AGILE: GIScience Series, 5, 46, 2024. https://doi.org/10.5194/agile-giss-5-46-2024
Proceedings of the 27th AGILE Conference on Geographic Information Science, 4–7 Sept. 2024.
Editors: Alison Heppenstall, Mingshu Wang, Urska Demsar, Rob Lemmens, and Jing Yao.
This contribution underwent peer review based on a full paper submission.

Estimating hyperlocal traffic CO_2 by customizing spatial relationship: An analysis from Digital Footprint data

Ye Tian, Joao Porto de Albuquerque*, and Nick Bailey

Urban Big Data Centre, University of Glasgow, Glasgow, UK

Correspondence: Joao Porto de Albuquerque (joao.porto@glasgow.ac.uk)

Abstract. Globally, transport accounts for around one-fifth of CO_2 emissions. However, , leveraging the DF data in modeling hyperlocal traffic CO_2 and exploring the potential environmental justice is still underexplored. Here, we first extract traffic flows from the DF data, including individual GPS tracks, traffic counts, and car ownership rates in Glasgow, UK, then redefine the spatial relationship by incorporating traffic flows into the Spatial Weight Matrix (SWM), and finally predict the hyperlocal traffic CO_2 based on customized SWM. We find that, compared to traditional distance-based SWM, incorporating the real traffic flows into the SWM could better predict hyperlocal traffic emissions, with the Manski model performing best ($R^2 = 0.62$). Besides, the Manski model shows that income and car ownership rates are dominant factors related to traffic CO_2. Based on this, we reveal two aspects of environmental justice: 1). Distribution inequality - the high-income areas also have higher levels of car ownership rates, indicating higher barriers and challenges for low-income communities; 2). Contributor inequality - most high traffic CO_2 emissions are produced by nearby affluent areas with high car ownership rates, whereas the low-income areas suffer more traffic emissions produced by them, which indicates that disadvantaged groups bear the costs of emissions disproportionately generated by the advantaged. This pilot study explores the application of DF data in environmental monitoring, carbon justice, and climate mitigation to create an equitable and sustainable living environment.

Keywords. Digital Footprint (DF) Data, Traffic CO2, Spatial Weight Matrix (SWM), Manski Model, Environmental Justice

1 Introduction

Global warming, primarily caused by CO_2 emissions, is a great threat to human health and the environment. The average surface temperature in the UK has risen by 1.2°C since pre-industrial times ((IPCC), 2021). Therefore, many scholars and policymakers have focused on addressing the impacts of global warming through the flow of human mobility, such as urbanization, economic growth, and population migration (Y. Cai et al., 2023). In addition, urban design and planning have also been regarded as a third pathway beyond technological and market methods in reducing CO_2 in cities (Cervero et al., 2017).

For human mobility, direct measurements (e.g., portable GPS sensors) are cost-, time-, and labor-intensive, especially for long-term studies. As a representation of crowdsourced data, Digital Footprint (DF), referring to the "digital traces people leave behind as they conduct their lives" (Weaver & Gahegan, 2007), has emerged as an emerging approach to quantify human mobility due to its flexibility, large- scale, low-cost, and real-time. These geolocated DF data, such as cell phone data, represent people's presence in a given area and their movement between areas that make cities what they are (Dong et al., 2024) and serve as a great tool for hyperlocal traffic emissions estimation.

Besides human mobility, urban planning is responsible for around 40% of the total energy use and 17% of emissions in the UK (Climate Change Committee, 2022). As a result, scholars have recognized optimizing urban form is a potentially effective way to mitigate carbon emissions (Wang et al., 2018). Urban form represents the spatial distribution of urban elements inside a city (physical environment) and the interaction of urban factors within a specific urban system (perceived environment) (Fan et al., 2018). To illustrate, urban form affects how cities expand, as well as how land use, architecture, transportation, and infrastructure are allocated (Fang et al., 2016). Given that urban form influences mode choice, distances traveled, and trip frequency, with an impact on the overall energy needs (Silva et al., 2017), it has become an essential predictor in traffic CO_2 estimation. Neglecting urban form in carbon emissions modeling may influence the model accuracy and lead to an incomplete understanding of the urban form impact on further planning strategies (M. Cai et al., 2023). However, previous studies regarding the impacts of urban form on CO_2 emissions have not reached a consensus with both positive, negative, U- shaped, and non-significant correlation, which could partly be attributed to modulators, such as analysis scale and urban form measurement (Parshall et al., 2010).

