SUPPORTING INFORMATION

# Spatial distribution of U.S. household carbon footprints reveals suburbanization undermines GHG benefits of urban population density

*Christopher Jones*[*,1], *and Daniel M. Kammen*[*,1, 2, 3†]

## Summary of methods, figures, and tables

SUPPORTING INFORMATION

## Additional Figures

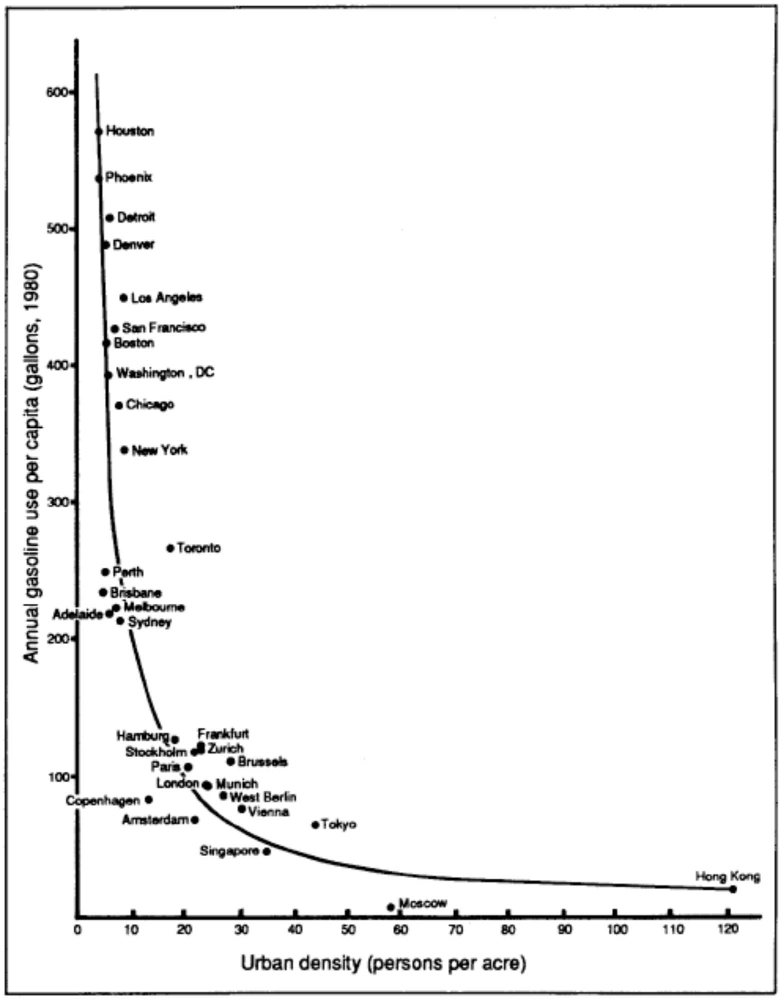


**Figure S-1.** Gasoline per capita vs population density (1980) reprinted from Newman and Kenworthy, 1989.[1]

For comparison, we have superimposed our results in Figure S-1b. The much larger set of cities included in our dataset shows a large range of gasoline consumption and very low correlation with population density ($R^2$=0.11). However, if only considering average cities for each population density (red diamonds), there is a strong correlation ($R^2$=0.86). This comparison demonstrates the effect of including all cities vs. only selected cities in such an analysis.

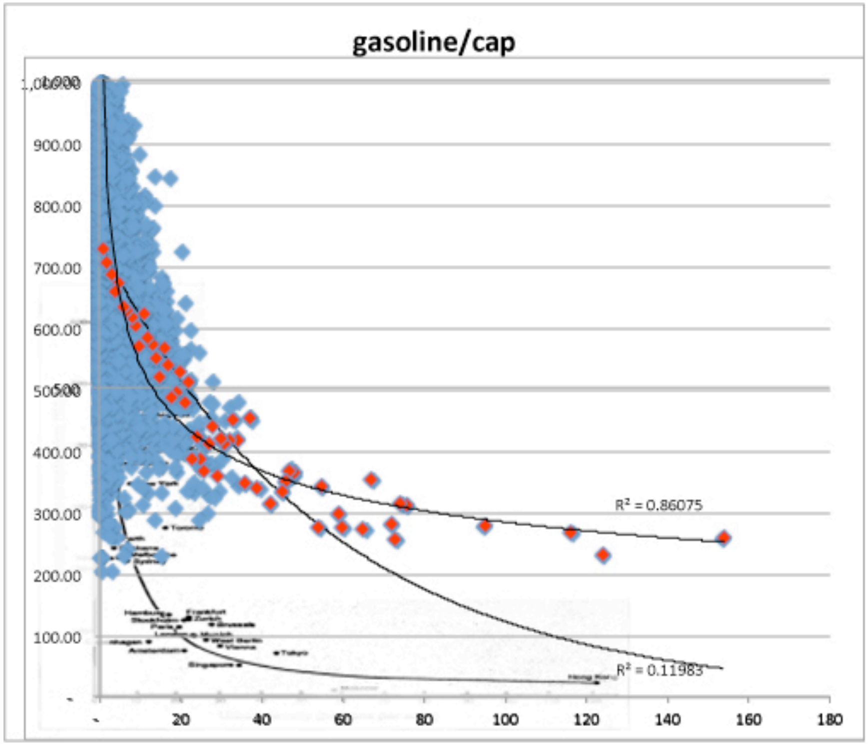


**Figure S-1b.** Gallons per capita in U.S. cities (blue diamond), average gallons per capita binned at increments of 0.1 on the x-axis (red diamonds) and results from Newman and Kenworthy (bottom line).

