A Dynamic Spatial-Temporal Model of Urban Carbon Emissions for Data-Driven Climate Action by Cities

Constantine Kontokosta
Asst. Prof (Urban Informatics)
New York University
(CUSP & Tandon)
ckontokosta@nyu.edu

Yuan Lai
PhD Candidate (Urban Systems)
New York University
(CUSP & Tandon)
yuan.lai@nyu.edu

Bartosz Bonczak
Associate Research Scientist
New York University
(CUSP)
bartosz.bonczak@nyu.edu

Sokratis Papadopoulos
PhD Candidate (Civil Eng.)
New York University
(CUSP & Tandon)
sokratis.papadopoulos@nyu.edu

Boyeong Hong
PhD Candidate (Urban Systems)
New York University
(CUSP & Tandon)
boyeong.hong@nyu.edu

Nicholas Johnson
PhD Candidate
(Dept. Computer Science)
University of Warwick,
Coventry, UK
nicholas.johnson@nyu.edu

Awais Malik
PhD Candidate (Civil Eng.)
New York University
(CUSP & Tandon)
awais.malik@nyu.edu

ABSTRACT

The reduction of energy use and greenhouse gas (GHG) emissions in the urban built environment has emerged as one of the primary grand challenges facing society in the 21st century. The Paris Climate Agreement calls on the global community to limit global temperature rise to 1.5 degrees Celsius through significant reductions in carbon emissions. Given the need for immediate action, cities and local governments are increasingly taking the lead in addressing this challenge, as cities are positioned to make substantial impacts through improvements to building and transit efficiency, and face dramatic consequences of inaction through increased risk from sea-level rise and extreme events. To achieve these goals, new data-driven methodologies are needed to identify and target energy efficiency and carbon reduction opportunities in the built environment at the building, neighborhood, and city-scale.

We propose a new high spatial-temporal resolution model of urban GHG emissions that combines data science, engineering, and urban planning methods and expertise to leverage new streams of data from public, private, and citizen-generated sources. Our aim is to advance carbon action in cities in a way that is efficient, scalable, and rapidly deployable. Our approach integrates numerous "big" data sources and develops a quantitative methodology that combines data-driven statistical and physical models of energy use and carbon emissions from buildings and transportation to generate a first-of-its-kind dynamic estimation of urban carbon emissions. Our data sources include the NYC Mayor's Office of Sustainability, NYC Department of Environmental Protection, NYC Department of Transportation, NYS Department of Transportation, U.S. Department of Energy, Earth Networks, PlumeLabs, Crimson Hexagon, and others. We then combine our model output with quality-of-life measures to better understand how urban emissions are associated with localized heat island effects, social sentiment, air quality, and socioeconomic disparities across New York City. This tool is designed to support city leaders and urban policymakers with an unprecedented view of localized carbon emissions to enable evidenced-based climate action policies based on rigorous scientific models.

Keywords

urban computing; data integration; greenhouse gas emissions; climate action; urban modeling

1. INTRODUCTION

The reduction of energy use and greenhouse gases (GHGs) emissions in the urban built environment has emerged as one of the primary grand challenges facing society in the 21st century. The Paris Climate Agreement calls on the global community to limit global temperature rise to 1.5°C through significant reductions in carbon emissions. Cities and urban areas are increasingly taking the lead in addressing this challenge given their ability to make substantial improvements to building and transit efficiency, and as a result of their vulnerability to the consequences of climate change such as increased risk of sea-level rise and extreme weather events. New York City, for example, recently announced a commitment to align with the Paris Agreement as part of its aggressive mandate to reduce GHG emissions by 80% from 2005 levels by the year 2050 [34]. Other cities have adopted similar goals including Los Angeles, Chicago, Boston, London, and Tokyo. To achieve these goals, however, new data-driven methodologies are needed that can identify and tar-

Bloomberg Data for Good Exchange Conference.
16-Sep-2018, New York City, NY, USA.

get energy efficiency and carbon reduction opportunities in the built environment at the building, neighborhood, and city-scale.

This work integrates numerous urban big data sources, including public and private administrative datasets and user-generated data, to develop a first-of-its-kind high resolution spatial-temporal model of urban carbon emissions. We combine data-driven models with physical models of energy usage and carbon emissions to estimate hourly city-wide emissions from building and transportation systems down to a spatial resolution of 500 meters. The results are then visualized through a web-based, interactive dashboard that aims to support city leaders and urban policymakers with an unprecedented, hyper-local view of carbon emissions to enable data-driven and evidenced-based climate action based on rigorous scientific models.