To bridge the gaps above, we 1) leverage DF data (e.g., car ownership, vehicle counts, and individual GPS tracks) to customize the SWM and apply them to construct seven spatial panel models. The models with distance-based SWM are used as control groups; 2) incorporate both 2D and 3D urban from metrics based on nine land use types and six landscape metrics to clarify the urban form effects on traffic CO_2; 3) predict the traffic emission in 747 data zones in Glasgow based on the spatial panel model with the best performance and quantify the environmental justice. We anticipate the focus of hyperlocal traffic CO_2 emissions to be towards inclusive and equitable practices for localizing the SDGs (Sustainable Development Goals) with sensitivity to clean energy (SDG 7), intra-urban inequalities (SDG 10). and climate mitigation (SDG 13).

2 Methods

This study includes five major steps:

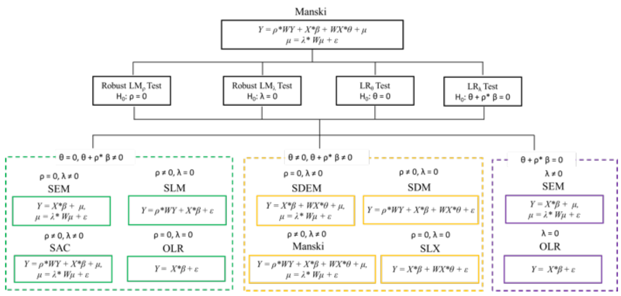
- Step 1: Extract traffic flow from individual GPS tracks, incorporate the traffic flow into SWM, and compute the 2D and 3D urban form with land use and meteorological datasets.
- Step 2: Estimate the traffic CO_2 in each DataZone based on traffic count and emission factors.
- Step 3: Construct the spatial panel models based on customized SWM and predict the traffic CO_2 in each Data Zone.
- Step 4: Interpret the traffic CO_2 distribution through hot spot analysis and the flow of socioeconomic factors with a strong correlation.

Table 1. Summary of data sources

Name	Data level	Periods	Temporal granularity	Spatial coverage	Source
Traffic Count	Station	2000 - 2021	Annually	UK	Department of Transport
CCTV	Station	2021	30 mins	Glasgow	UBDC
DVLA	LSOA	2021	Seasonal	Glasgow	Department of Transport
O-D Matrix (Huq)	Post Code	2021	Annually	Glasgow	UBDC
Public Transit Accessibility Indicators	DataZone	2021	Annually	UK	UBDC
Land use	Shapefile	2022	--	UK	Digimap
Meteorology	Station	2021	Hourly	UK	UK Met Office
Age groups	DataZone	2019	--	Scotland	National Records of Scotland
Population, income, education deprivation employment, mortality, and crime rates	DataZone	2020	Annually	Scotland	Scottish Government

2.1 Spatial panel model

Firstly, we test spatial autocorrelation. If spatial autocorrelation exists, we further conduct the Lagrange Multiplier (LM) test and Likelihood Ratio (LR) test to decide the selection of spatial panel models. Detailed procedures for model selection are shown in Figure 2. If spatial autocorrelation does not exist, we use the Ordinary Linear Regression (OLR), Random Forest (RF), and Artificial Neural Network (ANN) models directly.