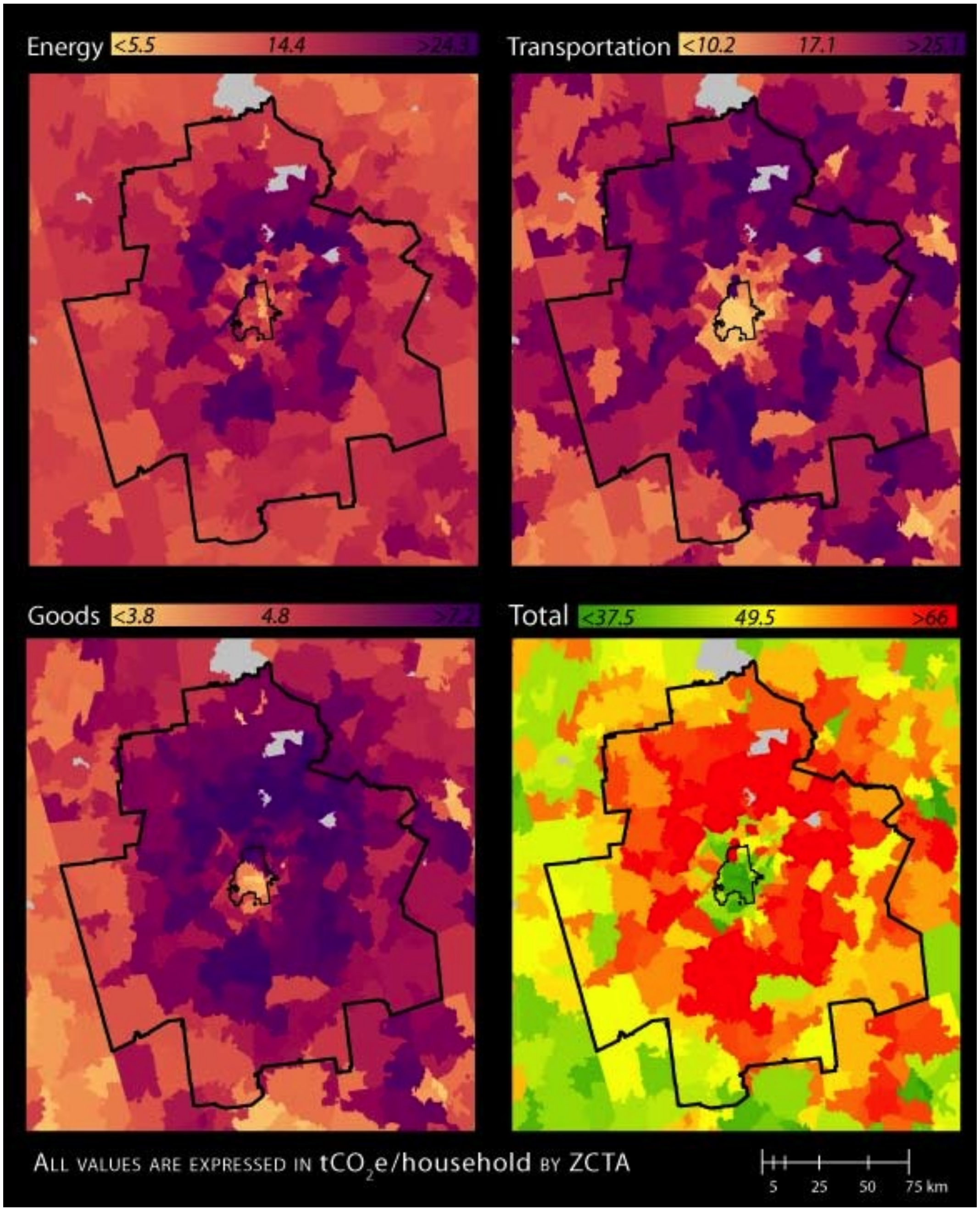


**Fig. S-2.** Composition of household carbon footprints in Atlanta, for Energy, Transportation, Goods and Total; see Figure 1 for definitions. All data are in metric tons $CO_2$e per household with colors reflecting the individual scales for each map, consistent with Figure 1. Outer dark line is the boundary of the 28-county metropolitan statistical area. Inner line is boundary of the city of Atlanta. The maps demonstrate relatively low carbon urban cores and high carbon suburbs for all major sources of household carbon footprints.

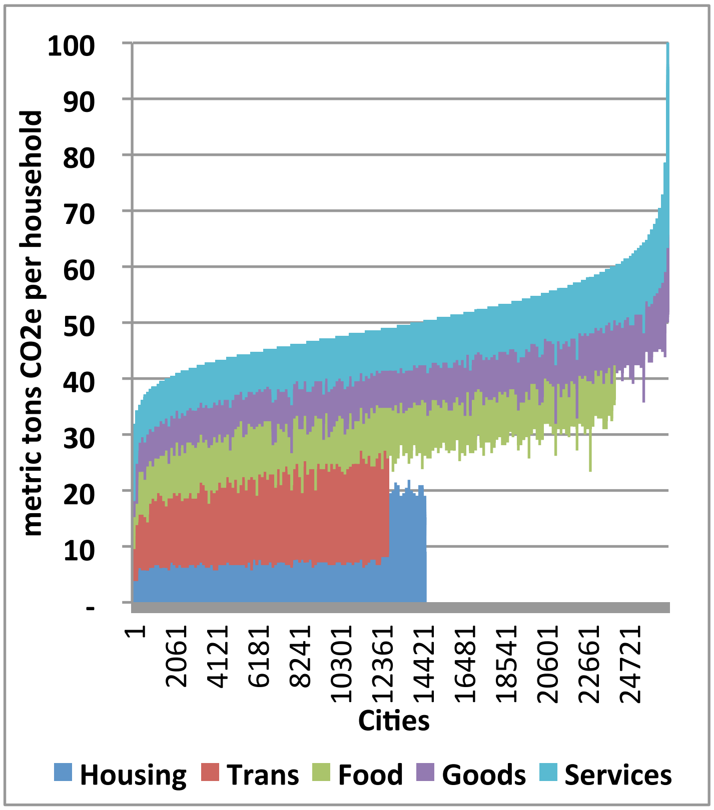


**Figure S-3.** Contribution of housing, transportation, food, goods and services to total household carbon footprints for 26,697 cities, sorted by total.

