Energy Policy 136 (2020) 111049

Contents lists available at ScienceDirect

Energy Policy

journal homepage: http://www.elsevier.com/locate/enpol

Urban form and transportation energy consumption

Nikhil Kaza

Department of City & Regional Planning, Environment, Ecology & Energy Program, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599-3140, USA

ARTICLE INFO

Keywords:
Transportation
Urban form
Cities
Landscape metrics

ABSTRACT

Transportation energy is a significant portion of the energy consumption of the US economy. While various policies such as changing the fuel mix and alternative fuels are proposed to make the system more efficient, the efficacy of land use policies such as changing the urban form and densification have been subject to considerable debate. In this paper, I use a rich dataset compiled from different sources to test the effectiveness of urban form on energy consumption in the transportation sector. I proxy the consumption with retail sales from gas stations for most of the conterminous United States at a county level. Using demographic, economic and landscape characteristics, I tease out the effect of different dimensions of urban form on energy consumption. I find that compact and contiguous urban form is modestly associated with lower energy consumption and is more important than demographic concentration in explaining the variance.

1. Introduction

In the United States (US), the transportation sector consumes about 29% of the total energy in 2017, rising from 23.5% in the 1960s (Bureau of Transportation Statistics, 2018) even while the energy efficiency of the economy increased. Much of this energy comes from liquid carbon-based fuels contributing to greenhouse gas emissions and bad air quality. Since the early work of Newman and Kenworthy (1989), planners have posited the role for urban planning and physical design of the cities in promoting more efficient transportation energy use. While the link between transportation behavior and land use patterns is unassailable, the effectiveness of using urban form to promote sustainable transportation patterns and therefore energy consumption is subject to considerable debate (e.g. Boarnet and Crane, 2001; Echenique et al., 2012; Handy, 2005; Stevens, 2017). The debate relies on the measurement of total distance travelled assuming a direct link between amount of travel and energy consumption (see e.g. Glaeser and Kahn, 2010; Lee and Lee, 2014). However, the consumption is also affected by other important factors such as, but not limited to, mode choice (e.g. Dieleman et al., 2002), elasticity of travel (e.g. Ding et al., 2017), fuel efficiency of the fleet (e.g. Lutsey and Sperling, 2005) and alternative fuel availability (e.g. Qiu and Kaza, 2017; Stephan and Sullivan, 2008). Thus, relating transportation energy consumption to urban form directly, rather than relating urban form to travel volume, is important. Different urban form characteristics, such as compactness, contiguity, density, and population distribution may have different impacts through different pathways on transportation energy consumption. In this analysis, we find that different urban form variables have different effects on energy consumption that is conditioned by the level of urbanisation.

Household travel surveys are often used to study the relationship between household characteristics, their location characteristics and travel behavior. For example, Brownstone and Golob (2009) use a subsample of the 2001 National Household Travel Survey to claim that households residing in lower density areas are prone to use higher transportation energy through increases in total travel and reduction in fuel efficiency (e.g. larger automobiles). On the other hand, Liu and Shen (2011) find no direct effect between urban form and energy consumption, but considerable indirect effect through mode choice (e.g. walking) and speed of travel using the same survey for the Baltimore subsample. Similarly, moving a step from energy consumption, using econometric models based on household surveys, Jones and Kammen (2014) find that suburbs contribute to half of the household carbon footprint. They find a non-linear association between population density and carbon footprint, primarily resulting from heterogenous intra city urban density variation. However, Stopher and Greaves (2007) point out that household surveys routinely miss 20–30% of the trips (as much as 60%) undertaken by the households. This suggests that we need to establish the relationships between energy consumption, travel and urban form using alternative datasets.

Furthermore, in these analyses the presumption is that household travel is a significant component of transportation energy consumption. In the US, energy consumption by trip types and purposes (e.g. person vs. freight) are not directly forthcoming. However, light duty vehicles with short wheel bases (passenger cars, vans, SUVs etc.) only accounted

E-mail address: nkaza@unc.edu.

https://doi.org/10.1016/j.enpol.2019.111049
Received 7 March 2019; Received in revised form 3 October 2019; Accepted 11 October 2019
Available online 25 October 2019

for 52% of highway transportation energy in 2016 (Bureau of Transportation Statistics, 2018 Table 4.06). Assuming other vehicle types are largely associated with non-household travel, relying simply on household surveys underestimates the total energy consumption in the system and the impact of urban form. In addition to these, household travel surveys rarely capture work related travel especially in employment sectors that require substantial travel (e.g. fire protection, solid waste, construction) or when they do (e.g. transit) they are often ignored in analyses that focus on travel distances (e.g. Jones and Kammen, 2014). Local freight is increasingly becoming important with the explosion of e-commerce (Rotem-Mindali and Weltevreden, 2013) as well as lengthening of supply chains (Hesse and Rodrigue, 2004). Largely because consumption data is not available at a sub-national scale and because urban form measures are scale dependent, there is a lack of literature that interrogates this relationship comprehensively. An exception is Brown et al. (2009), who model both freight and passenger miles using the Highway Performance Measurement System in 100 Metropolitan Statistical Areas (MSA). They suggest that both population density and concentration are associated with lower greenhouse gas emissions. Another is Liu et al. (2019), who suggest that increasing compactness of cities marginally reduces freight activity and emissions.

In addition, urban form is coarsely measured through population density and other demographic variables (see e.g. Bento et al., 2009; Lopez, 2014). However, there is a substantial literature that relates different dimensions of urban form (such as density, diversity, design, destination accessibility) to travel patterns (Cervero and Kockelman, 1997; Krizek, 2003) and therefore should be properly accounted for. In studies that do account for these different dimensions, the geographic structure of the economy of the place is largely unaccounted for. While it is true that residential sprawl is a significant dimension of urban form in the US, job sprawl and polycentricity are becoming increasingly important (Glaeser et al., 2001; Kneebone, 2009). In this study, we account for these dimensions of urban form.

In addition to these measures of urban form relying on census and administrative data, previous research has shown that satellite based imagery provides a complementary picture of urban form (Bereitschaft and Debbage, 2013; Herold et al., 2002; Burchfield et al., 2006). Because population and jobs are mobile, while urban development is durable and largely irreversible, changes to urban landscapes are largely orthogonal to changes in demographic and economic variables. Satellite based measurements can show that urban areas are fragmentary or contiguous irrespective of population density. Urban form measurements that do not incorporate land cover data miss an important dimension of urban form. These measurements have already shown to have correlations with air quality through non-point source pollution such as arising from transportation (Yuan et al., 2018).

Thus, local transportation energy consumption depends on local demographic and economic structure, urban form variables such as density and design and regional characteristics such as polycentricity and access to opportunity. While these local and regional characteristics are readily measured, the transportation energy consumption is often modelled rather than measured. To circumvent this issue, I take a novel approach of using sales in gas stations as a proxy for transportation energy consumption. Because much of gas station sales primarily from gasoline, diesel and gasohol, the sales are directly proportional to volume of fuel consumed. While there are limitations associated with using this data (detailed in the later sections), I find this the best proxy available at a subnational scale for the US. The contribution of the paper is to give an account of how different urban form variables are associated with transportation energy consumption in an area. In this paper, I construct an extensive database of urban form characteristics for every county in conterminous United States reflecting the different levels (rural to heavily urban) and patterns of urbanisation (e.g. fragmentary to compact). Thus, by relating these characteristics and dimensions of urban development, we can fashion land use policies that might make the transportation system more sustainable.

It is useful to use county as a unit of analysis rather than other geographies for few theoretical and practical reasons. First, much of the economic and gasoline sales data is available at this level of geography. Second, in addition to being relatively stable statistical units, counties (or their equivalents) are political units that have some influence (if not outright control) over land use patterns through annexation policies, service provision and have relationships with other local governments. Third, counties are also nested within states, whose fuel taxes and infrastructure investments may be important for explaining consumption patterns. Fourth, urban form variables employed in this analysis is suitable at this geographic scale because the degree of rurality/urbanity is often defined at this scale (Waldorf, 2006). Other higher level of geographic aggregations, such as MSAs, do not have these properties. It is important to understand the heterogenous effects of different variables on fuel consumption for different urbanisation levels to appropriately tailor policy responses.

The rest of the paper is as follows. I first describe the various data sources and methods used to construct different variables. Substantial effort is dedicated to creation of the variables from large and heterogenous datasets for all of the conterminous US. The statistical models employed are then explained in brief. The geographical pattern of the urban form and transportation energy consumption is used to motivate some of the hypotheses and modeling choices. I then describe the results and discuss the policy implications and results. The article concludes with limitations and future avenues of research.