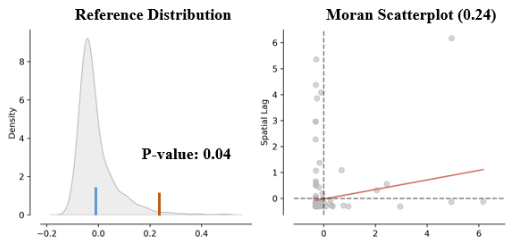
(W: Spatial Weight Matrix, μ and ε are error terms, LM: Lagrange Multiplier, LR: Likelihood Ratio, OLR: Ordinary Linear Regression, SEM: Spatial Error Model, SLM: Spatial Lag Model, SAC: Spatial Autoregressive Confused, SDEM: Spatial Dubin Error Model, SDM: Spatial Dubin Model, SLX: Spatially Lagged X Model)

Fig. 1 Top-down approach for selecting spatial panel models

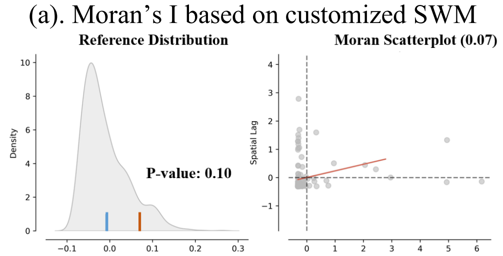
3 Results

3.1 Spatial autocorrelation test

From Figure 2, we find that there is a spatial autocorrelation issue of traffic emission based on our customized SWM (Figure 2a: Moran's I: 0.24, p-value: 0.04). However, the spatial autocorrelation issue is hidden (not statistically significant) by the traditional distance-based SWM (Figure 2b: Moran's I: 0.07, p-value: 0.10, the red line in 'Reference Distribution' is close to the blue line). This demonstrates that incorporating the car ownership and O-D flow into the SWM could better reflect the spatial autocorrelation of the residuals in traffic emission than the traditional distance-based SWM. Therefore, it is reasonable to use customized SWM to construct the spatial panel models for traffic emission estimation.



(a). Moran's I based on customized SWM



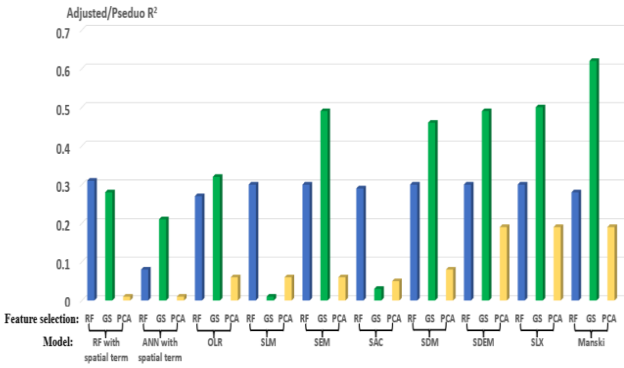
(b). Moran's I based on distance-based SWM

Fig. 2 Moran's I based on customized (a) and distance-based (b) SWM

3.2 Model Performance

We compared performances between seven spatial panel models (SEM, SLM, SAC, SDM, SDEM, SLX, and Manski) and two non-linear models (i.e., RF and ANN) which consider the flow of socioeconomic factors based on three feature selection methods (RF, PCA, and GS) in Figure 3. We also found that models constructed from customized SWM overall outperform the models from distance-based SWM, which reconfirms that incorporating the O-D flow into the SWM could predict traffic emissions better. The Lagrange Multiplier (LM) and Likelihood Ratio (LR) test results are all statistically significant.

Firstly, the model with GS feature selection outperforms the RF and PCA for all models. Therefore, we further compare the model performances based on GS. Secondly, incorporating the spatial autocorrelation of socioeconomic factors (*WX*), namely socioeconomic flow, improves the model performance substantially. For example, the R^2 increases from 0.32 (OLR) to 0.50 (SLX). Moreover, adding the *WX* with the spatial lag of emission (*WY*), SDM (R^2 = 0.46), or *WX* with the spatially autocorrelated error term (λ), SDEM (R^2 = 0.49), is also effective. However, the models only incorporate the *WY*, SLM (R^2 = 0.01), or *WY,* and λ, SAC (R^2 = 0.03), perform worse. This reconfirms that the flow of socioeconomic factors is indispensable. Additionally, we also compared the non-linear model RF (R^2 = 0.28) and ANN (R^2 = 0.21) by adding the *WX* and *WY*. Their performances are overall worse than other spatial panel models, particularly worse than the models that consider the λ, such as SEM and SDEM. To sum up, the Manski model, considering *WX*, *WY*, and λ, performs best (R^2 = 0.62).