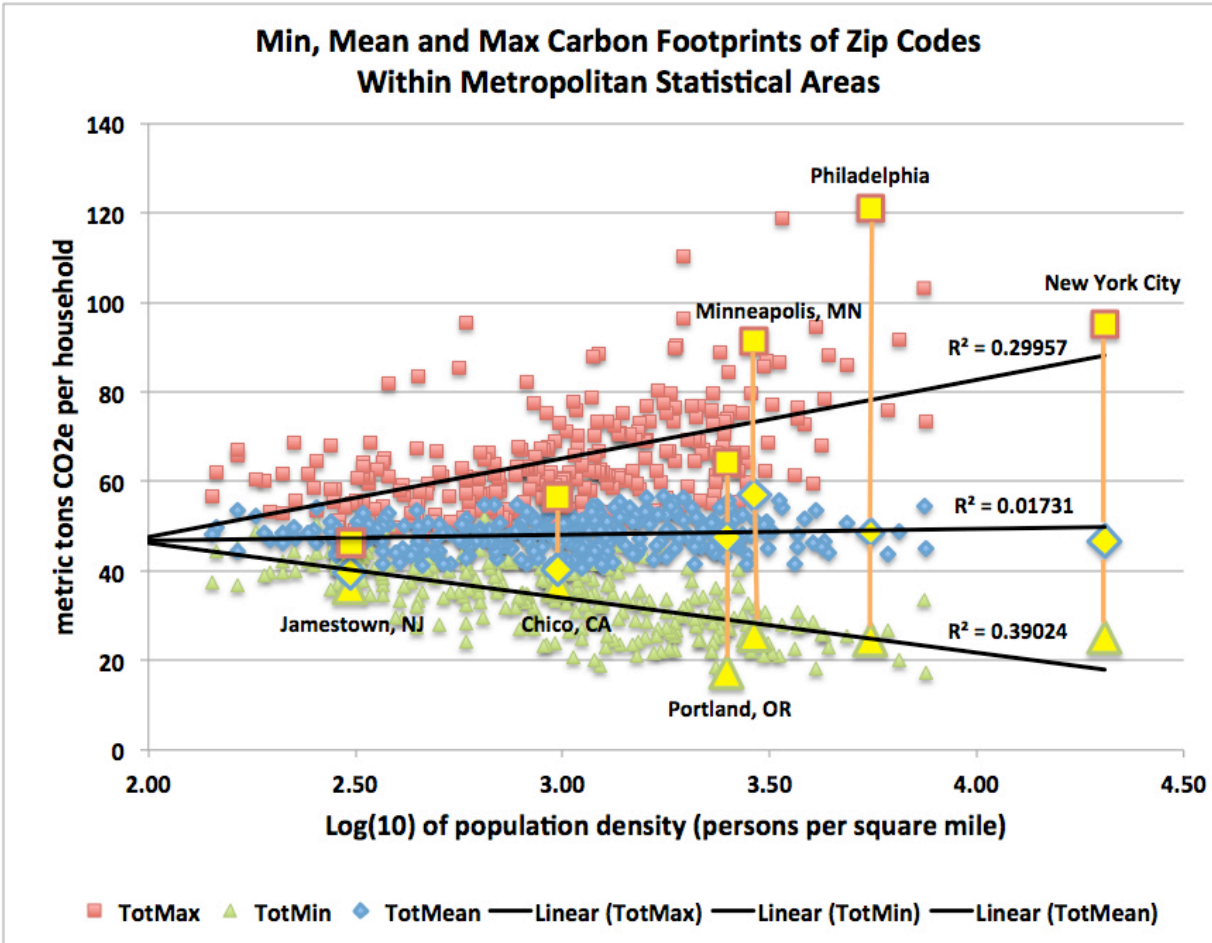


**Figure S-4.** Min, mean and max carbon footprints of zip codes within metropolitan statistical areas, ordered by log of population density (x-axis). Linear goodness of fit lines are drawn between min, mean and max values, including R-squared for each line. Results from six metropolitan regions are labeled, including Jamestown NJ (lowest mean HCF), Chico, CA (second lowest mean HCF), Portland, OR (includes zip code with lowest HCF), Minneapolis, MN (highest mean HCF), Philadelphia, PA (includes zip code with highest HCF), and New York, NY (highest population density).

## Materials and Methods

This study uses econometric analysis of national household survey results to estimate household consumption at the level of U.S. zip code tabulation areas (ZCTAs), roughly equivalent to U.S. zip codes. Model variables were chosen only if equivalent data from U.S. Census or other sources are known for ZCTAs. There are 31,914 ZCTAs in the model, covering essentially all populated areas of the 50 U.S. States. Eight separate linear and log-linear models were constructed for different categories of household consumption: one for vehicle miles traveled using the National Household Travel Survey,[2] five for household energy using the 2005 Residential Energy Consumption Survey,[3] one for food and another for other goods and services using the Consumer Expenditures Survey.[4] Additional datasets and methods were used to fill in average consumption values for water, waste, and building construction following previous work.[5]

The purpose of these models is prediction, and not inference. In order to improve the predictive power of the models we intentionally include collinear variables, e.g., the price of natural gas and the price of natural gas squared. As a result of collinearity, the correlation coefficients or t-statistics of independent variables should not be compared. Collinearity increases the goodness of fit of the model, which is necessary for more accurate prediction; however, collinearity confounds the relative contribution of independent variables as expressed by t-statistics or standardized beta coefficients. As long there is no interpretation of the regression coefficients, adding collinearity is a completely valid approach and is common in the literature for similar studies (e.g., Min et al, 2010).[6] In future iterations of the models we may choose to add interaction terms, which would further increase collinearity in order to enhance the goodness of fit.

# Electricity

Model results for Electricity are shown in Table S1.