2. LITERATURE REVIEW

Rapidly growing data streams generated in and about cities and increasing computing capability have enabled researchers to observe and model urban phenomena at the urban scale. Recent studies have explored novel data collection and analytical methods to model spatial-temporal dynamics of subsystems in cities. Multiple urban observational studies reveal a temporal regularity described as the 'pulse of the city' in transportation, energy consumption, social media activity, urban mobility, and waste generation [11,19,21,33]. Other studies focus on spatial dynamics of air pollution, water consumption, and public health and how they relate to urban form, building typologies, urban forestry, or neighborhood socio-economic characteristics [15, 18, 31, 40]. A growing literature emphasizes the importance of localizing urban GHG emissions, since they originate from economic production and human activity occurring in urban environments [1]. GHG-related policy actions, including mitigation and adaptation, are often designed and implemented at the local scale, by multiple stakeholders within specific physical, political, and social contexts [2, 3]. From a citizen engagement perspective, localized data can also provide actionable information for residents and community groups to promote behavioral change and public awareness through better knowledge and empowerment [14, 20]. However, accurately modeling the spatial-temporal dynamics of GHG emissions has been a challenge, in part because high resolution data sources across the urban systems that influence emissions - physical, environmental, economic, socio-behavioral - are limited [6, 10]. Integrating cross-domain data, at multiple spatial and temporal scales from a variety of sources, can confound attempts to generate hyperlocal urban models due to non-standard data formats, structures, definitions, and the non-linear complexities in urban phenomena that undermine validation [23]. If we view a city a complex 'System of Systems (SoS)' embedded within a socio-technical context with non-trivial bio-physical variations, analyzing local GHG emissions requires a robust model design involving both data-driven and systems modeling approaches [13, 26]. Necessary sub-system data are often generated and managed by different sectors - public, private, non-profit, and individuals - that operate in administrative and computing silos [39].

3. METHODOLOGY

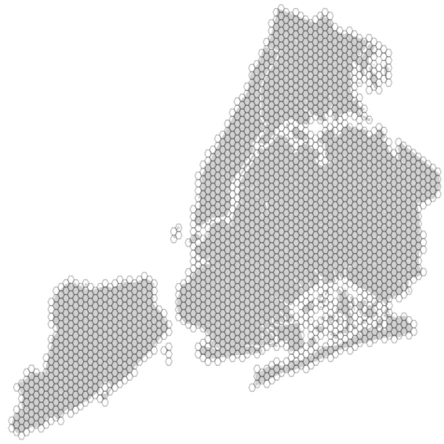

Figure 1: A 500m resolution hexagon-cell grid covering NYC.

3.1 Data integration

We begin by creating an indexed hexagonal tessellation grid to localize datasets available at various spatial resolutions. Our analysis is predicated on a 500 meter by 500 meter grid defined by the distance between the centroid points of two adjacent hexagonal cells. This grid contains a total of 2,169 cells covering the entire land area of New York City of approximately 300 square miles, represented by the WGS-84 Geographic Coordinate System (Figure 1). To quantify hourly GHG emissions for each cell ($GHGs_{i,t}$ in cell i at time t), we integrate numerous datasets from public and private sources, which can be further categorized as administrative data, sensor data, survey data, and transactional data (Table 1). Together, these represent the full spectrum of urban data, from static, point-based to dynamic, network-based typologies.

We collect administrative data from multiple agencies at the federal, state, and city levels, including the U.S. Department of Energy (building energy reference models), NYS Department of Transportation (traffic counts), NYC Department of City Planning (tax lot and zoning information), NYC Department of Buildings (building footprints and heights), NYC Department of Transportation (street segments, including length and width), and NYC Department of Parks and Recreation (street trees, parks and open space). We localize cross-domain parameters for each cell by spatial join, table join, aggregation, and downscaling to integrate different layers of information to the same spatial resolution (Figure 2). For instance, the Department of City Planning manages the Primary Land Use Tax Lot Output (PLUTO) database that contains detailed information on each of the more than 1,000,000 properties in NYC, with annual updates [25]. The data include information on land use type, zoning classification, floor area, building class, construction and alteration year, and assessed value, among other characteristics. By spatial join, we are able to identify specific buildings within each cell using the unique tax lot ID (Borough-Block-Lot or 'BBL' number) and can quantify aggregate land use and building parameters for the cell, such as total building gross square footage by land use type.

