, to avoid the “flip-flop” problem (\( \delta_j \theta_i = (-\delta_j)(-\theta_i) \)) and allow feasible identification, we constrained \( \delta_j \) to be positive. Similar approaches have been applied in past studies (Abellán et al., 2007; Congdon, 2016; Marí-Dell’Olmo et al., 2011). Also for identification purposes, the variance of \( \theta \) (denoted as \( \sigma_{\theta_j}^2 \)) is set to 1, equivalent to index standardization (Skrondal & Rabe-Hesketh, 2007). To test whether the spatial structure of restaurant assessment indicators is adequately captured by \( \theta_i \), we also fitted a model (Model II) by including a spatial random effect (\( \varphi_j \)) at Level 2. Similarly, an ICAR distribution with variance \( \sigma_{\varphi_j}^2 \) was specified to \( \varphi_j \).

In addition to the unknown parameters in the model, we also monitored the posterior probability that \( \theta_i \) falls inside the lowest quintile (denoted as \( PP_{\theta_i} \)) as a measure
for identifying neighborhoods that have least healthy NRE. Complementary to the point estimate of \( \theta_i \) (i.e., posterior mean), \( PP_{\theta_i} \) quantifies the uncertainty associated with \( \theta_i \) via taking into account the sampling variance of \( \theta_i \) and making use of the full posterior distribution of \( \theta_i \) (Richardson et al., 2004). Each neighborhood was given a binary indicator at each iteration (one if \( \theta_i \) falls into the lowest 20%; otherwise zero). \( PP_{\theta_i} \) is the fraction of one’s of all iterations. The higher the value of \( PP_{\theta_i} \), the stronger evidence that neighborhood \( i \) has a least healthy NRE.

To determine which restaurant indicator is most relevant to NRE healthfulness, we examined (i) the magnitude of factor loading, \( \delta_j \), and (ii) a proxy for the usual ‘proportion of variance explained’ statistic, estimated here as the ratio \( (\rho_j) \) of the empirical variance of \( \theta_i \) (denoted as \( s_{\theta_i}^2 \)) and an estimate of the total variance: \( \rho_j = \frac{s_{\theta_i}^2}{s_{\theta_i}^2 + \sigma_{e_i}^2} \) (Abellan et al., 2007). Higher values of \( \delta_j \) and \( \rho_j \) suggest stronger relevance between the restaurant indicator and NRE healthfulness.

Both models were fitted in WinBUGS (Lunn et al., 2000) with two parallel chains. Trace plots, history plots, autocorrelation plots, and Gelman-Rubin plots were visually examined for checking convergence. Models converged after 50,000 iterations. We ran each chain for another 100,000 iterations and retained every 10th sample, resulting in an acceptable Monte Carlo error (<5% of sample posterior deviation). A final 20,000 samples were obtained for posterior estimations. Model I and Model II were compared based on Deviance Information Criteria (DIC) (Spiegelhalter et al., 2002). The model favored by the data was the one with lower DIC. We conducted sensitivity analysis by specifying a prior of Uniform(0, 100) directly to variance parameters \( (\sigma_j^2 \) and \( \sigma_e^2 \)) in the model with lower DIC. Similar results were obtained and DIC difference was smaller than 5, indicating that inferential results were essentially insensitive to prior selections. The WinBUGS code is provided in the Appendix.