**Table S-1.** Log-linear multivariate regression model of electricity consumption**.**

| | Unstanardized coefficients | | Standardized coefficients | | |
|---|---|---|---|---|---|
| | B | Std. error | Beta | t | Sig |
| (Constant) | 4.82 | 0.64 | | 7.6 | 0.000 |
| PRICEKWH | -6.93 | 0.28 | -0.43 | -24.5 | 0.000 |
| LNNHSLD | 0.31 | 0.01 | 0.25 | 21.5 | 0.000 |
| HEATKWH | 0.35 | 0.02 | 0.22 | 19.8 | 0.000 |
| TOTROOMS | 0.09 | 0.01 | 0.24 | 18.6 | 0.000 |
| CD65 | 0.00 | 0.00 | 0.21 | 17.5 | 0.000 |
| KWHPSQU | 1.93 | 0.18 | 0.17 | 10.9 | 0.000 |
| CA | -0.26 | 0.03 | -0.11 | -10.3 | 0.000 |
| LNINCOME | 0.09 | 0.01 | 0.11 | 9.7 | 0.000 |
| RURAL | 0.14 | 0.02 | 0.08 | 7.9 | 0.000 |
| DIV8 | -0.218 | 0.03 | -0.08 | -7.8 | 0.000 |
| DETTACHED | 0.14 | 0.02 | 0.09 | 7.3 | 0.000 |
| WHITE | 0.122 | 0.02 | 0.08 | 6.0 | 0.000 |
| REG2 | -0.09 | 0.02 | -0.06 | -5.1 | 0.000 |
| YEAR | 0.001 | 0.00 | 0.05 | 4.4 | 0.000 |
| BLACK | 0.110 | 0.03 | 0.05 | 4.2 | 0.000 |
| FL | -0.113 | 0.03 | -0.04 | -3.4 | 0.001 |
| OWN | 0.065 | 0.02 | 0.04 | 3.4 | 0.001 |
| NY | 0.086 | 0.03 | 0.03 | 2.5 | 0.011 |
| AGEHHMEM1 | -0.001 | 0.00 | -0.03 | -2.2 | 0.030 |

Dependent Variable: LNKWH

Weighted Least Squares Regression - Weighted by NWEIGHT

Model Summary

| R | R Square | Adjusted R Square |
|---|---|---|
| 0.78 | 0.608 | 0.607 |

ANOVA

| | Sum of squares | df | Mean Square | F | Sig |
|---|---|---|---|---|---|
| Regression | 33,916,985 | 19 | 1,785,104 | 357 | 0.000 |
| Residual | 21,839,244 | 4,362 | 5,007 | | |
| Total | 55,756,229 | 4,381 | | | |

Dependent variable: LNKWH = natural log of electricity consumption in kWh/year

Independent variables:

- PRICEKWH = price of electricity in $/kWh (EIA average residential price for the year 2005 at level of U.S. states)

- LNNHSLD = natural log of number of household members
- HEATKWH = household heats with electricity
- TOTROOMS = number of rooms
- CD65 = cooling degree days, base 65 (NOAA, 1971-2000 30-year Climate Normal of over 5,000 weather stations interpolated to each zip code in GIS)[7]
- KWHPSQU = price of electricity squared
- CA = Dummy variable for state of California (chosen for inclusion because it was a significant variable)
- LNINCOME = natural log of household income
- RURAL = percentage of households categorized as rural
- DIV8 = U.S. Census division 8
- DETTACHED = percentage of single detached homes
- WHITE = percentage of households headed by race coded as White / Caucasian
- REG2 = U.S. Census region 2
- YEAR = year home built
- Black = percentage of households headed by race coded as Black / African American

- FL = Dummy variable for U.S. state of Florida (chosen for inclusion because it was a significant variable)
- OWN = percentage of households owned by occupant
- NY = Dummy variable for U.S. state of New York
- AGEHHMEM1 = Age of the head of household

Greenhouse gas emission factors for electricity are provided by the eGRID database at the level of eGRID subregions.[8] The eGRID database aggregates emissions for each generator for thousands of power plants in the United States. Indirect “well-to-plug” emissions are assumed to increase generation emissions by 20% for electricity and natural gas, following the GREET model.[9]

## Natural gas

Heating fuel type varies significantly by region. Piped natural gas is typically not available in rural areas, and some large regions, including most of Florida. In order not to overestimate the use of natural gas in regions without natural gas connections we created three separate models to account for the fraction of homes in each zip code with different heating fuels.

Dependent variable: LNBTUNG = natural log of natural gas consumption in BTU/year

Independent variables:

- PNGSQ = price of natural gas squared (EIA average residential price for the year 2005 at level of U.S. states)[10]
- TOTROOMS = number of rooms
- HD65 = heating degree days, base 65 (NOAA, 1971-2000 30-year Climate Normal of over 5,000 weather stations interpolated to each zip code in GIS)
- YEAR = year home built
- LNNHSLD = natural log of number of household members
- BLACK = percentage of households headed by race coded as Black / African American
- HD65SQ = heating degree days, base 65, squared
- REG1 = Census region 1
- CD65 = cooling degree days, base 65
- DETTACHED = percentage of single detached homes
- AGEHHMEM1 = Age of the head of household
- DIV 4 = Census division 4

- OWN = percentage of households owned by occupant

- DIV7 = Census division 7

- DIV8 = Census division 8

- LNINCOME = natural log of household income

**Table S-2.** Log-linear model of annual natural gas consumption for percentage of homes with electric heating (RECS, 2005)

| | Unstandardized Coefficients | | Standardized Coefficients | | |
|---|---|---|---|---|---|
| | B | Std. Error | Beta | t | Sig. |
| (Constant) | 8.4 | 0.3 | | 27.5 | 0.000 |
| PNGSQ | -1,767 | 421.1 | -0.690 | -4.2 | 0.000 |
| TOTROOMS | 0.1 | 0.0 | 0.215 | 3.6 | 0.000 |
| PRICENG | 80.8 | 24.9 | 0.535 | 3.3 | 0.001 |
| CA | 0.4 | 0.2 | 0.164 | 2.8 | 0.006 |
| LNNHSLD | 0.3 | 0.1 | 0.159 | 2.7 | 0.007 |
| FL | -0.7 | 0.3 | -0.151 | -2.6 | 0.010 |