2. Data & methods

This cross-sectional analysis uses data circa 2011, compiled from a number of sources for the conterminous United States. These raw datasets include landcover data from the US Geological Survey, elevation data from Mapzen, demographic data and from the US Census and some economic data from the Bureau of Economic Analysis. The construction of the variables used in the study are described below. Since much of the raw data is large and is assembled from multiple sources, the data processing is done in a multi-node, multi-core, Linux based cluster computing environment in R.

2.1. Transportation energy consumption

I use the 2012 Economic Census by the US Census Bureau to construct a proxy for transportation energy consumption, by analyzing the sales at gas stations in each county (or equivalent areas) of the United States. The Bureau collects extensive data on businesses every 5 years. I use the data that is reported at a county level for the retail sector (North American Industrial Classification System (NAICS) code 44–45), in particular, the total sales receipts from the gas stations (NAICS 447). Due to confidentiality concerns, data for 344 counties are not reported.

Gas stations in the US are routinely accompanied by convenience stores. Thus, the sales receipts include non-energy expenditures in the county. To test the external validity of using the sales to proxy for energy consumption, I compared the sales in gas stations that do not have convenience stores (NAICS 44719). Data is reported only for 835 counties for these type of gas stations, due to confidentiality concerns. For this subset of counties, the correlation between sales from all gas stations and those from gas stations without convenience stores is 0.82, suggesting that the receipts are not heavily biased by the non-energy expenditures.

2.2. Urban form measures

The land cover data, circa 2011, was produced by the US Geological Survey and retrieved from the Multi-Resolution Land Characteristics Consortium website. Each of the 30m pixels (~9 billion for the contiguous US) was classified into 16 National Land Cover Data (NLCD) level II land cover classes (United States Geological Survey, 2014). I use urban land cover classes (21–24) to determine the urban land cover in each county. County boundaries are from US Census for 2010. Calculation of urban form measures such as total urban land cover and its fragmentation is marred by the classification of roads as urban land cover in the raw dataset. Expanding on the procedures described in McCarty & Kaza (2015), I reclassified the urban pixels that are within the buffer (based on the number lanes) of each link in the road network, as non-urban. The road network used is 2011 National Highway Planning Network published as part of the National Transportation Atlas Database (Bureau of Transportation Statistics, 2015). Highway lane miles in the county are also calculated from this dataset. Differences in the positional accuracy of the vector and raster road networks (e.g. curves in the roads) lead to small erroneous residual urban patches in rural areas. To mitigate against these, I use a morphological operator called 'opening' (see e.g. Gonzalez and Woods, 2017). Opening is combination of erosion followed by a dilation using a structuring element. I use a 3×3 square as the structuring element. To remove any other noise, I also reclassify the collections of pixels whose core area is less than 10% of the total patch area. Core area is defined as collection of pixels within a contiguous patch that are completely surrounded by the urban pixels on all sides, whereas the edges are those pixels that abut any non-urban pixels. Elongated patches that are remnants of linear features such as roads have very little core area. These procedures produce a map of urban area that is conducive to measure urban form (see e.g. Fig. 1).

Once the urban patches (contiguous clusters of pixels) are identified, the urban form indices are calculated using SDMTools (VanDerWal et al., 2014). The key urban form metrics are number of patches, mean and standard deviation of patch area (e.g. Kaza, 2013; Seto and Fragkias, 2005) and index of moment of inertia (W. Li, Goodchild, and Church, 2013). High number of patches reflects an overall fragmentary urban pattern, whereas high standard deviation of the patch area is indicative of a fragmentary pattern at the edges (satellite non-contiguous urbanization).

Index of Moment of Inertia (IMI) is a measure of shape compactness; an inverse of the ratio of the moments of inertia (MI) of the urban area relative to the centroid to that of the most compact urban form of the same area (circle). If A is the area of the urban area and then the $A^2r^2/2\pi$ is the MI of the circle. The MI of the urban area is given by

$$\sum_i \left(\frac{r^2}{6} + d_i^2\right) \mathbb{I}_i r^2 \tag{1}$$

where $\mathbb{I}_i$ is the indicator function which is equal to 1 when the cell i is urban and 0 otherwise, r is the resolution of the raster and d_i is the distance of the cell i from the centroid. This can be derived from parallel axis theorem and MI of a square. This measure does not require contiguity of urban form though large urban patches farther from the centroid are more heavily penalized compared to those that are closer. IMI is closer to 0, when the urban form is heavily dispersed and closer to 1 when it is most compact for a given area. This is more effective measure of compactness of urban form than other metrics such as perimeter-area ratio or Reock index (ratio of urban area to the area of minimum bounding circle) because urban form is fragmentary, non-contiguous and polycentric (Kaza, 2019). This metric has been used to study, among other things, the compactness of political districts (Fan et al., 2015) and the effect of land configuration on urban heat islands (X. Li et al., 2016).

The ruggedness of a county is measured by the root squared deviation from the median of the elevation. This is a better metric than root mean squared deviation, because mountainous areas have a right skewed distribution and means are susceptible to outliers. The elevation data is from Mapzen, which is in turn derived from 3DEP data courtesy of the U.S. Geological Survey (Sugarbaker et al., 2017). The nominal ground resolution of the elevation raster is 150m representing 360 million pixels for the conterminous US.

Of the 220,000 block groups, I identify the outliers in employment density using the McMillen (2003) method of subcenter delineation. The outliers are substantially different from the predictions of a locally weighted regression that accounts for the position of the centroid of each block group. The block group level employment data is from Longitudinal Employer-Household Dynamics (LEHD) data by the US Census. These outliers are the employment centers. The proportion of the employment in these centers relative to the total employment is considered the index of concentration of employment. Values closer to 1, suggest a more concentrated employment opportunity within the county and closer to 0 suggest job sprawl.

Similar methods cannot be used for quantify the concentration of population as the population density has a much flatter spatial profile than employment density, which has marked by its peaks and plains. Thus, to account for population concentration, I use Gini index of population density using block groups within the county. I use the 2010 decennial census because population statistics are available at block group level. Area of the block group is taken to be the land area. Values closer to 1 implies the county has large areas that are unpopulated and fewer areas that have significant population density (an uneven distribution).

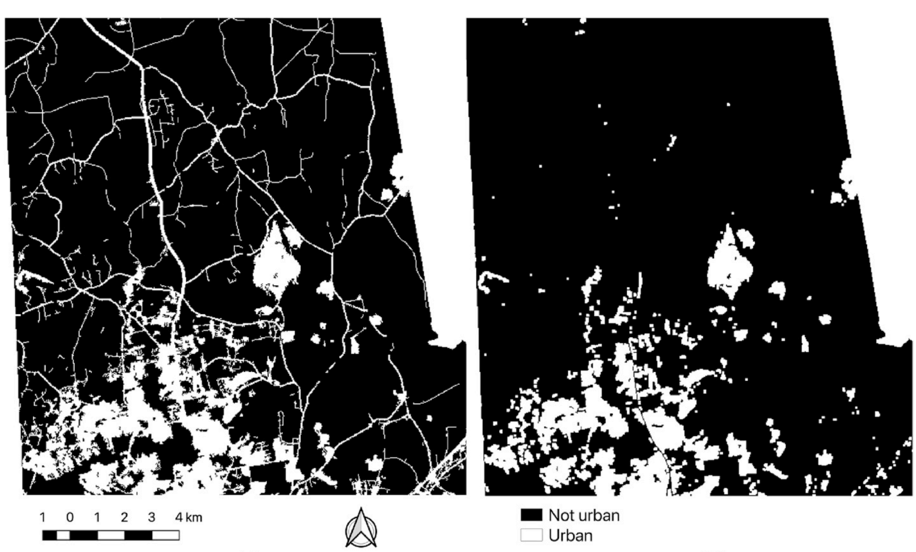

Fig. 1. Example of processing land cover data to derive urban form metrics a) Raw NLCD data b) Processed data.

2.3. Other control variables

Total and sectoral employment for the county is from the Local Area Income tables published by the Bureau of Economic Analysis for 2011. I define freight related employment as total employment in each county that is in goods producing sectors (Agriculture, Mining, Manufacturing and Construction) as well as other freight dependent sectors (Wholesale Trade, Retail trade and Transportation & Warehousing). I use the proportion of the employment in these sectors as one of the variables in the models. Total population and median income are from the American Community Survey 5-year tables.

