Upscaling urban data science for global climate solutions

Felix Creutzig¹,², Steffen Lohrey², Xuemei Bai³, Richard Dawson⁴, Shobhakar Dhakal⁵, William Lamb¹, Timon McPhearson⁶, Jan Minx¹, Esteban Munoz⁷, Brenna Walsh⁸

¹Mercator Research Institute on Global Commons and Climate Change, Berlin
²Sustainability Economics of Human Settlements, Technical University Berlin
³Australian National University
⁴University of Newcastle
⁵Asian Institute of Technology, Bangkok
⁷UN Environment, Paris
⁸Future Earth, Montreal

Abstract. Cities have an increasingly integral role in addressing climate change and maintaining quality of life. To gain a common understanding on solutions, we require an adequate data presentation of urban areas, related GHG emissions, climate threats and socio-economic contexts. Here we review the current state of urban data science that is concerned with climate change. We outline three routes for upscaling urban data science for global climate solutions: 1) Mainstreaming and harmonizing data collection in cities worldwide; 2) Exploiting big data to upscale agent and/or building scale solutions while maintaining privacy; 3) Using high-resolution remote sensing data to automatize computation of first-order climate effects and solutions. We also argue that even in the absence of crucial background data, cities should systematically share knowledge and experiences on climate solution strategies. Collaborative efforts towards a joint data platform would provide the quantitative foundations of an emerging global urban sustainability science.
Introduction: Manhattan, Berlin, and everywhere else

“First we take Manhattan, then we take Berlin”. The locations of choice in Leonhard Cohen’s terrorist phantasm are a nightmare for the urban greenhouse gas emission accountant. Manhattan inhabitants emit low amounts of CO₂ per capita; but as part of the larger New York context, with ample commuting from less efficient building densities in surrounding areas, this positive example is compromised. Researchers are similarly perplexed as to how Berlin’s GHG emissions can be properly accounted, given its embedding in the wider Brandenburg area. Even Berlin’s area is unclear: should the administrative boundaries be chosen, including numerous parks and forests? Or only the built-up area, that itself requires proper definition? Or even the Brandenburg hinterland, which produces large amounts of renewable energy for the city itself?

These kind of urban data challenges derive from the open system character of cities. It is one of the main reasons that a coherent global urban data-based sustainability science is still in its infancy. But the urgency to tackle this conundrum could not be greater. With an expected increase in global urban population to 66% by 2050 (UN-DESA 2015), cities will play a decisive role in the reduction of global CO₂ emissions. By 2010, emissions from urban building and transport amounted to about 10 GtCO₂, of which 2.8 GtCO₂ came from transport, and 6.8 GtCO₂ from buildings (Creutzig et al 2016). By 2030, buildings could account for 12.6 GtCO₂ (UN Environment 2017). And still, as cities will play an increasingly important role both for local adaption but also global mitigation efforts, a systematic data-driven global urban sustainability science is still missing.

Gathering and interpreting urban emissions, climate, and ancillary data is troublesome for at least four reasons. First, boundaries of analysis are ambiguous and often inconsistent, making comparison on an equal footing challenging. This applies to all cities, not only to Manhattan and Berlin.

Second, data gathering is cumbersome, and can easily turn into multi-million-dollar efforts. The best data exist for the richest and most committed cities such as Paris and Los Angeles, which leads to a huge bias in the data representation of cities. The expected urban agglomerations with the highest growth rate are medium-sized cities and cities with a population lower than 1 million in Asia and Africa (UN-DESA 2015), and the group of 47 least developed countries (LDCs) has the biggest population growth rate at 2.4 per cent per year (UN-DESA 2017). Precisely these cities are those without the economic resources and infrastructure capacities to systematically collect and assess the required urban data.

Third, the data collected is often not the most useful for addressing urban climate solutions. For example, cities need building-stock data in order to understand and quantify the impact of retrofitting their buildings, and expanding their urban built-up area. And travel surveys remain often an insufficient base to address urban mobility challenges. This is also closely related to the first issue: the data taken usually are of varying quality and make intercity comparisons cumbersome.

Forth, qualitative data are often missing in data gathering. Some of the most policy-relevant information are not captured in the quantitative data, but rather in narratives- e.g. who, why and how cities do things the way they do, that can provide the context and causal relations of the actions to the end result that are captured by the quantitative data. Such information, both for individual cities and systematic analysis across cities, is crucial for building up transferable knowledge, enabling cross-city learning, and guiding urban policy and practice (Bai et al 2010).