4. Results

Table 3 shows the values of DIC and pD (effective parameters) from the two fitted models. Although Model II has a higher pD, the DIC difference is only 2937.88 - 2937.36 = 0.52, indicating that the two models fit the dataset equally well. Thus, the parsimonious Model I was chosen as the final model. We report below results from Model I.
Table 3: DIC and pD values from two fitted models

<table>
<thead>
<tr>
<th>Model</th>
<th>DIC</th>
<th>pD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model I: Without spatial residuals (φ_i)</td>
<td>2937.88</td>
<td>39.004</td>
</tr>
<tr>
<td>Model II: With spatial residuals (φ_o)</td>
<td>2937.36</td>
<td>42.24</td>
</tr>
</tbody>
</table>

4.1. Factor loadings

Loadings of availability, affordability, and facilitator/barrier on the common factor (the composite index θ_i) are presented in Table 4. All three indicators are significantly associated with the composite index since the 95% CrI of the factor loadings do not cover zero, suggesting that each indicator is a meaningful manifestation of the underlying concept – the ‘healthfulness’ of NRE. Facilitator/barrier (1.036, 95% CrI: [0.525, 1.715]) has the highest magnitude of loading factor, followed by availability (0.823, 95% CrI: [0.321, 1.443]) and affordability (-0.675, 95% CrI: [-1.127, -0.280]). While availability and facilitator/barrier are positively associated with the index, a negative association was found between affordability and the composite NRE index, indicating that the low affordability as noted above is discounting NRE healthfulness. Values of the ratio, ρ_j (a variance-explained statistic as noted above), for availability, affordability, and facilitator/barrier are 0.955, 0.906, and 0.980, respectively, which are in agreement with the factor loadings of each indicator on the index. These high values of ρ_j suggest that the variances of the estimated availability, affordability, and facilitator/barrier for each neighborhood are well explained by the composite index θ_i.

Table 4: Factor loadings (δ_j) and the ratio (ρ_j) from Model I

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Factor loading (95% CrI)</th>
<th>Proportion Variance Explained (ρ_j)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Availability of healthy eating option</td>
<td>0.823 (0.321,1.443)</td>
<td>0.955</td>
</tr>
<tr>
<td>Affordability of healthy eating</td>
<td>-0.675 (-1.127, -0.280)</td>
<td>0.906</td>
</tr>
<tr>
<td>Facilitator/barrier of healthy eating</td>
<td>1.036 (0.525, 1.715)</td>
<td>0.980</td>
</tr>
</tbody>
</table>

4.2. Posterior estimations of the composite NRE index

Posterior means and 95% CrI of the composite index θ_i, which represents NRE healthfulness, are plotted in Figure 2. Varying NRE ‘healthfulness’ is observed among neighborhoods. Notably, neighborhoods with similar posterior means (shown in red dots) could have different 95% CrI thus associated with different degrees of uncertainty in
identifying neighborhoods with least healthy NRE. Such uncertainties are also reflected by the fraction of the 95% CrI that falls within the lowest quintile (Fig.2). Posterior means of $\theta_i$ are further mapped (Fig.3a). Four distinct clusters of neighborhoods locating at west, northwest, north, and northeast Kitchener are identified as having least healthy NRE. These neighborhoods simultaneously suffer from deprived availability, affordability, and facilitator/barrier, or in other words, lower relative availability of healthy eating options, higher relative prices of healthy eating, and higher/lower levels of facilitators/barriers to healthy eating.

We also map the posterior probability of $\theta_i$ that falls inside the lowest quintile, $PP_{\theta_i}$ (Fig.3b). Following the approach of Marí-Dell’Olmo et al. (2011) for classifying deprivation and considering that the maximum of $PP_{\theta_i}$ is 0.634, we categorized $PP_{\theta_i}$ into three groups, representing neighborhoods that ‘probably suffer from least healthy NRE’ ($PP_{\theta_i} > 0.5$), ‘probably do not suffer from least healthy NRE’ ($0.05 < PP_{\theta_i} \leq 0.5$), and ‘have low probability of least healthy NRE’ ($PP_{\theta_i} \leq 0.05$). Two clusters of neighborhoods locating at west and towards northwest Kitchener as well as several neighborhoods scattering across the region are identified as ‘probably suffer from least healthy NRE’. These neighborhoods all fall inside the lowest quintile based on the posterior mean of $\theta_i$ (Fig.3a). Compared with their counterparts in the same quintile, they have a NRE that is more likely to be least healthy, which, again, shows the unreliability of using a point estimate (i.e., the posterior mean) to evaluate NRE healthfulness.