Table 1: Data Collection

Data	Source	Period	Spatial unit	Temporal unit	Format
Weather sensor data	Earth Networks	2015	Geopoint	5-minutes	csv
Weather stattion locations	Earth Networks	2015	Geopoint	Annual	csv
Land Use (PLUTO*)	NYC DCP	2015	Lot	Annual	shapefile
Energy & Water Consumption (LL84** Disclosure)	NYC MOS	2015	Tax Lot	Annual	csv
Reference Building Model Output	U.S. Dept. of Energy	2015	Regional	N/A	csv
Annual Average Daily Traffic	NYS DOT	2015	Road Segment	Annual	shapefile
Traffic short counts	NYS DOT	2015	Geopoint	Annual	shapefile
Roadway traffic count report	NYS DOT	Varies	Geopoint	Varies	pdf
Street Tree Census	NYC Dept. Parks & Recreation	2015	Geopoint	Decennial	shapefile
Air Quality	Plume Labs	2017	City	Hourly	csv
American Community Survey	U.S. Census Bureau	2015	Census Block Group	Annual (5-year average)	csv

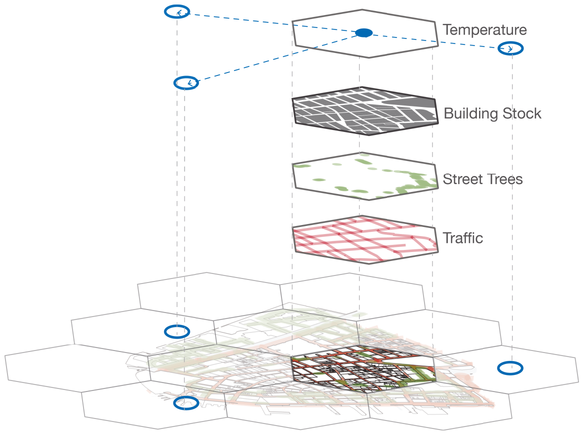

Figure 2: Data integration process.

To estimate GHG emissions from buildings and transport, we integrate multiple consumption and emissions datasets from from public and private sources. To model building GHG emissions, we use a combination of data-driven and engineering approaches. For all buildings larger than 50,000 square feet, we use NYC's Local Law 84 (LL84) energy disclosure data. LL84 mandates that large properties annually report their energy and water usage. The mandate applies to more than 20,000 privately-owned buildings, and all municipal buildings, accounting for approximately 45% of the citywide total building area, energy use, and carbon emissions [5]. We then use U.S. Department of Energy (DOE) reference building energy models, matched to specific building types, to estimate hourly consumption patterns, and adjust these figures based on building-specific LL84 energy efficiency data [9]. For buildings without LL84 data, we use the unadjusted reference model output. For traffic-related GHG emissions, the New York State Department of Transportation (NYSDOT) reports 2015 Annual Average Daily Traffic (AADT) data for major road segments statewide [28], including NYC. To define hourly vehicular traffic volume profiles, we use NYSDOT Short Counts data and the Roadway Traffic Count Hourly Reports associated with selected traffic counting locations. We then apply these counts to the NYC street network by identifying the street segment type for each street using the NYC-specific LION Single Line Street Base Map (LION) data available from NYC DCP [29].

To localize our building GHG emission models, we collect historical weather data from Earth Networks, one of the largest monitoring networks worldwide providing near-real-time localized weather data. This network covers NYC with 227 weather stations collecting data on temperature, wind speed, humidity, and light intensity at five-minute intervals [12]. We use Apache® Spark™ to parse and aggregate the 20+ GB *csv* files into hourly temperature measurements by station from 2015 to 2017. We then use the MapPLUTO dataset from NYC DCP to extract built area and building usage at the tax lot level to model and extrapolate localized weather in areas not covered by the sensor network [25].

We utilize several ancillary, privately-held data sets to explore the relationship between localized GHG emissions and quality of life across different neighborhoods and socioeconomic groups. We use the Plume Labs API to query City air quality data [30]. Plume Labs tracks hourly concentration levels of nitrogen dioxide (NO_2), sulfur dioxide (SO_2), ozone (O_3), particulate matter (PM2.5 and PM10), and several air quality indices, including the American Air Quality Index (AQI), the Chinese Air Quality Index (AQI CN), the Common Air Quality Index (CAQI), and the custom-designed Plume Index (PI). Due to the lack of access to historical data, we are only able to utilize air quality measurements for the corresponding period of time in 2017 as an representation of this potential functionality. We incorporate City-wide sentiment on air quality based on social media activity, by using Crimson Hexagon® âĂŹs social media analytics platform [8]. Using "air quality" and "air pollution" as key words and NYC as the spatial boundary, we query the database using a random sample of posts on social media platforms, including Twitter, Instagram, and Facebook. Within Crimson Hexagon, the ForSight™ analytical tool is a sentiment classifier that labels each post or tweet as 'positive', 'negative', or 'neutral' based on the post content. We then calculate a ratio between negative and positive posts to represent the hourly City-wide public sentiment on air quality. Finally, to identify socio-economic disparities in the spatial patterns of GHG emissions, we collect median household income data at the Census Block Group level from the U.S. Census American Community Survey for 2015, and localize data to each hexagonal cell by spatial join and aggregation.