Overcoming these issues is central to developing knowledge-based climate solutions that can be up-scaled individual cities to cities worldwide, while still respecting the differences between cities. A harmonized and large-scale data infrastructure is needed to pave the way.
In this paper, we extensively review the efforts of different communities to collect and make use of urban data for addressing climate challenges. We then outline a vision: a platform for data and methods available to urban decision makers worldwide. To achieve this, we suggest a focus on harmonizing data gathering, exploiting big data, making best use of remote sensing data, and synthesizing research insights on solutions.

Urbanization is a megatrend of the 21st century and has huge impact on climate change and other global environmental challenges, such as land use change. At the same time, cities are the focal points of social challenges as enshrined in Sustainable Development Goal 11 – making cities and human settlements inclusive, safe, resilient and sustainable. Researchers awake to these trends and repeatedly call for building a global urban science (Bai 2007b, Solecki et al 2013, Creutzig 2015, Solecki et al 2015, Bai et al 2016, Acuto et al 2018). However, development towards such global urban science remains stuck in well-trodden paths, as even conferences dedicated to the global dimension of urbanization continue to focus on case studies. Key barriers towards globalizing urban sciences involve data inconsistencies and inadequate model development. This paper focuses on the quantitative foundations and will help to create the envisaged Global Urban Sustainability Science.

Current state of data-based efforts at the urban-climate-change nexus
A number of disciplines and epistemic communities attempt to provide the data and understanding of urban characteristics pertinent for climate mitigation and adaptation. Here we review and organize these approaches by looking into:

- Accounting for urban greenhouse gas emissions relying on urban metabolism studies, data science approaches and assessment studies
- Making use of spatially gridded data based on remote sensing at global scale to characterize and typologize cities according to climate and other characteristics
- Utilizing Big data approaches at the scale of individual cities
- Data-based approaches for urban climate policies, based on insights from econometrics, urban economics and planning, and ex-post policy analysis.

Accounting for urban greenhouse gas emissions
Spatial boundaries of analysis are crucial and determine the quality and quantity of metrics assessed. The New York example demonstrates that differing the spatial scope changes interpretation notably: the efficiency of Manhattan is offset by high-emitting surround areas (Jones and Kammen 2014). Similarly in Europe supposedly green and compact cities, such as Freiburg and Barcelona, display low emissions in inner-city households, while the largest part of their transport emissions originate in commuters from areas far outside administrative boundaries (Creutzig et al 2012). A household consumption based analysis in Xiamen City in China shows up to 70% of the emission can be attributed to regional and national activities that support household consumptions (Lin et al 2013). Urban density and the distribution of activities also shape the urban metabolism. For example, the dense city center of Paris exports all of its waste but concentrates food consumption, whereas its surrounding areas consume high levels of construction materials and fuel (Barles 2009). A similar case can be made for infrastructures that often cross city boundaries. Airports, power plants or wastewater facilities may often be located outside of urban administrative boundaries (Hillman and Ramaswami 2010, Ramaswami et al 2008). Economic disparity within the city is a key factor that causes large disparities in household carbon footprints and underlying spatial resource flows (Lin et al 2013, Baiocchi et al 2015, Wiedenhofer et al 2017).

Essentially, hence, cities are open systems, and can be well described by investigating in- and outward flows of energy, material and emissions, which is subject of urban metabolism studies.
(Kennedy et al 2007) (Kennedy and Hoornweg 2012). Over the last two decades or so, urban metabolism studies have accumulated ample empirical and quantitative evidence to understanding cities as open system, e.g. energy and material budget and pathways; flow intensity; energy and material efficiency; rate of resource depletion, accumulation and transformation; self-sufficiency or external dependency; intra-system heterogeneity; intersystem and temporal variation; and regulating mechanism and governing capacity (Bai 2016). The importance of such a systemic assessment of all direct and indirect flows increases as the size of the unit of observation decreases (Minx 2017), i.e. smaller spatial units like cities are less self-sufficient are more dependent on trade and the GHG emissions arising outside the city boundaries in the supply chain of traded products. Urban metabolism researchers distinguish three major frameworks for assessing GHG emissions: 1) territorial or production-based accounting, 2) consumption-based footprint (CBF) and 3) trans-boundary supply chain footprint (TBIF) methodology (Ramaswami et al 2011, Chavez and Ramaswami 2011). There is no superior choice among any of these frameworks for urban analysis, rather they complement each other and provide a more complete picture for devising effective urban emission reduction strategies.