From the smart location database published by the US Environmental Protection Agency (Ramsey and Bell, 2014), I also calculate the county level indicators for intersection density of pedestrian- oriented intersections having four or more legs. These, in turn, are derived from the Navteq street network databases.

According to the Energy Information Administration, 70 percent of the retail gasoline price is due to price of crude oil and refining costs. The regional variation in prices is because of marketing, distribution, taxes and regulation. For example, California's requirement for unique blend partially explains the higher prices in that state compared to its neighbors. Because average annual prices at a county level are not readily available, I proxy the variation by accounting for number of gas stations (US Census), and distance from nearest petroleum terminal (Department of Homeland Security). State level taxes on gasoline, representing the variation in prices across states is from the 2011 state level statistical abstracts published by Federal Highway Administration (2014).

Distance to nearest Central Business District (CBD) is calculated using Great Circle Distance using the Haversine method (Sinnott, 1984) from the population weighted centroid of the county. The population weighted center of the county is obtained from US Census. The locations of the Central Business Districts are from Fee and Hartley (2012).

2.4. Statistical methods

Because the distribution of the variables are highly skewed and are of different scales, I used Inverse Hyperbolic Sine (IHS) transformation for variables that are not constrained in a limited interval (Burbidge et al., 1988). Unlike log transformation, IHS is defined on 0 and the interpretation of the coefficient is the same as interpreting a coefficient of a log transformed variable as (semi) elasticities. In particular, the log-likelihood function is defined for this transformation with 0 values, which is important for accounting for spatial autocorrelation.

I use regression analysis through ordinary least squares to first test the association between transportation energy consumption and urban form. In addition to the various variables described in the above subsections, I also include a dummy variable to represent each state as fuel taxes and other state level energy policies affect fuel prices in the state. The presence of spatial autocorrelation necessitates also the use of spatial error model (selected using the Lagrange Multiplier test). This is of the form.

$$Y = \alpha + \delta_i + \beta X + \xi; \xi = \lambda W \xi + \varepsilon; \varepsilon \tilde{} N(0, \sigma^2 I) \quad (2)$$

I use queen contiguity to represent the spatial weights matrix W. ξ is the spatial component of the error term. α, δ_i (state fixed effects), β and λ are estimated using Maximum Likelihood Estimation using 'spdep' (Bivand and Piras, 2015). I present the results from both models.

While above tests produce information about the 'theoretical importance', i.e. importance of the changes in the dependent variable by changes in the independent variables, we also need to explain the variance in the dependent variable (Achen, 1982). Johnson and LeBreton (2004 p. 238) define relative/dispersion importance as, "the proportionate contribution each predictor makes to R^2, considering both the unique contribution of each predictor by itself and its incremental contribution when combined with the other predictors." To estimate the dispersion importance, I use a metric proposed by Lindeman et al. (1980) by averaging sequential sums of squares over all orderings of explanatory variables. I use 'relaimpo' for this part of the analysis (Grömping, 2006).

3. Patterns of transportation energy consumption and urban form

Of all the states, Wyoming and North Dakota are among the top of the per capita expenditure in gasoline stations, followed by the states in the Midwest. These states are characterized by low population and vast open spaces. The populous regions in the US, the Northeast and the Pacific are at the bottom of the per capita expenditures. District of Columbia has only $418 expenditures per capita suggesting potential explanations of large commuting population from nearby states (Virginia and Maryland in particular), extensive public transportation infrastructure and high population density.

Counties outside metropolitan statistical areas have 41.2% more per capita sales than those within them. While these counties account for only 16.2% of the total population, this suggests that urbanization is associated with lower per capita consumption due to proximity of destinations and increased economic development. Finer urban type classification of the counties from National Center for Health Statistics (NCHS) reveal an even starker pattern. Large central metro counties, on average, have two fifths of the per capita sales of the non-core rural counties. As the urbanization intensifies, per capita consumption decreases (see Table 1), even though large and medium metro counties account for more than two-thirds of the total sales.

Of the metropolitan areas Cheyenne, WY, Winchester, VA and Joplin, MO spend more than 4,000 USD per capita in gas stations, suggesting a large variation within urban areas. Corvallis, OR has the least, with both New York, NY and Boulder CO following closely (<1,000 USD per capita). This suggests dense urban environments with high transit amenities may result in lower spending. Metros along the Gulf coast and in states like Arizona and South Carolina exhibit large expenditures.

The county level data reveal that much of the Mountain west have high expenditures per capita (see Fig. 2). However, it is noticeable that the top five in per capita sales counties are in Texas, Colorado and New Mexico all have population less than 10,000. In particular, Culberson county in Texas has a population of 2,400 and reported over 125 million USD in sales with only 8 gas stations. Since it is not clear if data collection and reporting errors explain this, I ignore the counties in top 1 percentile (per capita expenditures more than ~10,000 USD per annum) in the statistical analyses.

Cities are more fragmented in the South and the West as is evidenced from the number of patches (see Fig. 3). In particular, counties in Florida and California have both large number patches and large average patch size. This is indicative of large urban subdivisions in the green fields, rather than coalescent infill urbanization.

The IMI, which is a measure of compactness suggests a different pattern. Compared to other parts of the country, the major metropolitan areas are more compact. This is largely due to the fact IMI measures both

Table 1
Transportation energy expenditures by type of county (Source: US Census & NCHS).

County Type	Population (in 1000s)	Annual Sales (in Millions USD)	Per capita Sales (in USD)
Non-core	14.4	47.3	3,275
Micropolitan	42.0	110.1	2,622
Small Metro	79.2	190.1	2,400
Medium Metro	169.0	333.8	1,975
Large Fringe Metro	204.9	363.7	1,775
Large Central Metro	1,364.7	1,753.2	1,285

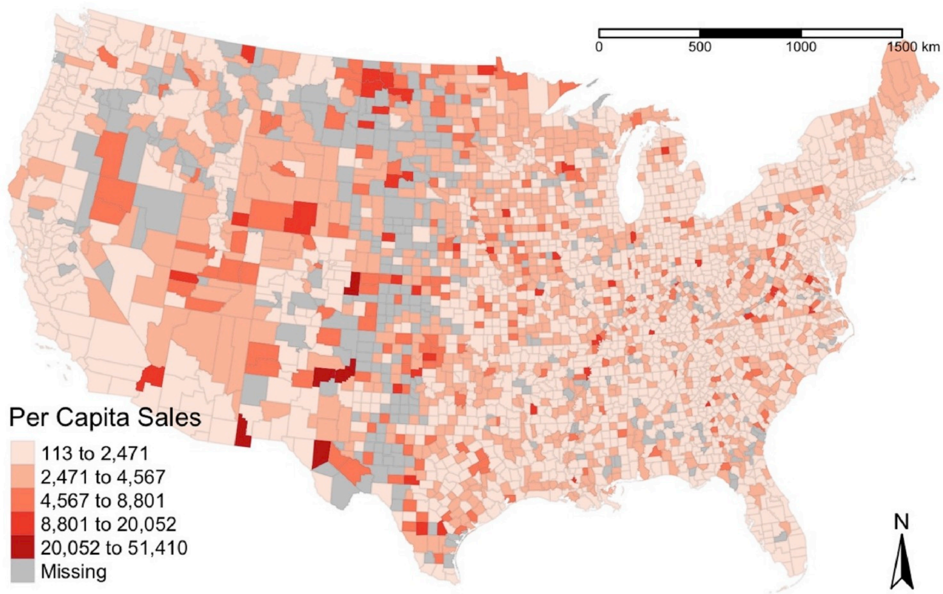

Fig. 2. Gas station sales in 2012 at the county level (Source: US Census).

distance to the centroid and the size of the urban patch. Large urban areas close to center, which are more likely to be the case in cities, receive a higher compactness score, while suburbs and rural counties receive lower. It should also be noted that large central metros have higher IMI and skewed to the right, while the rest of the counties have left skewed distributions (see Fig. 4). This is probably because these large central counties have relatively fewer patches compared to other types of counties, including fringe and medium metros. Of all the large central metro counties, on average, West North Central are most compact, while Mountain counties are least compact. This result can partially be explained by the variation in topography. Terrain ruggedness, potentially prevents compact development. In the large fringe metro category, Middle Atlantic counties, on average, are most compact. Predictably Western US with its large mountainous ranges have more elevation differences. Other parts of the US with higher ruggedness are also in the Appalachia, the Adirondacks and the Ozarks. These are also places with comparatively low economic development. However, the converse is not true. Many parts of the eastern Carolinas, Northern California and Midwest have low ruggedness but also relatively low economic development. While the terrain might explain some constraints on economic activity, it is not determinative.