![Figure 2. Caterpillar plot of the posterior mean and 95% credible interval of composite index ($\theta_i$)](image-url)
Figure 3. Quantile map of (a) the composite NRE index ($\theta_i$) and (b) the posterior probability of $\theta_i$ falling into the lowest quintile ($PP_{\theta_i}$).

5. Discussion

5.1. Dissecting uncertainties associated with descriptive measures for quantifying NRE ‘healthfulness’

As noted above, using just the mean NEMS-R score to quantify NRE healthfulness ignores the variability associated with this statistic, and thus has limited ability to address the following questions.

- Do two neighborhoods with the same mean NEMS-R score but different numbers of accessible restaurants have the same level of healthfulness (scenario A)?
- Is a neighborhood with higher mean NEMS-R score but lower number of accessible restaurants necessarily healthier than a neighborhood with lower mean NEMS-R score but higher number of accessible restaurants (scenario B)?
- Which neighborhood of the two has a healthier NRE: a neighborhood without accessible restaurants or a neighborhood with accessible restaurants that have low scores of availability, affordability, and facilitator/barrier (scenario C)?

On the other hand, these questions can all be addressed using the full posterior distribution of the parameters. The mean NEMS-R score of accessible restaurants for each neighborhood is mapped in Figure 4. Neighborhoods without accessible restaurants
are highlighted with hatch lines. We also highlight and label three groups of neighborhoods corresponding to the three scenarios, and demonstrate how the applied BSFA approach quantifies the aforementioned uncertainties.

Figure 4. Mean NEMS-R scores at the dissemination area level, 2010

Table 5 presents the descriptive statistics and posterior estimates for selected neighborhoods. Under scenario A, the mean NEMS-R scores of neighborhoods A₁ and A₂ are the same (14); however, the posterior means for A₁ and A₂ are -0.185 and 0.245, respectively. This difference is not surprising given that the estimations incorporate NEMS-R information from adjacent neighborhoods, which are usually different, thus enabling the differentiation between two neighborhoods with the same mean NEMS-R score. ‘Borrowing strength’ from neighbors is reasonable since it strengthens NRE healthfulness assessment via accounting for the possibility that residents could walk beyond their own neighborhoods (Luan et al., 2015). Furthermore, the uncertainty associated with varied total number of accessible restaurants is reflected by the 95% CrI of the index. The index of neighborhoods with smaller numbers of accessible restaurants usually has a wider 95% CrI range. For example, the range for A₁ (only one accessible restaurant) is 1.613 (=0.621+0.992), wider than that (1.076, =0.255+0.821) of A₂ (5 accessible restaurants), suggesting that there is greater uncertainty associated with the assessment for A₁. With smaller sample size (i.e., observed accessible restaurants) providing limited information, the posterior estimation is largely determined by the prior distribution, which in our case is vague, leaving the posterior estimation with a wide 95% CrI.
Under scenario B, neighborhood B1 has a higher mean NEMS-R score (12) than B2 (9.94). However, the former can access to one restaurant only while the latter 87. Not surprisingly, the 95% CrI for B1 is wider than that of B2 (1.529 versus 1.377), indicating greater uncertainty of NRE healthfulness assessment for B1 for the same reason as mentioned above. Interestingly, B1 has a lower posterior mean of the composite index $\theta_i$ and a higher $PP_{\theta_i}$ than B2 (-0.291 versus -0.270 and 0.324 versus 0.276, respectively), suggesting that B1 has a less healthy NRE although its mean NEMS-R score is higher.