3.2 Building GHG emissions estimation

We implement two approaches to estimate building-related GHG emissions: (a) statistical modeling and (b) physics-based modeling. The U.S. Department of Energy has developed energy models for 45 building typologies using *EnergyPlus*, a widely adopted building energy simulation engine [7].

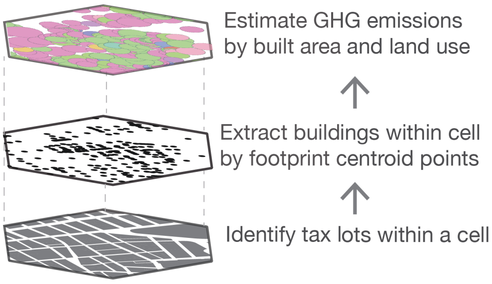

Figure 3: Illustration of building data aggregation by hexagonal cell.

For each typology, we simulate the annual building energy use intensity (EUI) in hourly intervals, expressed as:

$$EUI_t = kBtu_t / ft^2 \tag{1}$$

where $kBtu_t$ is the building's energy consumption in time t, and $t \in I\{1, ..., 8760\}$. We break down energy consumption by the two main fuel types encountered in NYC's building stock (electricity and natural gas) to obtain $EUIelec_t$ and $EUIgas_t$, similar to (Eq. 1). We also calculate the building annual load profile, as ratios of the hourly EUIs over the annual EUI (Eq. 2.

$$EUIratio_t = \frac{EUI_t}{\sum_{n=1}^{8760} EUI_t} \tag{2}$$

Using building class information extracted from PULTO data, we match each property across the five boroughs of NYC to the appropriate reference building model typology. Almost 95% of the more than 1,000,000 properties were successfully assigned to a corresponding energy model. Then, we multiply each building's modeled EUI values by its size to compute hourly energy consumption, based on the DOE model output (Eq. 3b). If the building is covered by LL84, we multiply its annual energy consumption by the typology's $EUIratio_t$ to estimate its hourly energy consumption.

If LL84 energy data available then

$$EC_{i,t} = \text{Annual EC} * EUIratio_t \tag{3a}$$

otherwise

$$EC_{i,t} = EUI_t * \text{Total Area} \tag{3b}$$

The result is an hourly energy use prediction for each building in the City, adjusted based on available measured energy efficiency data from LL84. Using the individual tax lot BBL identifier and data from MapPLUTO (the spatial PLUTO file that contains tax lot geometries), we calculate centroids of each property expressed by geographic coordinates in the WGS-84 geographic coordinate system. The location of each property includes an associated reference model output. We spatial join property centroid points with the hexagonal grid to aggregate energy consumption output for each cell (Figure 3). Finally, we estimate the hexagon's aggregate hourly GHG emissions in CO_2 equivalents (kg) using the appropriate U.S. Environmental Protection Agency (EPA) emissions coefficients for different fuel types (Eq. 4):

$$GHG_{X,t} = \textstyle\sum_i^N (ECelec_{i,t} * \beta_3) + (ECgas_{i,t} * \beta_4) \tag{4}$$

where $ECelec_{i,t}$ and $ECgas_{i,t}$ are the electricity and gas consumption for building i at time t, and β_3, β_4 are the GHG emission conversion coefficients for electricity and gas, respectively.

3.3 Transportation GHG emissions estimation

We utilize traffic monitoring data and spatial interpolation methods to estimate localized traffic GHG emissions. As mentioned in the data integration section, we use Annual Average Daily Traffic (AADT) data provided by the New York State Traffic Data Viewer (TDV) as our primary dataset and LION data as an ancillary dataset for the typologies of individual street segments. We also use AADT reports for specific road networks, which include measured hourly traffic volumes. Hourly GHG emissions derived from each street segment are then estimated through several steps. First, we classify all road segments into two types - vehicle-only and pedestrian-accessible - based on the LION data. A sample street for each road type - (1) vehicle only: FDR Drive and (2) pedestrian accessible: 14th street between 6th Avenue and 7th Avenue in Manhattan - is selected as a reference model for hourly traffic change patterns on all other street segments.

Once we have calculated the hourly traffic of the road network, we divide every road segment into 100-foot sub-segments associated with computed hourly traffic. This process is used to derive a homogeneous spatial unit for the entire road network to allow for aggregation to the unified hexagon cell grid. GHG emissions associated with the traffic network are computed using the Vehicle Miles Traveled (VMT) measure defined by the U.S. DOT to assess the total miles traveled by vehicles within a specified region for a specified time period. The âĂIJGreenhouse Gas Equivalences CalculatorâĂİ created by National Oceanic and Atmospheric Administration (NOAA) is used to calculate GHG emissions for each of the individual computed road sub-segments [27]. The tool converts VMT values into metric tons of CO2 equivalent (MTCO2e). VMT is calculated by multiplying the hourly traffic of each sub-segment to the length of the sub-segments (Eq. 5):