The concentration of employment and population is very different in different types of counties (see Fig. 3). Central metro counties are more evenly populated (lower concentration) but have higher employment concentrations. This is also evident in the spatial patterns (see Fig. 2). In fact, population in rural counties is more heavily concentrated and employment is more heavily dispersed. The differences in the distributional patterns of urban form metrics and demographic variables reveal that they are complementary to one another.

Unsurprisingly, metropolitan counties have higher employment compared to other types of counties. Due to the documented trend of relocation of manufacturing first to the suburbs and then outside the country, large metropolitan counties are more heavily service oriented. Non-metropolitan and micropolitan counties still have substantially high proportion ($\sim$ 40%) of employment in freight-oriented employment (goods producing, trade and transportation sectors). These counties also have substantially fewer gas stations and are likely to be further away from the petroleum distribution terminals than metropolitan counties. Intermodal freight terminals are more common in large metro counties and counties with ports, with Cook County (Chicago) having the highest number (109). Unexpectedly, medium size metro counties, on average, have more such terminals than large fringe metro counties.

Transportation infrastructure such as highway lane miles are predictably more concentrated in large metro areas. However, the pedestrian orientation of the transport infrastructure is not. The bivariate relationship between employment/population size of a county and the density of pedestrian oriented intersections are positive for all counties except large central metro counties. This is not surprising because pedestrian oriented intersections are relatively sparse in large counties such as Riverside and Maricopa that dominate the set of large central metro counties.

4. Results

Summary statistics for the different variables included in the models are in Table 2. The models provide evidence for the stated hypotheses in the earlier sections (see Table 3). Both employment and population have substantial effect on the total sales in a county. But more significantly increase in share of employment that is related to freight (goods producing, trade, transportation and warehousing) have a large effect. The proportion of freight related employment is decreases as county moves from urban to rural suggesting the importance of service-oriented employment to urban counties. This could potentially explain the larger per capita consumption in rural counties. Nonetheless, the number of intermodal freight terminals have no effect. This suggests that the intra-county freight is largely responsible for the variation of sales within the county and the effect of inter-county freight is either captured by other demand variables such as population or supply variables such as highway provision. Incidentally, highway lane miles variable has the largest elasticity among the variables studied.

The distance to the petroleum distribution terminals and the number of gas stations, serves as a proxy for retail price variation and function as control variables. Larger distances, a proxy for higher prices, reduce sales. The effect, however is small. Higher number of gas stations, imply increased competition and lower prices; this does not seem to have any effect on the sales. Non-core and micropolitan counties are generally much further away from the terminals and this could explain both fewer stations, higher prices and higher per capita expenditures.

While, as control variables, both number of jobs and total population have large and robust effect on energy consumption, the demographic and economic concentrations have little effect. This suggests that

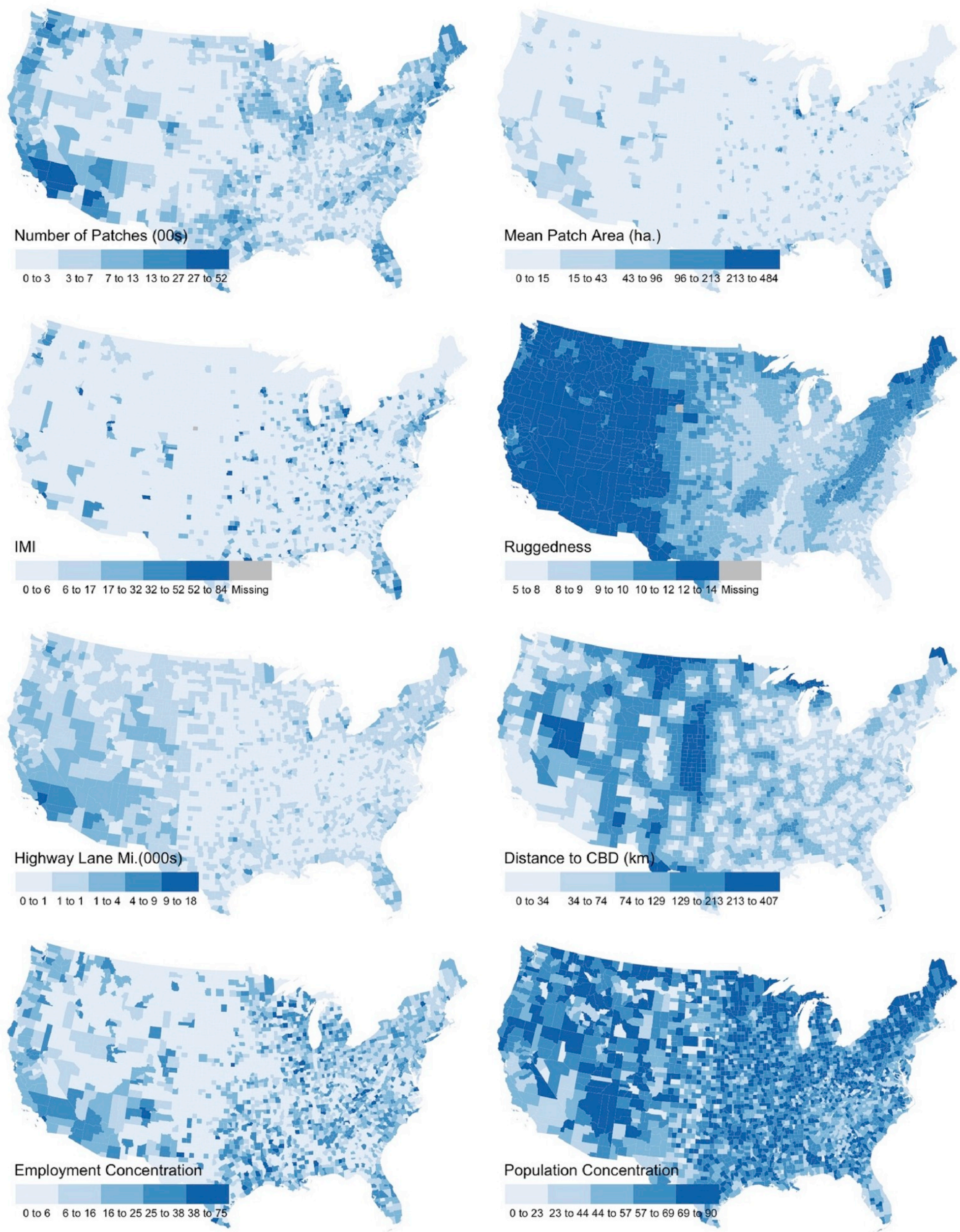

Fig. 3. Spatial patterns of some key explanatory variables circa 2011. Note that some variables are linearly scaled for map legibility. Ruggedness is IHS transformed. (Sources: Various. Refer to the text).

centrality of employment, even polycentricity, has little effect on the gasoline consumption. Proportion of concentrated employment, on average, is highest in large central metros (close to 30%). In other types of counties, the proportion, on average, is closer to 20%; employment is widely dispersed in the United States. The lack of effect of the concentrated employment could potentially be due to these low levels.

Controlling for the above variables, the urban form and spatial structure variables have expected impact. Fragmented urban counties characterized by high number of patches have higher expenditures. However, fragmentation at the edges as a characterized by increased standard deviation, does not seem to have any independent effect on the fuel consumption. Counties with large urban patches (on average) are associated with higher energy consumption profiles. Counties with large urban patches are usually associated with more intense and more expansive urbanization patterns. The more compact the urban area is (evidenced by IMI), the lower the energy consumption. Controlling for the total urban area, IMI is an index of spread around the centroid. Counties that have only few large urban areas which are concentrated near the center are associated with lower expenditures. Increase in pedestrian oriented intersection density has a negative effect on energy consumption. Furthermore, counties whose population weighted centroids that are further from the CBD have higher expenditures. This suggests CBDs still hold a significant role despite recent rise in polycentric development patterns. This significance of this variable also points to effects of the spatial structure of the region.