Comparing the posterior estimations of C1 and C2, we found that neighborhoods without access to restaurants do not necessarily have a lower composite index or a higher $PP_{\theta_i}$ (i.e., NRE is more likely to be least healthy) compared to neighborhoods with accessible restaurants. Nevertheless, according to the 95% CrI of $\theta_i$, greater uncertainties are associated with the posterior estimation for neighborhoods without accessible restaurants. Additional comparison between C2 and C3 (both do not have access to restaurants) highlights the impact of spatial lag 5 on posterior estimations for neighborhoods without access to restaurants. Neighborhoods with a higher spatial lag have a higher composite index and are less likely to have least healthy NRE (i.e., lower $PP_{\theta_i}$).

Table 5: Dissecting uncertainties under different scenarios

<table>
<thead>
<tr>
<th>ID</th>
<th># of accessible restaurants</th>
<th>Mean NEMS-R score</th>
<th>Spatial lag</th>
<th>Posterior mean (95% CrI)</th>
<th>$PP_{\theta_i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scenario A: Neighborhoods with the same mean NEMS-R score but different numbers of accessible restaurants</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td>1</td>
<td>14</td>
<td>11.64</td>
<td>-0.185 (-0.992, 0.621)</td>
<td>0.247</td>
</tr>
<tr>
<td>A2</td>
<td>5</td>
<td>14</td>
<td>18.59</td>
<td>0.245 (-0.255, 0.821)</td>
<td>0.003</td>
</tr>
<tr>
<td><strong>Scenario B: Neighborhoods with low number of accessible restaurants have higher mean NEMS-R score than neighborhoods with higher number of accessible restaurants</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>1</td>
<td>12</td>
<td>7.56</td>
<td>-0.291 (-1.068, 0.461)</td>
<td>0.324</td>
</tr>
<tr>
<td>B2</td>
<td>87</td>
<td>9.94</td>
<td>10.51</td>
<td>-0.270 (-0.967, 0.410)</td>
<td>0.276</td>
</tr>
<tr>
<td><strong>Scenario C: Neighborhoods with and without accessible restaurants</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>2</td>
<td>-0.5</td>
<td>8.54</td>
<td>-0.087 (-0.834, 0.658)</td>
<td>0.163</td>
</tr>
<tr>
<td>C2</td>
<td>0</td>
<td>NA a</td>
<td>13.98</td>
<td>0.352 (-0.522, 1.227)</td>
<td>0.037</td>
</tr>
<tr>
<td>C3</td>
<td>0</td>
<td>NA a</td>
<td>2</td>
<td>-0.226 (-1.57, 1.107)</td>
<td>0.359</td>
</tr>
</tbody>
</table>

a NA: not available

5 Spatial lag refers to the mean NEMS-R score in adjacent neighborhoods.
5.2. From community to consumer nutrition environment: policy and planning implications

Findings from our study are informative for developing in-restaurant feature based interventions for planning and improving NRE. Past restaurant interventions are predominantly implemented at the community nutrition environment level, for example, banning the construction of fast-food restaurants to encourage establishments of restaurants with more healthy eating options (Mair et al., 2005; Stephens, 2007), probably attributable to the lack of primary consumer nutrition environment data that support sound spatial statistics for NRE assessment. Our modeling however, provides information in terms of what indicator to prioritize and where the interventions should be targeted.