$$VMT_s = HT_s * L_s \tag{5}$$

Where HT_s is hourly traffic volume of the street sub-segment and L_s is the length of the sub-segment, set to 100 feet. On average, one (1) VMT is equivalent to 0.00053 MTCO2e. Consequently, each street sub-segment is associated with hourly GHG emissions according to (Eq. 6):

$$GHG_s = VMT_s * 0.00053(MTCO2e) \tag{6}$$

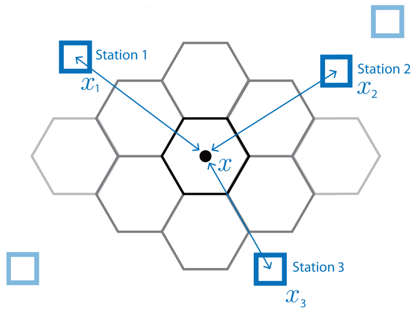

Figure 4: Spatial estimation of temperature by inverse distance weighting (IDW) interpolation.

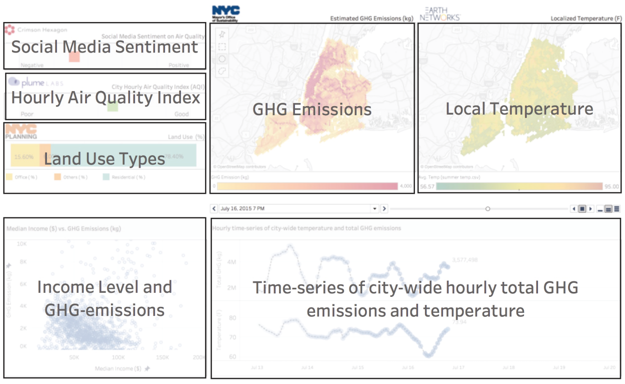

Figure 5: Dashboard interface components.

Finally, we aggregate the GHG emission values for each cell to estimate the hexagonâĂŹs hourly GHG emission (Eq. 7):

$$GHG_{X,t} = \sum_{i}^{N}(GHG_s) \tag{7}$$

3.4 Localizing weather sensor data

We develop a weather localization algorithm to estimate weather conditions at any geo-location in NYC by interpolating data from proximate weather stations (Figure. 4). For any given location x, we estimate its local temperature through Inverse Distance Weighting (IDW) interpolation. We first search the three nearest stations activated at time t to capture the temperature readings $u_i = u(x_i)$ for $i = 1,2,3$ (Eq. 8):

$$u(x) = \frac{\sum_{i=1}^{3} w_i(x)u_i}{\sum_{i=1}^{3} w_i} \tag{8}$$

where $w_i(x) = \frac{1}{d(x,x_i)}$ and d is the euclidean distance from interpolated point x to interpolating point x_i.

3.5 Developing the interactive visualization

We create our data visualization tool using Tableau® and integrate the model dashboard on a jQuery® enabled html web page [1]. Considering the size and comprehensiveness of our data and methodology, the web page includes multiple sections for users to navigate, including an introduction, dashboards, video tutorial, an interactive story board explaining the methodology, and a Carto® interactive map showing Lower Manhattan as a neighborhood case study. The dashboard is the primary interface for the hourly data visualization, which includes social media sentiment, air quality index, GHG emissions, temperature, and household income levels (Figure. 5). To demonstrate our output, we show two weeks of data (Monday - Sunday) for the periods from January 12th to the 19th and from July 13th to the 20th in 2017. Users can switch dashboards to explore how real-time GHG emissions vary between winter and summer seasons. Currently, we use these two one-week periods as inputs for the visualization tool due to limitations of the beta version of the interactive web-based environment. In the future, we plan to improve the current interface by integrating Tableau® dashboards and interactive data visualization using D3.js and Leaflet.js.

[1] http://www.urbanintelligencelab.org/wp-content/uploads/2017/11/Final_cut_short.mp4

4. RESULTS & DISCUSSION

Our model provides data-driven insights into local GHG emissions at high spatial and temporal resolutions. As a validation of our model estimations, we use official City statistics provided by the Inventory of New York City Greenhouse Gas Emissions for 2015 [6]. Compared to the reported annual citywide emissions, our model overestimated building emissions by less than 0.2% and underestimated traffic emissions by 38%. The latter can be explained, in part, by the fact that the AADT values do not cover the entire street network in NYC, and we also do not account for differences in emissions based on vehicle types. As expected, we find significant variations across NYC, which follow the range of land use types and building intensities, as well as the relative traffic volumes and congestion patterns along certain key segments of the surface transportation network (Figure. 6). It should be noted that buildings account for more than 70% of the City's total emissions, given the relative efficiency of the public transit system and thus lower per capita vehicle usage. We also note non-trivial variations in localized temperature, which highlights the significance of the urban heat island effect in dense urban environments. These variations are attributable to local wind patterns, traffic activity, built environment morphology and use types, and the presence of street trees and other natural features.