Dependent Variable: LNBTUNG
Weighted Least Squares Regression - Weighted by NWEIGHT

Model Summary

| R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|
| 0.474 | 0.225 | 0.206 | 1.016 |

ANOVA

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Regression | 73 | 6 | 12.086 | 11.707 | 0.000 |
| Residual | 250 | 242 | 1.032 | | |

Dependent variable: LNBTUNG = natural log of natural gas consumption in BTU/year

Independent variables:

- PNGSQ = price of natural gas squared

- TOTROOMS = number of rooms

- PRICENG = price of natural gas (EIA average residential price for the year 2005 at level of U.S. states)

- CA = Dummy variable for the state of California

- LNNHSLD = natural log of number of household members

- FL = Dummy variable for the state of Florida

**Table S-3.** Log-linear model of annual natural gas consumption for percentage of homes with heating oil as main heating fuel (RECS, 2005)

| | Unstandardized Coefficients | | Standardized Coefficients | | |
|---|---|---|---|---|---|
| | B | Std. Error | Beta | t | Sig. |
| (Constant) | 10.63 | 0.43 | | 24.74 | 0.00 |
| REG2 | -3.95 | 0.56 | -0.62 | -7.04 | 0.00 |
| NY | -1.11 | 0.28 | -0.39 | -4.02 | 0.00 |
| PNGSQ | -44.75 | 19.71 | -0.20 | -2.27 | 0.03 |
| TOTROOMS | -0.13 | 0.06 | -0.19 | -2.00 | 0.05 |

Dependent Variable: LNBTUNG

Weighted Least Squares Regression - Weighted by NWEIGHT

Model Summary

| R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|
| 0.686 | 0.471 | 0.442 | 1.063 |

ANOVA

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Regression | 72.5 | 4.0 | 18.135 | 16.038 | 0.00 |
| Residual | 81.4 | 72.0 | 1.131 | | |
| Total | 154.0 | 76.0 | | | |

Z

Dependent variable: LNBTUNG = natural log of natural gas consumption in BTU/year

Independent variables:

- REG2 = Census region 2

- NY = Dummy variable for state of New York

- PNGSQ = price of natural gas squared (EIA average residential price for the year 2005 at level of U.S. states)

- TOTROOMS = number of rooms

# Fuel Oil

**Table S-4.** Linear model of annual fuel oil consumption (RECS, 2005)

| | Unstandardized Coefficients | | Standardized Coefficients | | |
|---|---|---|---|---|---|
| | B | Std. Error | Beta | t | Sig. |
| (Constant) | 228.95 | 90.022 | | 2.543 | 0.011 |
| TOTROOMS | 75.846 | 8.985 | 0.434 | 8.441 | 0 |
| REG1 | 128.895 | 55.587 | 0.118 | 2.319 | 0.021 |
| OWN | -164.353 | 48.794 | -0.187 | -3.368 | 0.001 |
| AGEHHMEM1 | 2.836 | 1.043 | 0.134 | 2.72 | 0.007 |
| DIV1 | 99.678 | 39.622 | 0.133 | 2.516 | 0.012 |
| WHITE | -94.047 | 47.165 | -0.099 | -1.994 | 0.047 |

Dependent Variable: GALLONFO

Model Summary

| R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|
| 0.454 | 0.206 | 0.193 | 337.761 |

ANOVA

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Regression | 11,178,223 | 6 | 1,863,037 | 16 | 0.000 |
| Residual | 43,123,236 | 378 | 114,083 | | |
| Total | 54,301,460 | 384 | | | |

# Transportation

## Motor vehicles

Vehicle miles traveled (VMT) of motor vehicles driven by residents is modeled using the National Household Travel Survey.[2] Similar to household energy, only variable available at the level of ZCTAs are included in the model.

**Table S-5.** Log-linear model of household vehicle miles traveled (National Household Travel Survey, 2007)

| | Unstandardized Coefficients | | Standardized Coefficients | | |
|---|---|---|---|---|---|
| | B | Std. Error | Beta | t | Sig. |
| (Constant) | 8.16 | 0.10 | | 84.90 | 0.000 |
| HHVEHCNT | 0.34 | 0.01 | 0.39 | 45.60 | 0.000 |
| LNHHINCV | 0.17 | 0.01 | 0.16 | 20.29 | 0.000 |
| AVE_TIMETOWK | 0.00 | 0.00 | 0.10 | 14.16 | 0.000 |
| LNHHSIZE | 0.17 | 0.01 | 0.11 | 12.58 | 0.000 |
| RACEWHITE | 0.15 | 0.01 | 0.09 | 10.28 | 0.000 |
| PCT_DRIVE | 0.00 | 0.00 | 0.03 | 3.83 | 0.000 |
| Cen_d_wsc | 0.07 | 0.02 | 0.02 | 2.88 | 0.004 |
| CEN_D_NE | -0.06 | 0.03 | -0.02 | -2.19 | 0.028 |
| FOOD | 0.00 | 0.00 | -0.02 | -2.67 | 0.008 |
| CEN_D_W | -0.05 | 0.02 | -0.02 | -2.81 | 0.005 |
| REC | 0.00 | 0.00 | -0.02 | -2.90 | 0.004 |
| CEN_D_M | -0.09 | 0.03 | -0.02 | -3.17 | 0.002 |
| CEN_D_MA | -0.10 | 0.02 | -0.04 | -4.82 | 0.000 |
| lnHTPPOPD | -0.03 | 0.00 | -0.07 | -8.18 | 0.000 |

Dependent Variable: LNVMT
Weighted Least Squares Regression - Weighted by WTHHFIN

Model Summary

| R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|
| 0.621 | 0.385 | 0.384 | 38.25 |

ANOVA

| Model | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Regression | 10,748,754 | 14 | 767,768 | 525 | 0.000 |
| Residual | 17,158,757 | 11,730 | 1,463 | | |
| Total | 27,907,512 | 11,744 | | | |