While considering the effect of these urban form variables on a national sample is interesting, the generalizability may be limited because of heterogenous effects within counties at different stages of urbanisation. Collapsing the six categories of counties into three, different urban form variables have different impacts in these subsamples (see Table 4). While pedestrian intersection density seems to have an effect in large metro counties, it has no effect in other types of counties. Perhaps this is

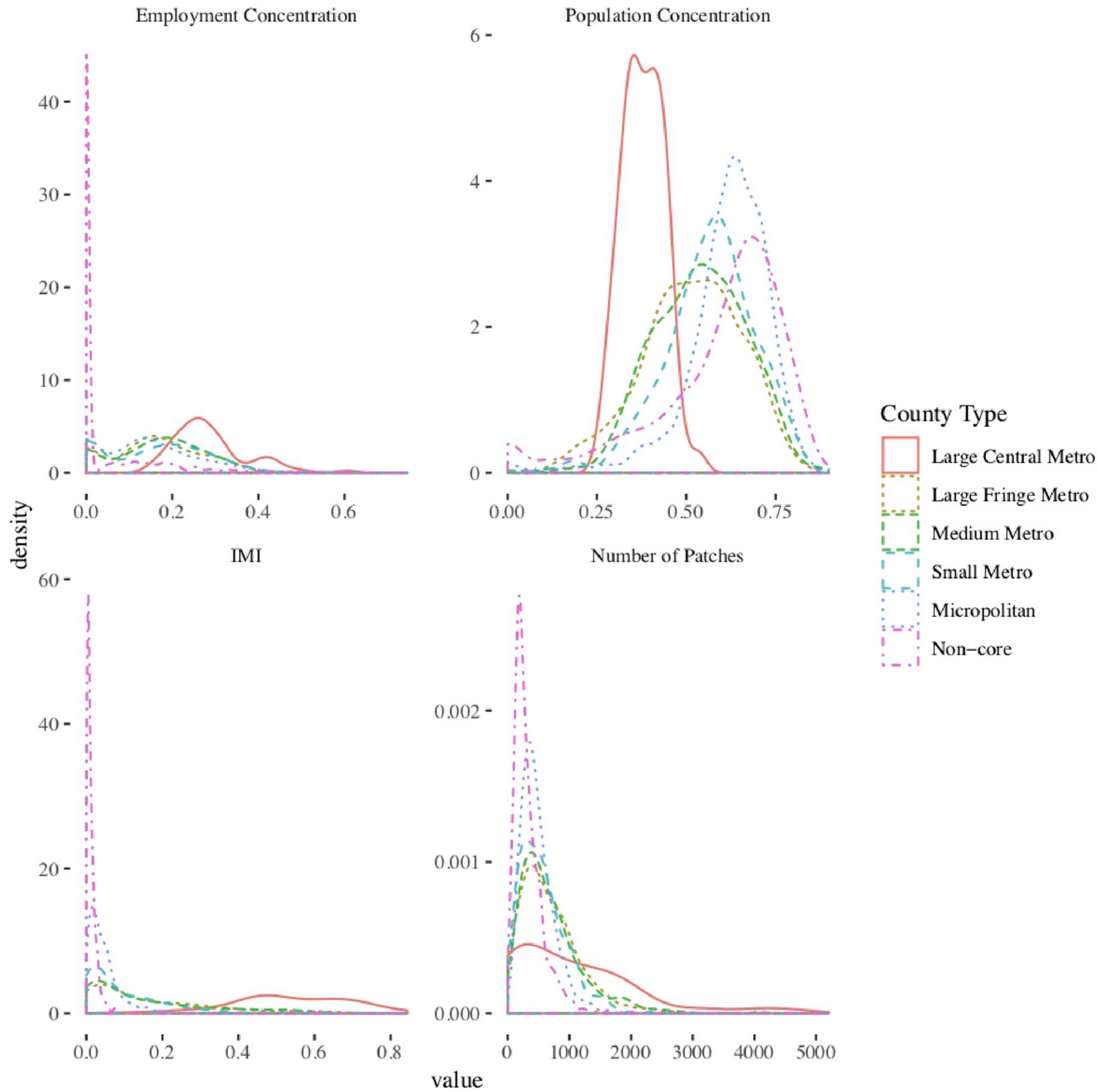

Fig. 4. Distribution of some key explanatory variables. County type is from NCHS.

Table 2
Summary Statistics (before IHS transformation).

Variable	n	Mean	St. Dev.	Min	Max
Sales in Gas stations (000s)	2,729	196,316.70	452,414.60	1,304	12,158,749
Population	2,729	109,848.40	331,201.00	731	9,787,747
Median Household Income (USD)	2,729	45,397.53	11,907.45	19,936	120,332
Employment	2,681	63,464.51	206,397.80	625	5,479,684
Proportion of Freight-Oriented Employment	2,681	0.381	0.09	0.05	0.74
Employment Concentration	2,729	0.122	0.123	0	1
Population Concentration	2,729	0.588	0.141	0	0.899
Distance to Nearest Petroleum Distribution Terminal (km)	2,729	46.73	35.20	0.23	241.23
Number of Gas Stations	2,729	40.73	83.50	3	1,887
Distance to CBD (km)	2,729	63.961	51.623	0.32	356
Highway Lane Miles	2,729	708.23	678.92	12.68	18,323.89
Number of Intermodal Freight Facilities	2,729	1.16	4.02	0	109
Density of Pedestrian Oriented Intersections	2,727	1.27	3.51	0.01	64.47
Terrain Ruggedness	2,728	20,490	47,994	88	686,335
Number of Urban Patches	2,729	526.56	429.50	8	5,206
Mean Urban Patch Area (sq.m.)	2,729	136,676.00	246,636.30	9,454	4,844,206
Std. Dev. of Patch Area (sq.m.)	2,729	858,364.20	1,365,706.00	18,703.93	28,487,123
IMI	2,729	0.08	0.14	0.00	0.84

Table 3
Results of the statistical models.[a].

	Dependent Variable: Annual Sales in Gas Stations in a County	
	OLS	Spatial Error Model
Population	0.298*** (0.041)	0.292*** (0.041)
Median Income	0.055 (0.048)	0.037 (0.050)
Employment	0.416*** (0.041)	0.421*** (0.041)
Proportion of Freight-Oriented Employment◆	0.256*** (0.103)	0.276*** (0.103)
Employment Concentration◆	-0.141 (0.089)	-0.134 (0.088)
Population Concentration◆	0.069 (0.072)	0.064 (0.071)
Number of Intermodal Freight Facilities	-0.001 (0.015)	-0.002 (0.015)
Distance to CBD	0.020* (0.011)	0.021** (0.011)
Highway Lane Miles	0.242*** (0.025)	0.240*** (0.025)
Density of Pedestrian Oriented Intersections	-0.147*** (0.031)	-0.148*** (0.031)
Terrain Ruggedness	-0.058*** (0.012)	-0.061*** (0.013)
Distance to Nearest Petroleum Distribution Terminal	-0.021* (0.011)	-0.020* (0.011)
Number of Gas Stations	0.0002 (0.0001)	0.0001 (0.0001)
Number of Urban Patches	0.191*** (0.025)	0.194*** (0.025)
Mean Urban Patch Area	0.197*** (0.044)	0.206*** (0.044)
Std. Dev. of Patch Area	0.046* (0.027)	0.045* (0.027)
IMI◆	-0.475*** (0.150)	-0.463*** (0.148)
Constant	-1.819*** (0.603)	-1.682*** (0.610)
λ		0.100*** (0.031)
Observations	2,679	2,679
(Adjusted/pseudo) R^2	0.897	0.900

Note: *p < 0.1 **p < 0.05 ***p < 0.01.
Note: a) All variables except marked with ◆ are IHS transformed. State dummies are included in both the models but are not shown.

Table 4
Select results for subsamples.[b]

	Dependent Variable: Annual Sales in Gas Stations		
	Large Metro Counties	Small/Medium Metro Counties●	Non-Core/ Micropolitan Counties
Density of Pedestrian Oriented Intersections	-0.140** (0.063)	0.032 (0.055)	-0.093 (0.061)
Number of Urban Patches	0.282*** (0.068)	0.122** (0.051)	0.100*** (0.035)
Mean Urban Patch Area	0.372*** (0.113)	0.107 (0.082)	0.056 (0.058)
Std.Dev Urban Patch Area	-0.089 (0.071)	0.131*** (0.049)	0.043 (0.036)
IMI◆	-0.146 (0.267)	-0.797*** (0.218)	0.201 (0.531)
Observations	404	653	1,622

Note: *p < 0.1 **p < 0.05 ***p < 0.01.
[b] All variables except marked with ◆ are IHS transformed. Only select variables are shown; for other variables in the models please refer to Table 3. Models marked with ● are spatial error models, due the presence of spatial autocorrelation. Standard errors are in the parentheses.

due to the fact that metro counties have an order of magnitude higher density of pedestrian intersections; while large central metropolitan counties have almost 9 intersections per sq. mile (median), other types of counties have less than 1. Similarly, the higher compactness is associated with lower energy expenditures in small and medium metropolitan counties, while having no effect in other types of counties. This suggests that promoting compact development will have much more significant effect in second tier urban areas than the big metropolitan areas. The positive correlation of the variance in patch sizes (indicative of fragmentation at the urban edge) in small and medium metro counties also provide support to this claim. Nonetheless, fragmentation as measured by total number of urban patches is significant across different types of counties.