We make a number of simplifying assumptions in our proposed methodology, which result in certain limitations of our current approach. In the building emissions methodology, we rely on U.S. DOE prototypical building models, which correspond to a limited number of residential and commercial building types. Therefore, we are not capturing the entire stock of industrial properties, which can significantly contribute to overall GHG emissions. Similarly, many New York City properties are mixed-use, which do not match well with the generic DOE reference models. In addition, the DOE models provide outputs only for electricity and natural gas consumption. While these are the two major energy sources in the City, other fuels such as steam and heavy fuel oils are widely used and can have a differential impact on local emissions. Future versions of the our model will adjust and modify the models using more detailed building-specific data inputs, such as actual build-

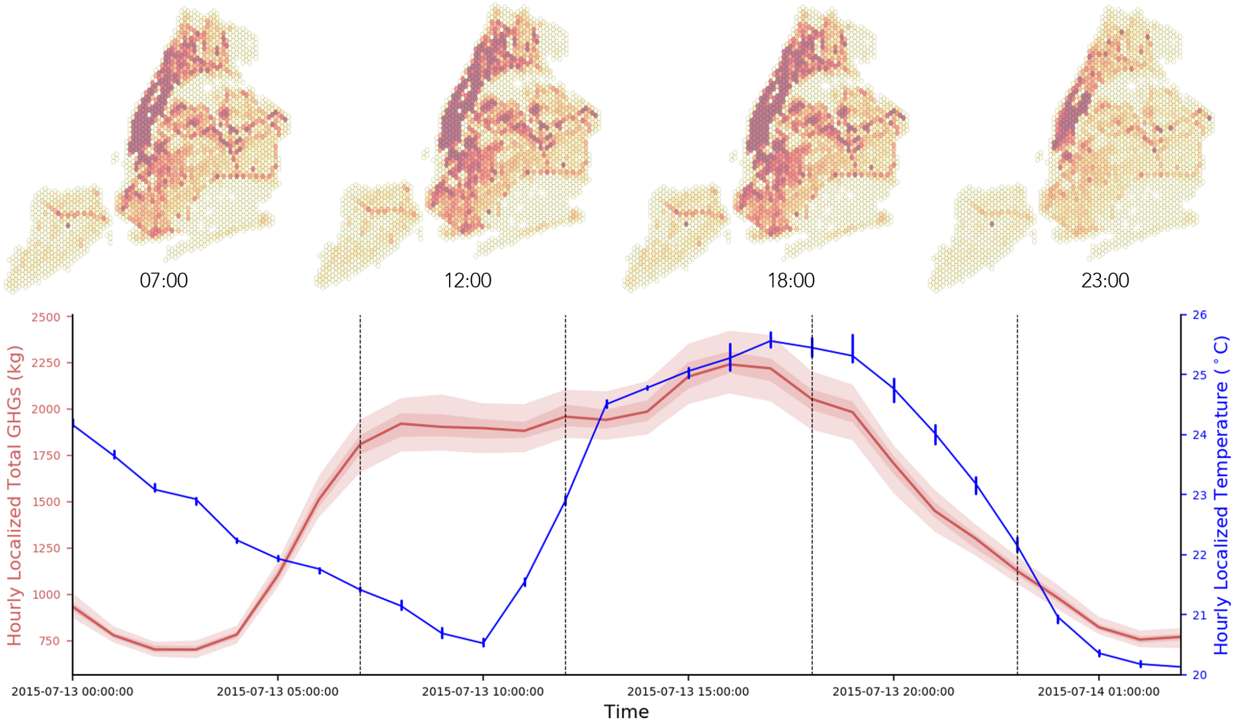

Figure 6: Spatial-temporal dynamics of local GHG emissions and temperature. Time-series plot shows intra-day hourly localized GHG emissions and hourly temperature with a 90% confidence interval, including hourly variations (red bands for GHGs and blue bars for temperature).

ing morphology parameters and hourly occupancy data, to improve the accuracy of the reference models and our estimates. Our ongoing work also involves developing detailed machine learning models to infer Citywide building energy consumption using property-specific data that can complement physics-based energy model hourly consumption estimates.

Our estimate of traffic GHG emissions relies on static, annual counts of traffic volume on major roads in the city. This approach does not account for seasonal variations, and we base our hourly traffic volume profiles drawn from a limited set of daily observations. Increasing the number and specificity of road typologies used can alleviate this problem to some extent. Another limitation is the fact that our current methodology does not distinguish between different vehicle classes, which impact emissions factors. We are working with Waze data to better account for real-time congestion and street segment speeds to adjust, or replace, the DOT counts for actual conditions. In addition, the DOT's traffic cameras, when parsed with computer vision techniques, could potentially improve the overall accuracy of our estimations by providing real-time traffic counts by segment for streets covered by these cameras.