## Description of Variables

- HHVEHCNT = number of vehicles per household
- LNHHINVCV = natural log of annual household income
- AVE_TIMETOWK = average minutes commuting to work

- LNHHSIZE = natural log of number of people in household
- RACEWHITE = percentage of residents whose race is “white” according to the U.S. Census
- PCT_DRIVE = percentage of commuters who drive to work instead of other modes
- Cen_d_wsc = West South Central Census Division
- CEN_D_NE = Northeast Census Division
- FOOD = number of food establishments in zip code
- CEN_D_W = West Census Division
- REC = number of recreation establishments in zip code
- CEN_D_M = Middle Census Division
- CEN_D_MA = Mid Atlantic Census Division
- LnPPOPD = natural log of population density (residents per square mile)

Annual $CO_2e$ = VMT / mpg * EF gasoline

Where,

VMT = vehicle miles traveled

MPG = 22 miles per gallon

EF gasoline = emissions factor of gasoline

## All other household emissions

All other sources of household emissions are extrapolated from Jones and Kammen,[11] which categorizes average household consumption and carbon footprints for 6 household sizes, 12 income brackets and 28 metropolitan regions using the 2008 Consumer Expenditures Survey. All combinations of income bracket and household size total 72 distinct household types, with corresponding consumption profiles from the detailed survey. We apply a simple linear regression model using household size and income as the two independent variables for each of the following dependent variables:

- air travel
- public transit
    - bus
    - light rail
    - heavy rail
    - electric rail
- water
- waste
- home construction
- food
    - meat & fish
        - beef

        - chicken and poultry
        - fish and seafood
        - lamb, pork
        - other meat
    - dairy
    - fruits & vegetables
    - grains and baked goods
    - Other food
- Goods
    - Clothing
    - Furniture and appliances
    - Other goods
- All services

The Consumer Expenditures Survey of the Bureau of Labor Statistics[4] provides estimates of consumer behavior across all categories of consumer spending. The survey consists of a national quarterly sample of ~15,000 in-person interviews and 3,200 detailed diaries. It is important economic instrument developed by the BLS to maintain the Consumer Price Index (CPI) as well as the CES, which is widely used in economic studies, including consumption-based greenhouse gas accounting. Following our previous work[11] food is estimated at 3 metric tons $CO_2$e per person.

## Validation of Model Results

Evaluating the predictive power of the model for all geographic locations is not possible due to lack of comparable data. We therefore compared model results to several existing studies and datasets in order to better understand how well the model predicts consumption and emissions at different geographic scales. Figure S2 summarizes results for four model comparisons for household electricity and natural gas for California counties (upper figures), vehicle miles traveled for U.S. states (lower left), and total household carbon footprints for 28 metropolitan regions (lower right). Actual natural gas consumption is within 20% of predicted values for 26 of the 30 counties, and within 15% for 23 counties. Actual electricity consumption is within 20% of predicted values for 23 of 30 counties. Actual VMT is within 20% of predicted values for 80% of U.S. states. The model tends to somewhat underestimate electricity and VMT for locations with relatively high values, thus differences between urban cores and suburbs described in the main paper are likely larger than estimated in this study. Total emissions for metropolitan statistical areas are well aligned, but somewhat higher in the current study compared to our previous work, which relied on the Consumer Expenditures Survey (CES) to estimate consumption. The difference may be due to weighting of sampled data, e.g., the CES may have included more persons in urban cores, while our current dataset is a population-weighted average of all persons in all zip codes within metropolitan statistical areas. Model results are comparable to other published studies. In particular, energy results are quite similar to Min et al. (2010)[6] and the goodness of fit (r-squared) of the transportation model is similar to Glaeser and Kahn (2010).[12]