While the statistical significance is of importance, the importance of a variable in explaining the variance of the dependent variable an often overlooked. While population and employment are significant in explaining the variance, the urban form indicators based on landscape metrics are close (see Fig. 5). It should be noted that variables that are not statistically significant in Table 3 can still be important (see Feldman, 2005). For example, while the standard deviation of the patch area is only marginally significant in the statistical models, it is the third most important variable in explaining the variance of fuel consumption. Such is usually the case when variables are strongly correlated with other variables in the model. At the same time, while ruggedness is highly significant and negatively correlated with fuel consumption, it explains the variance the least. However, as Grömping (2006) points out, while statistical significance is important for predictive purposes, one should still evaluate the effect of the insignificant theorized variable for potential causal purposes. Moreover, the variance inflation factors (VIF) calculated with 'car' (Fox and Weisberg, 2011) are less than 5 for all the variables except the controls (population and employment). Because VIFs are below the accepted thresholds for the variables of interest, the conclusions about them are not sensitive to multicollinearity problems (see Allison, 2012).

5. Limitations & further work

Just as any other empirical work, we ought to consider the conclusions from this research in the context of its limitations. This study makes no claims about causal relationships, but provides evidence of associations. It supplements existing literature on transport energy and broadly confirms the relationship of urban form to transportation energy consumption.

The construct validity of using sales in gas stations in a county as an indicator of transportation energy consumption needs to be more closely examined. Gas stations sell numerous energy products including diesel and gasoline at different prices. This has implications for total energy consumption and ultimately the greenhouse gas emissions associated with the transportation sector.

Because transportation, by definition, is highly mobile it is not clear what percentage of the sales in each county can be associated with travel within the county and what is a function of regional travel and transportation throughput. Because highly disaggregated data on sales and locations of gas stations are not readily available, I cannot tease out these effects. Future work involving digital traces of household expenditures or logistics tracking could perhaps be used to elucidate these effects. Larger geographic agglomerations such as Metropolitan Statistical Areas are potentially a unit of analysis. However, they suffer from the same boundary problem as counties. There are also other objections. MSAs frequently span state boundaries, which creates problems for including key state level variables such as fuel taxes. Focusing the

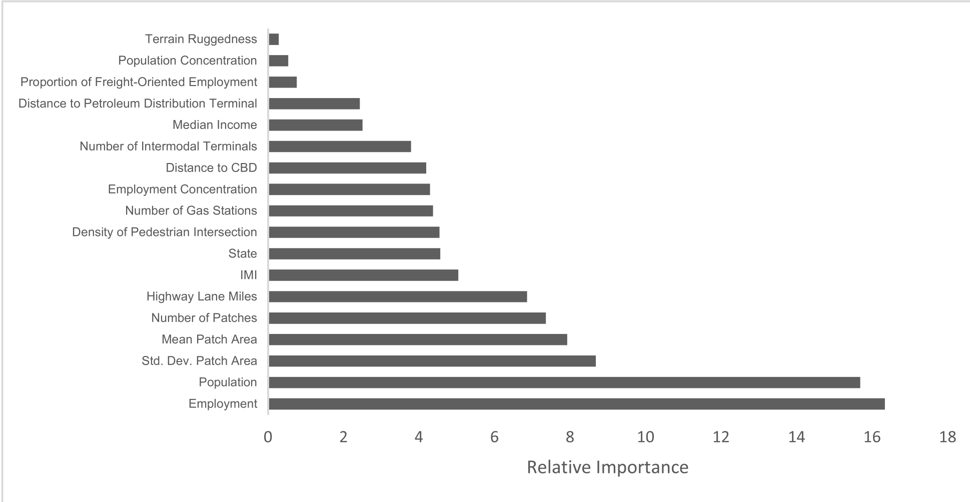

Fig. 5. Relative importance of explanatory variables.

analysis on metropolitan areas ignore some of the high per-capita consumption counties that are non-core and micropolitan counties, which are 62% of the sample. Furthermore, there is significant heterogeneity of urbanisation within MSAs, i.e. large numbers of rural counties are included within metropolitan areas as the emphasis is primarily on the connection between centers and periphery rather than distinctions between urban and rural. Isserman (2005) argues that majority of the rural residents (~30 million) are within metropolitan areas. This skews the urban form metrics, we are interested in, rendering them less explanatory. Nonetheless, secondary analysis conducted at the metropolitan level suggested that fragmentation of urban form increases energy consumption (results in Appendix A). However, MSAs are not the only geographic aggregation that could potentially be relevant (see e.g. Labor Market Areas (Fowler et al., 2018) or commuting clusters (He et al., 2019)). Future work should thoroughly examine the role of geographic scale and unit of analysis on both urban form indictors and the relationship of urban form to energy consumption.

A key variable that is missing in the analysis is the price of energy. Local level fuel prices are not readily available from Energy Information Administration or the Federal Highway Administration. One private source of information, Oil Price Information Service by IHS analytics, was prohibitively expensive to use in this analysis. Temporal variation on fuel prices is largely dependent on variation in crude oil prices. However, spatial variation in price is dependent on local and state level factors such as distribution costs and taxes. I proxy for the differences in the prices using a coarse variable of dummy variable for each state as well as distance to petroleum distribution centers. As a robustness check, I also acquired average gasoline taxes from Federal Highway Administration (2014) and ran the models by substituting the fixed effects of the state with the level of fuel taxes. The coefficient is not significant (results not shown). Other determinants of heterogeneity of local prices are proxied using distance to nearest petroleum distribution terminals and number of gas stations. These variables could be endogenous as the terminals are located close to demand centers. Future work should address the role of price fluctuation and geographic variation in energy demand.

Ultimately, transportation mode choices, distances travelled, trip frequencies and mode efficiencies all contribute to the energy consumption. While some of it observed at a household level in the household travel surveys, it is hard to relate individual household travel behavior to aggregate urban form characteristics as individuals are exposed to multiple urban forms and also might suffer from endogenous location choice. Estimates for non-household travel may also suffer from fleet composition, organizational policies and practices. Thus, we have to use aggregate units such as counties as unit of analysis. However, this aggregation brings its own problems of not being able to tease out the effect of various other policies avenues (such as subsidies for efficient vehicles) that may be more effective than urban form related policies in promoting sustainable transportation systems and behavior. Future work should combine both the aggregate geographic and micro level data to come up with better understanding of land use, mode, behavior and energy connections.

Wickham et al. (2017) report that the thematic accuracy of the 2011 NLCD is at 83% for Level II categories, and 89% for Level I categories. Since we are interested only in differentiation of urban from other landcover rather than differences within urban, the latter is relevant. However, as I have pointed out in section 2.2, raw NLCD does not fully capture the urban form characteristics. The positional accuracy of the road networks and the completeness of them introduce errors into the analysis. However, because there is no reason to suspect that errors are systematic, the estimates reported in the results are unbiased. Nonetheless, further research should examine the implications of using different geographical datasets with different accuracies on urban form.

Many other independent variables used in this work, as usual, have measurement errors that are both documented and undocumented. The American Community Survey data come with standard errors, but in this analysis, I only use the point estimates. Because the geographical unit is large (county), the errors are relatively minor, but should be noted. LEHD data excludes self-employed, federal/military/railroad workers and other employment exempt from unemployment insurance laws. It is estimated that LEHD underreports employment by 15% (Cambridge Systematics Inc, 2017). However, little is known about the systematic geographic errors. As long as there is no reason to suspect a systematic bias, we can provisionally accept the conclusions.

This study is focused on the association between energy consumption and urban form metrics and not on causality. Many of the urban form metrics are potentially endogenous to transportation infrastructure, including number of gas stations and gasoline sales. In addition, there may be other omitted variables that could influence the conclusions. Future work, could take advantage of the stability of the statistical units to study the changes in these variables to make some causal claims.