5. CONCLUSION

Our framework provides a scalable platform and methodology for estimating, evaluating, and visualizing local GHG emissions in cities. By combining data from public, private, and non-profit sources, together with citizen-generated data, we are able to integrate statistical and physics-based models to estimate hourly GHG emissions for buildings and vehicles down to 500m grid resolution. The tool is designed to provide city leaders and urban policy-makers, working across sectors and agencies, to understand highly-localized patterns of GHG emissions across their city. Our prototype model uses the data-rich environment of New York City as a test case for a generalizable methodology that can be used in a range of cities, thus expanding potential insights through comparative studies and analysis.

Going forward, as part of a larger funded project, we plan to expand the scope and impact of our methodology in three significant ways. First, we will incorporate additional data, such as real-time traffic congestion from Waze and LiDAR-derived building shape parameters, to increase the robustness and accuracy of our results, and account for the limitations described above. This will also entail revising and deepening our methods to increase the local precision of our estimates. Second, we intend to develop a similar models and visualizations for five other cities, representing a cross-section of city sizes, regions, and urban morphologies. Finally, using the data and model outputs, we can analyze causal relationships between different city phenomena and GHG emissions. The objective will be to create a dynamic policy scenario tool to provide decision-makers with a process to examine the potential emissions impacts of a range of regulatory, economic, and behavioral interventions. In addition to providing an impact evaluation tool for GHG emissions, we will also incorporate impacts on economic activity, public health, and climate justice, with a specific focus on social equity and the distributional effects of policy decisions.

As cities increasingly take their place on the front lines of the battle against climate change, city leaders need new data-driven and evidenced-based insights to evaluate future scenarios and competing policy options. We believe our work significantly adds to the climate action toolkit by creating a new method to understand and estimate high-resolution emissions from cities. Although this work utilizes a significant collection of data sources, non-trivial resource limitations mean that the described methodology and visualization platform are a prototype version, with opportunities for meaningful improvements. Nonetheless, we have developed a scientifically-rigorous and actionable platform to advance climate action and meaningfully address one of the most significant challenges facing society.

6. ACKNOWLEDGMENTS

The authors would like to thank the United Nations Global Pulse for organizing 2017 Data for Climate Action Challenge, which the author team was selected for the Best Visualization Project Award. The authors also thank the various data providers, including Earth Networks, Plume Labs, Google/Waze, Crimson Hexagon, and the New York City Mayor's Office of Sustainability. This material is based, in part, upon work supported by the National Science Foundation under Grant No. 1653772.

7. REFERENCES

[1] C. Aall, K. Groven, and G. Lindseth. The scope of action for local climate policy: the case of norway. *Global environmental politics*, 7(2):83–101, 2007.

[2] M. Bond. Localizing climate change: stepping up local climate action. *Management of Environmental Quality: An International Journal*, 21(2):214–225, 2010.

[3] H. Bulkeley and M. M. Betsill. *Cities and climate change: urban sustainability and global environmental governance*, volume 4. Psychology Press, 2005.

[4] City of New York. New york city local law 84 benchmarking report, september 2014. Technical report, Mayor's Office of Long Term Planning and Sustainability, 2014. Accessed October 2017.

[5] City of New York. 2016 energy and water data disclosure (data for calendar year 2015), 2016. Accessed October 2017.

[6] City of New York. Inventory of new york city greenhouse gas emissions in 2015. Technical report, Mayor's Office of Long Term Planning and Sustainability, 2017. Accessed October 2017.

[7] D. B. Crawley, J. W. Hand, M. Kummert, and B. T. Griffith. Contrasting the capabilities of building energy performance simulation programs. *Building and environment*, 43(4):661–673, 2008.

[8] Crimson-Hexagon. Crimson-hexagon platform, 2017. Accessed October 2017.

[9] M. Deru, K. Field, D. Studer, K. Benne, B. Griffith, P. Torcellini, B. Liu, M. Halverson, D. Winiarski, M. Rosenberg, et al. Us department of energy commercial reference building models of the national building stock. 2011.

[10] S. Dhakal. Ghg emissions from urbanization and opportunities for urban carbon mitigation. *Current Opinion in Environmental Sustainability*, 2(4):277–283, 2010.

[11] G. Dobler, M. Ghandehari, S. E. Koonin, R. Nazari, A. Patrinos, M. S. Sharma, A. Tafvizi, H. T. Vo, and J. S. Wurtele. Dynamics of the urban lightscape. *Information Systems*, 54:115–126, 2015.