In general, our findings suggest that restaurants in Kitchener city should increase the availability of healthy eating options, and decrease prices and barriers (or increase facilitators) of healthy eating because all three indicators are meaningful manifestations of NRE healthfulness (Table 4). Increasing/decreasing facilitator/barrier could be an intervention priority in that facilitator/barrier is most relevant to NRE healthfulness (i.e., highest values of factor loading ($\delta_j$) and the ratio ($\rho_j$), Table 4). This finding suggests that interventions such as implementing the regulation of menu labeling (e.g., labeling calorie, nutrient, and sodium) in Kitchener’s restaurants could potentially be effective for improving NRE healthfulness and promoting population-wide healthy eating. Mandatory menu labeling regulations have been implemented in several U.S. cities including New York City (Dumanovsky et al., 2010), but not in Ontario until January 01, 2017 (Ontario’s Regulatory Registry, 2016). Nevertheless, menu labeling has been found effective in reducing calorie and sodium intake and increasing awareness of healthy eating in the Region of Waterloo (Hammond et al., 2013) and other Canadian contexts (Girz et al., 2012; Scourboutakos et al., 2014; Vanderlee and Hammond, 2014). Such labeling regulations might need to couple with additional interventions, for example, removing barriers to healthy eating, to take effect since multiple facilitators and/or barriers could interact to impact eating behaviors (Haws and Liu, 2016). Interestingly, regulating calorie and nutrition-labeling also has the potential to affect other indicators of NRE healthfulness, for example, motivating restaurants to provide more healthy eating options (Namba et al., 2013) and improve signage for promoting healthy eating (Saelens et al., 2012), as evidenced by recent studies.

Neighborhoods with least healthy NRE (lowest quintile with darkest color in Fig.3a) should be prioritized for interventions in availability, affordability, and facilitator/barrier because they simultaneously suffer from these indicators as explained above. If resources are limited, priorities should be placed on the neighborhoods with higher $PP_q$ (i.e., > 0.5) (Fig.3b), where stronger evidence of least healthy NRE is present. The identification of neighborhoods with least healthy NRE is beneficial for community
food planning, which has recently emerged as a tool for improving the food environment and facilitating healthy eating. Although planners cannot control the food prices and what to sell in restaurants (Minaker et al., 2011), they can greatly contribute to restaurant environment improvement via zoning and licensing regulations. For example, Raja et al. (2008) suggested that fast-food restaurants should be required to provide a ‘healthy offerings check’, which certifies that healthy foods will be offered, from the local public health agency in the licensing process, when they are applying for a food establishment permit. In a similar fashion for Kitchener, municipalities and planners could request checks for availability, affordability, and facilitator/barrier from pending restaurants that are accessible to the neighborhoods with least healthy NRE (Fig. 3), ensuring that the new establishments could improve the NRE or at least maintain the healthfulness level in specific neighborhoods.

Interventions for neighborhoods without access to restaurants within a walkable distance, especially those with a high estimated composite index and low $PP_0$, require special attentions. The population density of these neighborhoods (64 in total; areas with hatch pattern, Figure 4) ranges from 87.86 to 5553.85 per km$^2$ (median: 2855.86), indicating that restaurants are inaccessible by walking to a substantial amount of residential neighborhoods in Kitchener. This inaccessibility probably results from zoning ordinances that prohibit the establishment of food outlets in residential neighborhoods or within a pre-designated distance (Black et al., 2011; Raja et al., 2008). The NRE healthfulness for these neighborhoods is estimated by pooling information from adjacent neighborhoods, which is usually associated with high uncertainties as noted above. While this approach is reasonable from a spatial statistical perspective since food access is a continuous phenomenon (Charreire et al., 2010), the estimation might not reflect the underlying needs of people residing in these neighborhoods, especially in the context that active transportation such as walking could be potentially effective for facilitating physical activity, thus reducing obesity rates. Future research surveying residents’ interests and desires in dining away from home within a walkable distance is warranted and survey results could be incorporated in the community food planning process for these neighborhoods.

5.3. Study strengths and limitations

Our research has several notable strengths. Instead of focusing on a proportion of restaurants such as fast-food restaurants and full-service restaurants, this study analyzes all restaurants, franchised, chain, or independent, in the study region. The analysis gives a holistic and more nuanced picture of NRE in Kitchener, which is essential for accurately targeting neighborhoods for interventions. In addition, rather than concentrating on the community nutrition environment, we explore the consumer nutrition environment. Compared with other measures based on restaurant types (e.g., of the density of fast-food
restaurants), the composite index constructed in this paper could be more meaningful and useful for determining NRE healthfulness, and evaluating opportunities for procuring and consuming healthy foods away from home, given that affordability and facilitator/barrier also influence residents’ eating behaviors other than availability as noted above. Finally, to our knowledge, our study is the first of its kind to analyze spatial patterns of NRE ‘healthfulness’ with in-restaurant indicators using a robust spatial statistical approach. This modeling approach advances the understanding of NRE by providing a more reliable measure of NRE healthfulness, which quantifies uncertainties associated with NRE assessment and could benefit food planning and interventions.