(RF: Random Forest, PCA: Principal Component Analysis, GS: Greedy Stepwise, ANN: Artificial Neural Network, OLR: Ordinary Linear Regression, SEM: Spatial Error Model SLM: Spatial Lag Model, SAC: Spatial Autoregressive Confused, SDM: Spatial Dubin Model, SDEM: Spatial Dubin Error Model, SLX: Spatially Lagged X Model)

Fig. 3 Comparison of model performances based on different feature selection methods

3.3 Model results

Due to the better performance of the Manski model, we summarized the Manski models for selected 30 predictors after the multicollinearity test and GS feature selection in Table 1.

For socioeconomic factors, the car ownership rate shows the most obvious positive correlation with traffic emissions in both focal (80.151^{**}) and nearby areas (177.498^{***}), namely spatial autocorrelation term, similar to population density (0.033^{***} and 0.017 for focal and nearby, respectively). Higher population density and car ownership rates are often associated with increased urbanization and economic activity (e.g., Glasgow city center) and tend to have more traffic due to commuting, commercial activities, and service (Figure 1). Besides, crime rates (0.273^{***} and 0.272^{***} for focal and nearby, respectively) are positively related to traffic emissions, respectively.

It is noticeable that the focal income shows negative associations (-99.907) with traffic emission, but a positive correlation (345.757^{**}) for the nearby income. This demonstrates that low-income communities tend to suffer more traffic emissions in the focal areas, whilst the rich are more responsible for producing the traffic emissions in the nearby areas. It also accords with previous studies that, in the UK, the poorest areas emitted the least NOx and PM, but the rich emitted the highest (Barnes et al., 2019).

However, such a trend is reversed for education deprivation rates (-324.178^{**} and 57.407 for focal and nearby, respectively), which indicates that the low-education level population could be more responsible for producing traffic emissions, whereas the high-education counterparts receive more emissions.

For urban form factors, even for the same land use, it is essential to apply multiple landscape metrics to comprehensively characterize the impacts of land use patterns on traffic emissions (Tian & Yao, 2022). For example, the Largest Patch Index of Waterway (LPI-Waterway) shows the strongest negative correlation (-67.361^{***}) and the Class Area of Waterway (CA-Waterway) shows the strongest positive correlation (32.352).This situation also happens for the primary road as well (-2.177^{**}, 10.251, and 2.856 for the AI-Other road, CA-Primary road, and LPI-Primary road, respectively). Additionally, the distance to the nearest hospital shows a positive correlation (4.973^{***}) with traffic emission, but the nearest distance to GPs shows a negative one (-2.237). Both hospitals and GPs provide medical services, but GPs are more spatially dense than hospitals in Glasgow, which indicates that people drive more to hospital and walk more to GPs given the emergency and convenience, respectively. Finally, the spatial lag of emissions also shows a negative correlation (-0.610^{***}), and the error displays positive spatial autocorrelation effects (0.020^{**}).