6. Conclusion and policy implications

The impact of urban form on transportation energy consumption has long been recognized. Yet, as this analysis shows, urban form characteristics that have hitherto not been considered play an important role in explaining some of the patterns. No single variable can capture the range of patterns, but combinations of them can paint a better picture.

Contrary to expectations, hilly counties are associated to have lower energy consumption. While terrain might only allow for fragmentary urbanization, it is also likely that the urban patterns within those fragments are likely to be more compact necessitating shorter travel and therefore less consumption. It is also likely that economic indicators such as jobs and median income are not capturing the full extent of economic distress in the county which could also explain the lower energy consumption. Ruggedness is largely associated with low economic development in the US due to difficulty of creating social, economic and physical infrastructures.

Nonetheless, most of the indicators of sprawling urban form have been shown to be associated with higher consumption. The effect of urban form variables, while significant, is not large compared to the levels of demographic and economic variables. However, they are more important than the concentrations of demographic and economic variables. This suggests that changing the patterns of development can be beneficial. While economic and demographic variables are important in aggregate, this research suggests that policies that modifying the spatial patterns of development might have definite impact. In particular, regulations such as urban growth boundaries and programs to promote infill development might reduce energy consumption patterns modestly.

The availability of highway infrastructure is also associated with increase in consumption and is an important indicator. This provides some evidence for induced travel. While the infrastructure availability is correlated with population and employment, its independent effect suggests that we should pay close attention to the decisions about road infrastructure. Lane miles are correlated to fragmentary patterns, especially in micropolitan and non-core counties, exacerbating the effect of urban form. Coupled with the fact, that large proportions of commuters use private automobiles, promoting alternative and less energy intensive and more healthy transportation modes such as biking and walking by providing more infrastructure for them would be useful. This conclusion is substantiated by the importance of pedestrian oriented intersections in the models.

Transportation contributes to about 28.5% of the greenhouse gas emissions in the US roughly equivalent to the emissions from the electricity sector. Unless dramatic shifts happen in the fuel mix, such as shift to electric vehicles, and the fuel efficiency of the fleet, these emissions are expected to grow. The prospect of autonomous vehicles may change the disincentives for single occupancy vehicles and incentives for public transportation and sustainable modes such as walking and biking with tremendous implications for energy consumption. This study provides evidence that some dimensions of urban form are effective in reducing energy consumption, even if the elasticities are small. Land use policies may provide some pathways to make the system more efficient. They should be treated as complementary to other transportation policies.

Declaration of competing interest

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

Acknowledgements

Much of this work has been done at the Habitat Research Center at the École polytechnique fédérale de Lausanne. I wish to thank two anonymous reviewers and Daniel Rodriguez who provided extensive comments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.enpol.2019.111049.

References

Achen, Christopher, 1982. Interpreting and Using Regression. SAGE Publications, Inc, Thousand Oaks, CA.

Allison, Paul, 2012. When can you safely ignore multicollinearity? Stat. Horiz. 2012 https://statisticalhorizons.com/multicollinearity.

Bento, A., Lowe, S., J Knaap, G., Chakraborty, A., 2009. Housing Market effects of inclusionary zoning. Cityscape: J. Policy. Dev. Res. 11 (2), 7–26.

Bereitschaft, Bradley, Debbage, Keith, 2013. Urban form, air pollution, and CO2 emissions in large U.S. Metropolitan areas. Prof. Geogr. 65 (4), 612–635. https://doi.org/10. 1080/00330124.2013.799991.

Bivand, Roger, Piras, Gianfranco, 2015. Comparing implementations of estimation methods for spatial econometrics. J. Stat. Softw. 63 (18), 1–36.

Boarnet, M., Crane, R., 2001. The influence of land use on travel behavior: specification and estimation strategies. Transp. Res. Part A 35 (9), 823–845.

Brown, Marilyn A., Frank, Southworth, Andrea, Sarzynski, 2009. The geography of metropolitan carbon footprints. Policy Soc. 27 (4), 285–304. https://doi.org/10.1016/j.polsoc.2009.01.001.

Brownstone, David, Golob, Thomas F., 2009. The impact of residential density on vehicle usage and energy consumption. J. Urban Econ. 65 (1), 91–98. https://doi.org/10.1016/j.jue.2008.09.002.

Burbidge, John B., Magee, Lonnie, Leslie Robb, A., 1988. Alternative transformations to handle extreme values of the dependent variable. J. Am. Stat. Assoc. 83 (401), 123–127. https://doi.org/10.1080/01621459.1988.10478575.

Burchfield, M., Overman, H.G., Puga, D., Turner, M.A., 2006. Causes of sprawl: a portrait from space. Q. J. Econ. 121 (2), 587–633.

Bureau of Transportation Statistics, 2015. National highway planning network, 2015. https://www.bts.gov/archive/publications/national_transportation_atlas_database/2011/index.

Bureau of Transportation Statistics, 2018. U.S. Energy consumption by the transportation sector, 2018. https://www.bts.gov/content/us-energy-consumption-transportation-sector.

Cambridge Systematics Inc, 2017. "Employment Data for Planning: A Resource Guide." NCHRP08-36 Task 127. American Association of State Highway and Transportation Officials Standing Committee on Planning, Medford, MA.

Cervero, Robert, Kockelman, Kara, 1997. Travel demand and the 3Ds: density, diversity, and design. Transp. Res. D Transp. Environ. 2 (3), 199–219. https://doi.org/10.1016/S1361-9209(97)00009-6.

Dieleman, F.M., Dijst, M., Burghouwt, G., 2002. Urban form and travel behaviour: micro-level household attributes and residential context. Urban Stud. 39 (3), 507–527. https://doi.org/10.1080/00420980220112801.

Ding, Chuan, Wang, Donggen, Liu, Chao, Zhang, Yi, Yang, Jiawen, 2017. Exploring the influence of built environment on travel mode choice considering the mediating effects of car ownership and travel distance. Transp. Res. Part A-Policy and Practice 100 65–80. https://doi.org/10.1016/j.tra.2017.04.008. June.

Echenique, Marcial H., Hargreaves, Anthony J., Mitchell, Gordon, Namdeo, Anil, 2012. Growing cities sustainably does urban form really matter? J. Am. Plan. Assoc. 78 (2), 121–137. https://doi.org/10.1080/01944363.2012.666731.

Fan, Chao, Li, Wenwen, Wolf, Levi J., Myint, Soe W., 2015. A spatiotemporal compactness pattern analysis of congressional districts to assess partisan gerrymandering: a case study with California and North Carolina. Ann. Assoc. Am. Geogr. 105 (4), 736–753. https://doi.org/10.1080/00045608.2015.1039109.

Fee, Kyle, Hartley, Daniel Aaron, 2012. The relationship between city center density and urban growth or decline. Working Paper 1213. Federal Reserve Bank of Cleveland. http://econpapers.repec.org/paper/fipfedcwp/1213.htm.

Feldman, Barry E., 2005. Relative Importance and Value." SSRN Scholarly Paper ID 2255827. Social Science Research Network, Rochester, NY. https://papers.ssrn.com/abstract=2255827.

Fowler, C.S., Jensen, L., Rhubart, D., 2018. Assessing U.S. Labor Market delineations for containment, economic core, and wage correlation.

Fox, John, Weisberg, Sanford, 2011. An R Companion to Applied Regression. Second. Sage, Thousand Oaks CA. http://socserv.socsci.mcmaster.ca/jfox/Books/Companion.

Glaeser, E.L., Kahn, M.E., 2010. The greenness of cities: carbon dioxide emissions and urban development. J. Urban Econ. 67 (3), 404–418. https://doi.org/10.1016/j.jue.2009.11.006.

Glaeser, E.L., Kahn, M.E., Chu, C., 2001. Job Sprawl: Employment Location in US Metropolitan Areas. Brookings Institution, Center on Urban and Metropolitan Policy.

Gonzalez, Rafael C., Woods, Richard E., 2017. Digital Image Processing. Pearson, New York.

Grömping, Ulrike, 2006. Relative importance for linear regression in R: the package relaimpo. J. Stat. Softw. 17 (1), 1–27.

Handy, Susan L., 2005. Smart growth and the transportation-land use connection: what does the research tell us? Int. Reg. Sci. Rev. 28 (2), 146–167.