Several limitations of this study should be acknowledged. First, we used geographic centroids rather than population centroids to determine whether a DA has access to a specific restaurant, although DA is relatively small such that geographic centroids and population centroids could be very close. Second, the uncertainty of NRE healthfulness assessment might be greater for periphery neighborhoods where the estimation cannot borrow strength from adjacent neighborhoods (which are outside Kitchener and not included in our dataset). Third, only availability, affordability, and facilitator/barrier are used to construct the composite index. Beyond these in-restaurant features collected via NEMS-R, additional consumer nutrition environment indicators could be incorporated in the model to refine the index. For example, when the index is intended to reflect NRE healthfulness for a specific group of population (e.g., Chinese, vegan, etc.), availability of culturally acceptable healthy foods should be included. Lastly, exploring the spatial patterns of NRE healthfulness is inherently exploratory. Socio-economic and socio-demographic environments should be incorporated into future NRE assessment in that residents with similarly healthful NRE but different socio-economic status could experience disparate eating patterns.

5.4. Future research

Future research could apply the proposed approach to the whole Region of Waterloo and other cities inside and outside Canada for assessing the healthfulness of NRE or the entire retail food environment. The derived composite NRE index could be further tested in terms of its usefulness for explaining geographical disparities of eating behaviors or diet-related health outcomes. The proposed approach is also useful for validating other indicators purported to measure the healthfulness of restaurants or food stores, especially given that increasing indicators are available for food environment measurement but validation approaches are lacking (Minaker et al., 2014).

Additionally, future research could analyze dynamic NRE healthfulness via spatio-temporal factor analysis by incorporating a temporal dimension. Availability, affordability, and facilitator/barrier change over short-term temporal scales including hours and weekdays due to restaurant opening-hour variations, and over long-term...
temporal scales including seasons and years attributable to the opening and closing of restaurants. Yet spatio-temporal analyses of the NRE require repeated assessment of in-restaurant features, which is costly and time-consuming. Alternative assessment tools, for example, the reduced-item audit tools (Partington et al., 2015) and mobile phone applications (Kanter et al., 2014) could be applied for rapid data collection in future research. Such spatio-temporal analyses could also be computationally challenging, for which fast but approximate inference methods for latent factor models, for example, the Integrated Nested Laplace Approximation approach (Blangiardo et al., 2013; Carroll et al., 2015; Rue and Martino, 2009), might be required. Finally, while this paper analyzes objective food environment and identifies neighborhoods with less healthy NRE from a statistical modeling perspective, future research could investigate how residents perceive the restaurant environment in their neighborhoods (Barnes et al., 2015) or how they are truly exposed to the restaurant environment based on activity space (Sadler & Gilliland, 2015).

6. Conclusion

This research illustrates a BSFA approach for assessing the healthfulness of restaurant environment at the neighborhood level, where healthfulness is a latent factor derived from three correlated restaurant assessment indicators: availability, affordability, and facilitator/barrier of healthy eating. Methodologically, uncertainties associated with the descriptive statistic (i.e., mean NEMS-R score) are modeled by accounting for the varying total number of accessible restaurants between neighborhoods, borrowing information of NRE healthfulness in adjacent neighborhoods, and incorporating variations of in-restaurant features within neighborhoods. These uncertainties are quantified with posterior estimates including the range of 95% CrI and the posterior probability of the composite index falling into the lowest quintile.