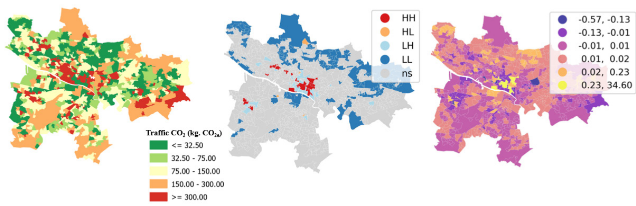
Table 1. Coefficients for the Manski model based on GS and customized SWM

Variable	Coefficient	Variable	Coefficient
Socioeconomic factors			
Focal		**Nearby**	
Car ownership rate	80.151^{**}	Lag-car ownership rate	177.498^{***}
Income	-99.907	Lag-Income	345.757^{**}
Education deprivation rate	-34.178^{**}	Lag-Education deprivation rate	57.407
Population density	0.033^{***}	Lag-Population density	0.017
Age 65 over rate	-2.842^{**}	Lag-Age 65 over rate	1.606
Crime rate	0.273^{***}	Lag-Crime rate	0.272^{***}
Urban form factors			
Building height variance	1.559	CA-Railway	4.808
Distance to nearest GPs	-2.237	PD-Agricultural land	0.286
Distance to nearest hospital	4.973^{***}	CA-Primary road	10.251
Distance to nearest primary school	-0.637	LPI-Primary road	2.856
Distance to nearest secondary school	0.862	AI-Primary road	-0.884^{***}
CA-Other road	-13.082	CA-Industrial land	-1.602
LPI-Other road	-11.012	PD-Industrial land	0.763
AI-Other road	-2.177^{***}	CA-Waterway	32.352
ED-Natural land	0.121	LPI-Waterway	-67.361^{***}
Other factors			
Lag-Emission (Lag-Y)	-0.610^{***}	λ in error term	0.020^{**}
Constant	78.825		

(***: p-value < 0.01, **: p-value < 0.05, *: p-value < 0.1 of z-statistics. Lag-: spatial autocorrelation term, GP: General Practice, CA: Class Area, LPI: Largest Patch Index, ED: Edge Density, PD: Patch Density, AI: Aggregation Index)

3.4 Prediction results

For predicted CO_2 distribution (Figure 4), higher CO_2 values mainly concentrate in the city center, northwestern, and eastern Glasgow, and lower values are located in the northern and western parts. As the hub for commercial and business activities, including shopping districts, corporate offices, and entertainment venues, the Glasgow city center typically has high mobility flows, vehicle usage, and traffic congestion, which all contribute to increased traffic emissions. Affluent neighborhoods (e.g., northwestern parts) with higher O-D flow and a preference for private transportation can also contribute to higher emissions. The higher positive and negative values of local Moran's I (Figure 3b) concentrate in the city center and northern Glasgow, respectively. It suggests the higher levels of spatial autocorrelation in these regions. The hot spot (Figure 3c) shows high clusters in the city center and some parts in northwestern and southwestern regions, which accords with previous studies on the spatial distribution of daily average flow in Glasgow (Li et al., 2022). Due to the high traffic emissions with short and local journeys (27% and 54% of trips in Glasgow in 2019 were less than 1km and 3km in length, respectively), the *Glasgow Transport Strategy* proposed to encourage more journeys on foot, by cycles and public transport, and shared mobility to create an efficient integrated system and decarbonize the traffic emissions (Glasgow City Council, 2022).



(a). Traffic CO_2 (b). Local Moran's I (c). Hot spot Map

Fig. 4 Prediction and hot spot results from LISA based on the customized SWM

4 Discussions

This study leverages the real O-D flows and car ownership rates to customize the SWM to predict hyperlocal traffic CO_2 emissions in Glasgow. We further illustrated the predicted emission distribution through hot spot analysis and environmental justice.