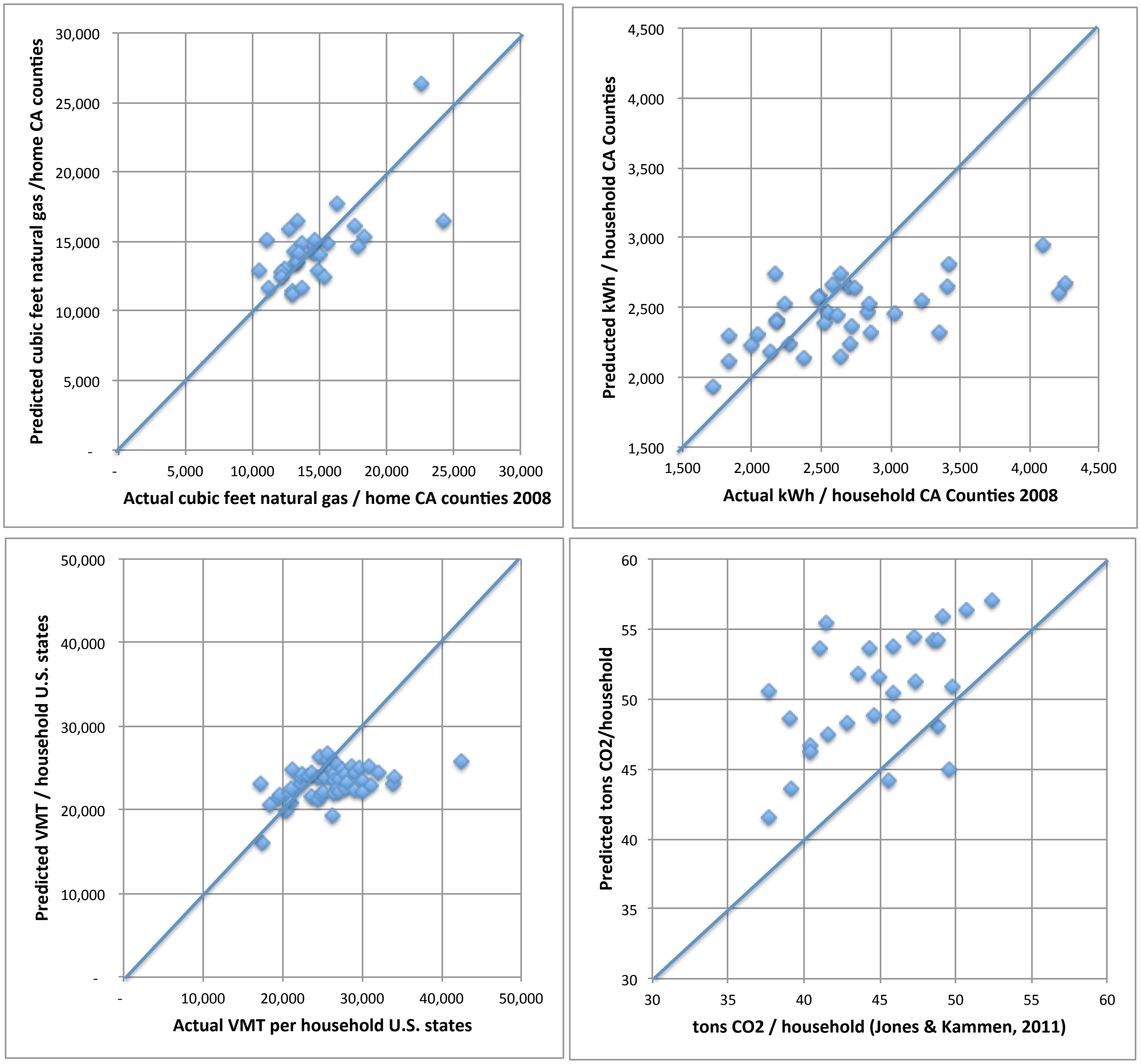


**Figure S-5**. Comparison of current results with other datasets. Upper left figure compares predicted natural gas consumption to average household natural gas consumption in 30 most populous California counties.[13] Upper right figure compares predicted electricity with California county data. Lower left figure compares predicted vehicle miles traveled to average household VMT for 50 U.S. states.[14] Lower right figure compares total household carbon footprints with results from previous work[11] using the Consumer Expenditures Survey for 28 metropolitan regions.

## Uncertainty

These results should be understood in the context of uncertainty and the methods used to derive the estimates. We have used national survey data to predict consumption at fine geographic scales and have used average GHG emission factors to estimate emissions. This approach hides important regional differences. For example, while we estimated vehicle miles traveled for every zip code in the U.S. using locally-available data, we have assumed average vehicle fuel economy for all locations. We have also assumed similar diets, housing construction, water and waste-related emissions due to lack of regionally specific data. The results should be considered benchmarks by which more accurate local assessments may be compared; such an analysis would be akin to determining level of efficiency compared to what might be expected from similar U.S. locations. The model shows expected consumption given the variables known at the level of zip codes. Local energy policies are reflected in the model only to a certain degree, by inclusion of some states as dummy variables.

The results from this analysis suggest sharp differences between urban and suburban households. The model likely understates these differences as it does not consider differences in motor vehicle fuel efficiency, which is likely higher in city centers that require smaller vehicles. City centers have also been shown to be more politically liberal and more likely to support climate change policy, including fuel-efficient vehicles, as well as other energy efficiency measures that are not captured in this model. Also, as noted under model validation above, the model tends to underestimate emissions for locations with relatively high consumption.

The primary purpose of our paper is not to highlight which model variables have the strongest impact on HCF, particularly since we have been selective about which variables to include (i.e., only those available at the zip code level). Rather, the purpose is to build the strongest model possible to predict energy consumption, VMT, consumption, etc, at fine geographic resolution. As a result of multicollinearity between several variables in our consumption models (Tables S1-S6), variables should not be ranked based on the relative importance.

In Table 3 in the main paper, on the other hand, we have attempted to explain relative contribution of independent variables on our results. For this purpose we created a single table of results from our regression models of energy, motor vehicles and consumption, in addition to the most important independent variables used in those models, for every zip code. We then created several multivariate linear regression models of these results in order to analyze the relative contribution of independent variables. We ran these models for the entire dataset of zip codes as well as for subsets of the data for principal cities (urban cores) and suburbs. Unlike our regression models of energy, transportation and consumption, for which prediction was the objective and collinearity was therefore not a relevant concern, the objective of this analysis was interpretation of causation so understanding of collinearity was essential. Coefficient correlations matrices and estimate of variance inflation factors is included in Table S7. The variance inflation factors (VIF) of independent variables for the full model and suburbs were never greater than 2.0 in the full model, and 3.2 when just considering principal cities. We chose models 2, 3 and 4 in Table 3 in the main paper to highlight the effect of colinearity, in particular the effects of population density and income. First, population density becomes not only a significant variable, controlling for all others, but it also becomes strongly correlated with rooms per household (a proxy for home size) and vehicles per household. The correlation between population density and income also becomes strongly negative, whereas in the full model the correlation is only slightly positive (i.e., relatively dense suburbs are richer and relatively dense principal cities are poorer). Entering variables in stepwise fashion was also helpful to understand interaction between variables. The variable with the strongest correlation with total $CO_2$ per household is number of vehicles per household. When controlling for income, the standardized beta coefficient decreases from 0.789 to 0.613 and further decreases to 0.338 (lower than the beta coefficient of income = 0.499) when controlling for all 5 other independent variables included the full model.  Income correlates positively with population density when considering all zip codes, but slightly negatively when only considering zip codes in principal cities. VIF for all variables is under 2.0 for the full dataset, less than 3.2 for cores and less than 2.4 for suburbs. VIF near 3 indicates some collinearity, as is apparent in the correlation matrices; however, the level of collinearity is not severe enough to warrant excluding any variables from the model. As a rule of thumb collinearity is considered severe if VIF is over 10, and even then may not require excluding variables from the model.[15]

**Table S-6.** Pearson's coefficient correlations matrices of results presented in Table 3 - Model 1 in main paper and VIF.