He, M., Glasser, J., Pritchard, N., Bhamidi, S., Kaza, N., 2019. Demarcating geographic regions using community detection in commuting networks [Physics], March. 1903.06029. http://arxiv.org/abs/1903.06029.

Herold, M., Scepan, J., C Clarke, K., 2002. The use of remote sensing and landscape metrics to describe structures and changes in urban land uses. Environ. Plan. 34 (8), 1443–1458.

Hesse, Markus, Rodrigue, Jean-Paul, 2004. The transport geography of logistics and freight distribution. J. Transp. Geogr. 12 (3), 171–184. https://doi.org/10.1016/j.jtrangeo.2003.12.004.

Federal Highway Administration, 2014. "2011 state statistical abstract." highway statistics series, 2014. https://www.fhwa.dot.gov/policyinformation/statistics/abstracts/2011/.

Isserman, A.M., 2005. In the national interest: defining rural and urban correctly in research and public policy. Int. Reg. Sci. Rev. 28 (4), 465–499.

Johnson, Jeff W., LeBreton, James M., 2004. History and use of relative importance indices in organizational research. Organ. Res. Methods 7 (3), 238–257. https://doi.org/10.1177/1094428104266510.

Jones, Christopher, Kammen, Daniel M., 2014. Spatial distribution of US household carbon footprints reveals suburbanization undermines greenhouse gas benefits of urban population density. Environ. Sci. Technol. 48 (2), 895–902. https://doi.org/10.1021/es4034364.

Kaza, N., 2013. The changing urban landscape of United States. Landsc. Urban Plan. 110, 74–86. https://doi.org/10.1016/j.landurbplan.2012.10.015.

Kaza, N., 2019. Landscape Shape Adjusted Compactness Index for Urban Areas (Working Paper).

Kneebone, Elizabeth, 2009. "Job Sprawl Revisited: the Changing Geography of Metropolitan Employment." Metro Economy Series. Brookings Institution, Washington, D. C. https://www.brookings.edu/research/job-sprawl-revisited-the-changing-geography-of-metropolitan-employment/.

Krizek, Kevin J., 2003. Residential relocation and changes in urban travel: does neighborhood-scale urban form matter? J. Am. Plan. Assoc. 69 (3), 265–281. https://doi.org/10.1080/01944360308978019.

Lee, Sungwon, Lee, Bumsoo, 2014. The influence of urban form on GHG emissions in the US household sector. Energy Policy 68 (May), 534–549. https://doi.org/10.1016/j.enpol.2014.01.024.

Li, Wenwen, Goodchild, Michael F., Church, Richard, 2013. An efficient measure of compactness for two-dimensional shapes and its application in regionalization problems. Int. J. Geogr. Inf. Sci. 27 (6), 1227–1250. https://doi.org/10.1080/13658816.2012.752093.

Li, Xiaoxiao, Li, Wenwen, Middel, A., Harlan, S.L., Brazel, A.J., Turner II, B.L., 2016. "Remote sensing of the surface urban heat island and land architecture in phoenix, Arizona: combined effects of land composition and configuration and cadastral–demographic–economic factors. Remote Sens. Environ. 174, 233–243. https://doi.org/10.1016/j.rse.2015.12.022. March.

Lindeman, R.H., Merenda, P.F., Gold, R.Z., 1980. Introduction to Bivariate and Multivariate Analysis. Glenview, Illinois (Scott, Foresman).

Liu, Chao, Shen, Qing, 2011. An empirical analysis of the influence of urban form on household travel and energy consumption. Comput. Environ. Urban Syst.Sustain. Urban. Dev. 35 (5), 347–357. https://doi.org/10.1016/j.compenvurbsys.2011.05.006.

Liu, Liang, Hwang, Taesung, Lee, Sungwon, Ouyang, Yanfeng, Lee, Bumsoo, Smith, Steven J., Tessum, Christopher W., et al., 2019. Health and climate impacts of future United States land freight modelled with global-to-urban models. Nat. Sustain. 2 (2), 105–112. https://doi.org/10.1038/s41893-019-0224-3.

Lopez, Russell, 2014. Urban sprawl in the United States: 1970-2010. Cities. Environ. (CATE) 7 (1). http://digitalcommons.lmu.edu/cate/vol7/iss1/7.

Lutsey, Nicholas, Sperling, Daniel, 2005. Energy efficiency, fuel economy, and policy implications. Transp. Res. Rec. 1941 (1), 8–17. https://doi.org/10.1177/0361198105194100102.

McCarty, J., Kaza, N., 2015. Urban form and air quality in the United States. Landsc. Urban Plan. 139, 168–179. https://doi.org/10.1016/j.landurbplan.2015.03.008. July.

McMillen, Daniel P., 2003. Identifying sub-centres using contiguity matrices. Urban Stud. 40 (1), 57–69. https://doi.org/10.1080/00420980220080161.

Newman, Peter W.G., Kenworthy, Jeffrey R., 1989. Gasoline consumption and cities. J. Am. Plan. Assoc. 55 (1), 24–37. https://doi.org/10.1080/01944368908975398.

Qiu, S., Kaza, N., 2017. Evaluating the impacts of the clean cities program. Sci. Total Environ. 579, 254–262. https://doi.org/10.1016/j.scitotenv.2016.11.119. February.

Ramsey, Kevin, Bell, Alexander, 2014. Smart Location Database Version 2.0 User Guide. Environmental Protection Agency. http://www.epa.gov/smartgrowth/pdf/sld_userguide.pdf.

Rotem-Mindali, Orit, Weltevreden, Jesse W.J., 2013. Transport effects of E-commerce: what can Be learned after years of research? Transportation 40 (5), 867–885. https://doi.org/10.1007/s11116-013-9457-6.

Seto, K.C., Fragkias, M., 2005. Quantifying spatiotemporal patterns of urban land-use change in four cities of China with time series landscape metrics. Landsc. Ecol. 20 (7), 871–888.

Sinnott, R.W., 1984. Virtues of the haversine. Sky Telesc. 68, 158. December.

Stephan, Craig H., Sullivan, John, 2008. Environmental and energy implications of plug-in hybrid-electric vehicles. Environ. Sci. Technol. 42 (4), 1185–1190. https://doi.org/10.1021/es062314d.

Stevens, Mark R., 2017. Does compact development make people drive less? J. Am. Plan. Assoc. 83 (1), 7–18. https://doi.org/10.1080/01944363.2016.1240044.

Stopher, Peter R., Greaves, Stephen P., 2007. "Household travel surveys: where are we going?" *Transportation research Part A: Policy and practice*, bridging research and practice. Synth. Best Pract. Travel. Demand. Model. 41 (5), 367–381. https://doi.org/10.1016/j.tra.2006.09.005.

Sugarbaker, Larry J., Eldridge, Diane F., Jason, Allyson L., Lukas, Vicki, Saghy, David L., Stoker, Jason M., Thunen, Diana R., 2017. "Status of the 3D Elevation Program, 2015." USGS Numbered Series 2016–1196. Open-File Report. U.S. Geological Survey, Reston VA. http://pubs.er.usgs.gov/publication/ofr20161196.

United States Geological Survey, 2014. NLCD 2011 Land Cover (2011 Edition, Amended 2014) - National Geospatial Data Asset (NGDA) Land Use Land Cover - ScienceBase-Catalog. U.S. Geological Survey. https://www.sciencebase.gov/catalog/item/581d050ce4b08da350d52363.

VanDerWal, Jeremy, Falconi, Lorena, Januchowski, Stephanie, Shoo, Luke, Storlie, Collin, 2014. SDMTools: species distribution modelling tools: tools for processing data associated with species distribution modelling exercises. https://CRAN.R-project.org/package=SDMTools.

Waldorf, Brigitte S., 2006. A continuous multi-dimensional measure of rurality: moving beyond threshold measures. In: Annual Meeting, American Agricultural Economics Association. Long Island, California.

Wickham, James, Stehman, Stephen V., Gass, Leila, Dewitz, Jon, Sorenson, Daniel G., Granneman, Brian J., Poss, Richard V., Baer, Lori Anne, 2017. Thematic accuracy assessment of the 2011 national land cover database (NLCD). Remote Sens. Environ. 191 https://doi.org/10.1016/j.rse.2016.12.026, 328341.

Yuan, Man, Song, Yan, Huang, Yaping, Hong, Shijian, Huang, Liejia, 2018. Exploring the association between urban form and air quality in China. J. Plan. Educ. Res. 38 (4), 413–426. https://doi.org/10.1177/0739456X17711516.