According to the spatial autocorrelation test, we find a statistically significant spatial autocorrelation issue based on customized SWM (Moran's I: 0.24, p-value: 0.04), whereas such phenomenon is hidden by the distance-based SWM (Moran's I: 0.02, p-value: 0.10). Therefore, it is necessary to incorporate the real O-D flow into the SWM to better reflect the spatial autocorrelation of traffic emission residuals for spatial panel models. Based on customized SWM, we reveal that incorporating the socioeconomic flow (*WX*) improves the model performance substantially and the Manski model, which incorporates the spatial lag term of both socioeconomic flow (*WX*), emissions (*WY*), and error terms (λ), performs best ($R^2 = 0.62$). It demonstrates that traffic emission is highly related to income (-99.07 and 345.757** for focal and nearby, respectively) and car ownership rates 80.151 and 177.498** for focal and nearby, respectively). The results suggest that the affluent and high car ownership regions are responsible for producing more traffic emissions, whilst low-income communities tend to suffer more traffic emissions. This is reconfirmed by the distribution of income and car ownership rates where high-income areas with high car ownership rates tend to produce more traffic CO_2. This suggests that the low-income communities areas are facing additional barriers and challenges in accessing essential services, opportunities for upward mobility, and a fair share of societal resources, which can lead to increased stress, limited social cohesion, a less productive and dynamic workforce, negative impacts on physical and mental health (Alegría et al., 2018). To mitigate such inequality, Glasgow receives a significant allocation due to its comparative deprivation and incidence of low-income households (Glasgow City Council, 2019).

For prediction results, we find a higher value of traffic emissions and the hot spot in Glasgow city center. A typical example to deal with this issue is the *Low-Emission Zones* in the city center proposed by *Glasgow City Council* and implemented in June 2023 (Glasgow City Council, 2019), which aimed to control traffic emissions. However, whether such a policy could reduce traffic emissions still needs further evaluation. For example, we do not know whether vehicles would bypass the zones to avoid penalties and produce more emissions. Due to the high traffic emissions with short and local journeys (27% and 54% of trips in Glasgow in 2019 were less than 1km and 3km in length, respectively), the *Glasgow Transport Strategy* proposed to encourage more journeys on foot, by cycles and public transport, and shared mobility to create an efficient integrated system and decarbonize the traffic emissions (Glasgow City Council, 2022). Additionally, different landscape metrics for the same land use may display a reversed correlation with traffic emissions. Therefore, it is essential to apply multiple landscape metrics and land uses simultaneously to comprehensively clarify the impacts of urban form on traffic emissions. This also explains why many previous studies have not reached a consensus on the urban form-traffic CO_2 relationship to a certain extent.

5 Conclusions

The motivation behind this research is to fill the gaps in applying the emergent DF data in the fields of environmental monitoring, social justice, and climate mitigation. In this study, we investigate how social inequality relates to traffic CO_2 emissions in Glasgow by leveraging real mobility data and improving hyperlocal traffic CO_2 prediction by redefining the spatial relationship. The conclusions show that, compared to traditional distance-based SWM, incorporating the real O-D flows into the SWM could better predict traffic emissions. For example, the nearby areas with high income and car ownership rates show a stronger correlation with traffic CO_2 than the focal areas. Given that income and car ownership rates are most correlated to traffic CO_2, the study also reveals two aspects of environmental justice issues: 1). There is inequality in income and car ownership distribution. The high-income areas also have higher levels of car ownership rates, and vice versa, which indicates additional barriers and challenges for low-income communities; and 2). There is inequality for emission contributors. The highest traffic CO_2 emissions are produced by nearby high-income and high car ownership areas which are away from the focal areas, and the low-income communities suffer more traffic emissions produced by them. It reveals that disadvantaged groups bear the costs of emissions disproportionately generated by the advantaged. The complex effects of socioeconomic factors on traffic CO_2 emissions and environmental justice require local governments to customize the measures for controlling CO_2 emissions to create healthy and equitable cities.

Acknowledgments

For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence (where permitted by UKRI 'Open Government Licence' or 'Creative Commons Attribution No- derivatives (CC BY-ND) licence may be stated instead) to any Author Accepted Manuscript version arising.

This work was supported by UKRI (UK Research and Innovation) - ESRC (Economic and Social Research Council) [ES/L011921/1 and ES/S007105/1]. Authors are thankful for the support of UBDC's staff, including Dr Rafael Verduzco for computing Public Transit Accessibility Indicators, Dr Faraz Malik Awanall for processing the GPS tracks from Huq and Tamoco, and other members of the Urban Big Data Centre (UBDC) for providing the data and feedback on the research work.