**All**

| | Total | # vehicles | annual hh income | gCO2/kWh | # rooms | log persons per hh | log pop. density | VIF |
|---|---|---|---|---|---|---|---|---|
| Total | 1.00 | 0.79 | 0.69 | 0.30 | 0.76 | 0.42 | (0.04) | |
| # vehicles | 0.79 | 1.00 | 0.39 | 0.21 | 0.61 | 0.37 | (0.07) | 1.88 |
| annual hh income | 0.69 | 0.39 | 1.00 | (0.17) | 0.55 | 0.18 | 0.34 | 1.87 |
| gCO2/kWh | 0.30 | 0.21 | (0.17) | 1.00 | 0.12 | 0.07 | (0.21) | 1.19 |
| # rooms | 0.76 | 0.61 | 0.55 | 0.12 | 1.00 | 0.25 | (0.02) | 2.03 |
| log persons per hh | 0.42 | 0.37 | 0.18 | (0.07) | 0.25 | 1.00 | 0.03 | 1.19 |
| log pop. density | (0.04) | (0.07) | 0.34 | (0.21) | (0.02) | 0.03 | 1.00 | 1.24 |

**Cores**

| | Total | # vehicles | annual hh income | gCO2/kWh | # rooms | log persons per hh | log pop. density | VIF |
|---|---|---|---|---|---|---|---|---|
| Total | 1.00 | 0.84 | 0.75 | 0.35 | 0.84 | 0.47 | (0.48) | |
| # vehicles | 0.84 | 1.00 | 0.56 | 0.21 | 0.73 | 0.43 | (0.56) | 3.21 |
| annual hh income | 0.75 | 0.56 | 1.00 | (0.01) | 0.56 | 0.09 | (0.19) | 1.85 |
| gCO2/kWh | 0.35 | 0.21 | (0.01) | 1.00 | 0.22 | (0.06) | (0.24) | 1.17 |
| # rooms | 0.84 | 0.73 | 0.56 | 0.22 | 1.00 | 0.41 | (0.42) | 2.56 |
| log persons per hh | 0.47 | 0.43 | 0.09 | (0.06) | 0.41 | 1.00 | (0.17) | 1.46 |
| log pop. density | (0.48) | 0.56 | (0.19) | (0.24) | (0.42) | (0.17) | 1.00 | 1.55 |

**Suburbs**

| | Total | # vehicles | annual hh income | gCO2/kWh | # rooms | log persons per hh | log pop. density | VIF |
|---|---|---|---|---|---|---|---|---|
| Total | 1.00 | 0.78 | 0.73 | 0.33 | 0.80 | 0.38 | (0.08) | |
| # vehicles | 0.78 | 1.00 | 0.42 | 0.19 | 0.58 | 0.39 | (0.14) | 1.84 |
| annual hh income | 0.73 | 0.42 | 1.00 | (0.12) | 0.67 | 0.15 | 0.29 | 2.29 |
| gCO2/kWh | 0.33 | 0.19 | (0.12) | 1.00 | 0.11 | (0.05) | (0.19) | 1.15 |
| # rooms | 0.80 | 0.58 | 0.67 | 0.11 | 1.00 | 0.21 | 0.00 | 2.38 |
| log persons per hh | 0.38 | 0.39 | 0.15 | (0.05) | 0.21 | 1.00 | (0.01) | 1.20 |
| log pop. density | (0.08) | (0.14) | 0.29 | (0.19) | 0.00 | (0.01) | 1.00 | 1.20 |

Quantification of uncertainty for the current dataset was not possible. While sampling error is available in national household surveys, including only this form of uncertainty would be misleading since many other sources of uncertainty exist, including measurement error, aggregation error associated with deriving average emission factors, model errors associated with using a limited number of variables, and other sources of error. The current paper draws limited conclusions that are strongly represented by the dataset and should not be greatly affected by uncertainty.

# References


(1) Newman, P. W. G.; Kenworthy, J. R. Gasoline Consumption and Cities. *J. Am. Plann. Assoc.* **1989**, *55*, 24–37.

(2) Highlights of the 2001 National Household Travel Survey http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/highlights_of_the_2001_national_household_travel_survey/index.html (accessed Jul 9, 2013).

(3) U.S. Energy Information Administration Residential Energy Consumption Survey, 2003.

(4) Bureau of Labor Statistics Consumer Expenditures Survey **2008**.

(5) C.M. Jones, W. Sullens, A. Dryden, J. Kantenbacher, D.M. Kammen *Measuring the Climate Impact of Residential Buildings: GreenPoint Rated Climate Calculator Version 2*; Report prepared for the California Air Resources Board 09-344; 2012.

(6) Min, J.; Hausfather, Z.; Lin, Q. F. A High-Resolution Statistical Model of Residential Energy End Use Characteristics for the United States. *J. Ind. Ecol.* **2010**, *14*, 791–807.

(7) National Ocean and Atmospheric Administration *Monthly Normals of Temperature, Precipitation, and Heating and Cooling Degree Days 1971-2000,*; Climatography of the United States NO. 81,; National Ocean and Atmospheric Administration, 2002.

(8) eGRID http://www.epa.gov/cleanenergy/energy-resources/egrid/index.html (accessed Aug 1, 2013).

(9) Argonne GREET Model http://greet.es.anl.gov/ (accessed Aug 1, 2013).

(10) U.S. Energy Information Administration (EIA) http://www.eia.gov/ (accessed Aug 1, 2013).

(11) Jones, C. M.; Kammen, D. M. Quantifying Carbon Footprint Reduction Opportunities for U.S. Households and Communities. *Environ. Sci. Technol.* **2011**, *45*, 4088–4095.

(12) Glaeser, E. L.; Kahn, M. E. The Greenness of Cities: Carbon Dioxide Emissions and Urban Development. *J. Urban Econ.* **2010**, *67*, 404–418.

(13) California Energy Consumption Database.

(14) Table 5-3: Highway Vehicle-Miles Traveled (VMT): 2005, 2010 http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/state_transportation_statistics/state_transportation_statistics_2011/html/table_05_03.html (accessed Oct 22, 2013).

(15) O'brien, R. M. A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Qual. Quant.* **2007**, *41*, 673–690.