[12] Earth Networks. Earth Networks web page, 2017.

[13] J. Fiksel. Sustainability and resilience: toward a systems approach. *Sustainability: Science, Practice and Policy*, 2(2):14–21, 2006.

[14] K. R. Gurney. Track urban emissions on a human scale: cities need to understand and manage their carbon footprint at the level of streets, buildings and communities, urge kevin robert gurney and colleagues. *Nature*, 525(7568):179–182, 2015.

[15] R. K. Jain, J. M. Moura, and C. E. Kontokosta. Big data+ big cities: Graph signals of urban air pollution [exploratory sp]. *Ieee Signal Processing Magazine*, 31(5):130–136, 2014.

[16] C. Kontokosta. Modeling the energy retrofit decision in commercial office buildings. *Energy and Buildings*, 131:1–20, 2016.

[17] C. E. Kontokosta. Energy disclosure, market behavior, and the building data ecosystem. *Annals of the New York Academy of Sciences*, 1295(1):34–43, 2013.

[18] C. E. Kontokosta and R. K. Jain. Modeling the determinants of large-scale building water use: Implications for data-driven urban sustainability policy. *Sustainable Cities and Society*, 18:44–55, 2015.

[19] C. E. Kontokosta and N. Johnson. Urban phenology: Toward a real-time census of the city using wi-fi data. *Computers, Environment and Urban Systems*, 64:144–153, 2017.

[20] C. E. Kontokosta, N. Johnson, and A. Schloss. The quantified community at red hook: Urban sensing and citizen science in low-income neighborhoods. *arXiv preprint arXiv:1609.08780*, 2016.

[21] C. E. Kontokosta and C. Tull. A data-driven predictive model of city-scale energy use in buildings. *Applied energy*, 197:303–317, 2017.

[22] Y. Lai and C. E. Kontokosta. Measuring the impact of urban street trees on air quality and respiratory illness. *arXiv preprint arXiv:1710.11046*, 2017.

[23] T. K. Lim, M. Ignatius, M. Miguel, N. H. Wong, and H.-M. H. Juang. Multi-scale urban system modeling for sustainable planning and design. *Energy and Buildings*, 157:78–91, 2017.

[24] Y. Liu, R. Stouffs, A. Tablada, N. H. Wong, and J. Zhang. Comparing micro-scale weather data to building energy consumption in singapore. *Energy and Buildings*, 152:776–791, 2017.

[25] MapPLUTO. PLUTO and MapPLUTO data, 2016.

[26] W. K. Michener, T. J. Baerwald, P. Firth, M. A. Palmer, J. L. Rosenberger, E. A. Sandlin, and H. Zimmerman. Defining and unraveling biocomplexity. *AIBS Bulletin*, 51(12):1018–1023, 2001.

[27] National Oceanic and Atmospheric Administration. Greenhouse gas equivalencies calculator, 2017. Accessed October 2017.

[28] New York State DOT. Traffic data viewer, 2017.

[29] NYC Department of City Planning. Lion single line

street base map, 2017.

[30] PLUME Labs. Plume.io, 2017. Accessed October 2017.

[31] Z. Ross, I. Kheirbek, J. E. Clougherty, K. Ito, T. Matte, S. Markowitz, and H. Eisl. Noise, air pollutants and traffic: continuous measurement and correlation at a high-traffic location in new york city. *Environmental research*, 111(8):1054–1063, 2011.

[32] N. B. Schulz. Delving into the carbon footprints of Singapore - comparing direct and indirect greenhouse gas emissions of a small and open economic system. *Energy Policy*, 38(9):4848–4855, 2010.

[33] J. Shang, Y. Zheng, W. Tong, E. Chang, and Y. Yu. Inferring gas consumption and pollution emission of vehicles throughout a city. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1027–1036. ACM, 2014.

[34] The City of New York, Mayor Bill de Blasio. Aligning new york city with the paris climate agreement, 2017. Accessed March 2018.

[35] United Nations DESA/ Population Division. World population prospects 2017, 2017. Accessed March 2018.

[36] U.S. Census Bureau. American community survey 5-year estimates 2011-2015, 2017. Accessed October 2017.

[37] US Department of Energy. Commercial reference buildings, 2017. Accessed October 2017.

[38] US Environmental Protection Agency and US Department of Energy. Portfolio manager technical reference: Greenhouse gas emissions. Technical report, 2017. Accessed October 2017.

[39] Y. Zheng. Methodologies for cross-domain data fusion: An overview. *IEEE transactions on big data*, 1(1):16–34, 2015.

[40] Y. Zheng, F. Liu, and H.-P. Hsieh. U-air: When urban air quality inference meets big data. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1436–1444. ACM, 2013.