References

Alegría, M., NeMoyer, A., Falgàs Bagué, I., Wang, Y., & Alvarez, K. (2018). Social determinants of mental health: where we are and where we need to go. *Current psychiatry reports*, *20*, 1-13.

Barnes, J. H., Chatterton, T. J., & Longhurst, J. W. (2019). Emissions vs exposure: Increasing injustice from road traffic-related air pollution in the United Kingdom. *Transportation Research Part D: Transport and Environment*, *73*, 56-66.

Cai, M., Ren, C., Shi, Y., Chen, G., Xie, J., & Ng, E. (2023). Modeling spatiotemporal carbon emissions for two mega-urban regions in China using urban form and panel data analysis. *Science of The Total Environment*, *857*, 159612.

Cai, Y., Brock, W., & Xepapadeas, A. (2023). Climate change impact on economic growth: Regional climate policy under cooperation and noncooperation. *Journal of the Association of Environmental and Resource Economists*, *10*(3), 569-605.

Cervero, R., Guerra, E., & Al, S. (2017). *Beyond mobility: Planning cities for people and places*. Island Press.

Climate Change Committee (2022). *The Sixth Carbon Budget Buildings*. https://www.theccc.org.uk/wp-content/uploads/2020/12/Sector-summary-Buildings.pdf

Dong, L., Duarte, F., Duranton, G., Santi, P., Barthelemy, M., Batty, M., Bettencourt, L., Goodchild, M., Hack, G., & Liu, Y. (2024). Defining a city— delineating urban areas using cell-phone data. *Nature Cities*, 1-9.

Fan, C., Tian, L., Zhou, L., Hou, D., Song, Y., Qiao, X., & Li, J. (2018). Examining the impacts of urban form on air pollutant emissions: Evidence from China. *Journal of environmental management*, *212*, 405-414.

Fang, C., Li, G., & Wang, S. (2016). Changing and differentiated urban landscape in China: spatiotemporal patterns and driving forces. *Environmental science & technology*, *50*(5), 2217-2227.

Glasgow City Council (2019). *Glasgow Climate Plan*. https://www.glasgow.gov.uk/CHttpHandler.ashx?id=50623&p=0

Glasgow City Council (2022). *Glasgow Transport Strategy*. https://www.glasgow.gov.uk/transportstrategy

IPCC (2021). *Intergovernmental Panel on Climate Change Sixth Assessment Report Working Group I*. https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Full_Report.pdf

Li, Y., Zhao, Q., & Wang, M. (2022). Analysis the Influencing Factors of Urban Traffic Flows by Using New and Emerging Urban Big Data and Deep Learning. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *43*, 537-543.

Parshall, L., Gurney, K., Hammer, S. A., Mendoza, D., Zhou, Y., & Geethakumar, S. (2010). Modeling energy consumption and CO2 emissions at the urban scale: Methodological challenges and insights from the United States. *Energy policy*, *38*(9), 4765-4782.

Scottish Government (2021). *Scottish Index of Multiple Deprivation 2020*. https://simd.scot/#/simd2020/BTTTFTT/12/-4.2755/55.8504/

Silva, M., Oliveira, V., & Leal, V. (2017). Urban form and energy demand: A review of energy-relevant urban attributes. *Journal of Planning literature*, *32*(4), 346-365.

Tian, Y., & Yao, X. (2022). Urban form, traffic volume, and air quality: A spatiotemporal stratified approach. *Environment and Planning B: Urban Analytics and City Science*, *49*(1), 92-113. https://doi.org/https://doi.org/10.1177/239980832199582

Wang, S.-H., Huang, S.-L., & Huang, P.-J. (2018). Can spatial planning really mitigate carbon dioxide emissions in urban areas? A case study in Taipei, Taiwan. *Landscape and Urban Planning*, *169*, 22-36.

Weaver, S. D., & Gahegan, M. (2007). Constructing, visualizing, and analyzing a digital footprint. *Geographical Review*, *97*(3), 324-350.