The applied modeling approach enables to identify neighborhoods with least healthy NRE and the in-restaurant feature that is most relevant to NRE healthfulness. Such information guides community food planning and interventions in terms of where and what restaurant indicators to intervene. In particular, neighborhoods with a composite NRE index in the lowest quintile (i.e., those with the darkest color and locate at west, northwest, north, and northeast Kitchener, Fig.3a) should be targeted for interventions, with prioritization of two clusters of neighborhoods at west and towards northwest Kitchener and several individual neighborhoods across the city (Fig.3b). The identification of facilitator/barrier with highest loading (compared to availability and affordability) on NRE healthfulness supports implementing interventions for increasing/decreasing facilitator/barrier of healthy eating such as mandatory menu labeling. While the applied modeling approach provides a tool for assessing NRE healthfulness of neighborhoods without accessible restaurants within a walkable distance, interventions for these neighborhoods warrant special attentions.
Acknowledgement

We are grateful to Dr. Leia Minaker, University of Waterloo, for providing the consumer nutrition environment dataset of the Region of Waterloo.
References


Dumanovsky T, Huang CY, Bassett MT, Silver LD. Consumer awareness of fast-food


Maddock J. The Relationship Between Obesity and the Prevalence of Fast Food

Mair JS, Pierce MW, Teret SP. The use of zoning to restrict fast food outlets: a potential strategy to combat obesity. 2005.


Nordström J, Thunström L. The impact of price reductions on individuals’ choice of healthy meals away from home. Appetite 2015; 89:103–11.


Ontario Professional Planners Institute, Healthy Communities and Planning for Active


Appendix: WinBUGS code used for the final model (Model I)

model
{
  ## Number of indicators: 3
  for(j in 1:J)
  {
    ## Number of restaurants: 351
    for(k in 1:K)
    {
      for(m in N[k]:N[k+1]-1))
      {
        sub[j,m] <- mu[j.ID[m]]
      }
      mu2[j,k] <- sum(sub[j,N[k]:N[k+1]-1])/TOTAL[k]
      Y[j,k] ~ dnorm(mu2[j,k], tau[j])
    }
  }
  ## Number of neighborhoods
  for(i in 1:I)
  {
    mu[i,j] <- alpha[j]+delta[j]*theta[i] + u[j,i]
    u[j,i] ~ dnorm(0,tau.uid)
  }
  alpha[j] ~ dflat()

  ## Random noise
  tau.uid ~ dgamma(0.5,0.0005)
  sigma.uid ~ sqrt(1/tau.uid)

  ## Indicator precision
  tau[j] ~ dgamma(0.5, 0.0005)
  sigma[j] ~ sqrt(1/tau[j])
}

delta[1] ~ dlnorm(0,0.01)
delta[2] ~ dnorm(0,0.001)
delta[3] ~ dnorm(0.0001)

## Identify the posterior probability that i^{th} neighborhood falls into the lowest quantile
for(j in 1:N)
{
  darank[j] <- rank(theta[j],)
  hotspot[j] <- step(-darank[j]+60)
}

##the value of the 20% threshold
ranked60 <- ranked(theta[], 60)
# variance explained

```r
for(j in 1:M) {
  var.theta[j] <- pow(delta[j],2)*pow(sd(theta[j]),2)
  var.noise[j] <- 1/tau.u[j]
  theta.explain[j] <- var.theta[j]/(var.theta[j]+var.noise[j])
}
theta[1:N] ~ car.normal(adj[], weights[], num[], 1)
for(k in 1:sumNumNeigh) {
  weights[k] <- 1
}
```

Note: \( m_2[j, k] \) in the code is the \( \frac{1}{n_k} \sum_{i=N_k}^{N} \mu_{ij} \) in Level 1 (Table2). This link was implemented by \( m_2[j, k] \) \( \leftarrow \) sum(sub[j, N[k]:N[k+1]])/TOTAL[k], where \( TOTAL[k] \) and \( sub \) refers to \( n_k \) and \( \mu \) in the model specification, respectively.