Territorial-based accounting is most common and aggregates all emissions occurring within urban administrative or functional boundaries (also denoted as scope 1 emissions) and is most straightforward to compute. Comparative studies investigate energy flows and greenhouse gas emissions in a number of cities and megacities (Kennedy et al 2009, 2015). If upstream emissions from electricity production are included, emissions are denoted as scope 2.

The CBF method accounts all GHG emissions resulting from consumption with the urban boundary (scope 3 emissions). Hence, it leaves out GHG emissions that are export-related, for example for energy use for producing goods that are consumed elsewhere. Consumption-based emissions have been best investigated in the UK context, where consumption emissions are typically higher than territorial emissions, but are also more homogenous, driven by socio-economic drivers, whereas territorial emissions vary with production facilities (Minx et al 2013). Comparing the emissions of cities in the UK, China, and the US, patterns of production vs consumption-based emissions similarly reveal that UK and US city emissions have higher consumption-based emissions (Sudmant et al 2018) (Figure 1).

The TBIF method accounts for all emissions that occur within the city boundary, plus all indirect emissions that serve the city and are relevant to its metabolism. The latter includes electricity, airline and commuter travel, water supply and other infrastructures (Ramaswami et al 2011). TBIF leaves out all life-cycle GHG emissions of non-infrastructure items, including goods and services consumption of households, but accounts for production and within the city. In a comparative TBIF study between Denver, USA, and Delhi, India (Chavez et al 2012) Denver shows much higher per-capita consumption in all sectors. Floor space emerges as a crucial unit: Household area in Denver is double that of Delhi, while per-capita floor space is seven times higher.

Most accounting efforts focus on CO₂ emissions. However, one fifth of all methane (CH₄) emissions, another potent GHG gas, have urban origin (Marcotullio et al 2013). Although, CH₄ emissions reductions are typically economically attractive, and technologically straightforward, there is little systematic accounting for urban CH₄ emissions, and important mitigation opportunities remain overlooked (Hopkins et al 2016).

Bottom-up accounts of urban GHG emissions and sector-specific drivers have been comprehensively assessed in global assessment studies such as the Global Energy Assessment (Grubler et al 2012, Kahn Ribeiro et al 2012) or the IPCC (Seto et al 2014, Lucon et al 2014, Sims et al 2014), and larger databases of urban energy and greenhouse gas emissions have been compiled. 60 – 80% of global
energy is used in cities (Grubler et al. 2012), highlighting the importance of focusing climate mitigation efforts on urban areas. A small but growing number of studies have made use of the aggregated data from (Grubler et al. 2012, Grubler and Fisk 2013) and other sources, such as the World Bank and national statistic bureaus, to investigate patterns of urban greenhouse gas emissions and energy use across hundreds of cities worldwide. A statistical analysis making use of decision tree methods investigates energy use and emission patterns 274 cities worldwide, identifying the consistent relevance of GDP per capita, population density, heating degree days and transport fuel prices as drivers of energy use, albeit differentiated across 8 statistically distinct types of cities (Creutzig et al. 2015a). Another quantitative paper aggregated data across multiple sources estimates that urban activities contribute directly 36.8 and 48.6 % of total greenhouse gas emissions, with significant differences in the urban share of greenhouse gas emissions between developed and developing countries as well as among source sectors for geographic regions (Marcotullio et al. 2013). The same study also finds that the 50 largest urban emitting areas account for 38.8 % of all urban greenhouse gas emissions (Marcotullio et al. 2013).

At a regional level, nationally-consistent data sets can be used to assess greenhouse gas emissions across metropolitan areas and over time, as demonstrated for the US (Markolf et al. 2017), and across all administrative units in England (Baiocchi et al. 2015) (see figure 2 for the London case). In other cases, such as India, spatialized large-scale household surveys enable a nationally consistent emission footprinting of household energy use across all locations (Ahmad et al. 2015).

The construction of urban infrastructure is also creating significant emissions in its own right, and the way future cities are built now determines their energy demand and emissions tomorrow, given the strong path dependence and lock-in potential of urban form (Creutzig et al. 2016). A high share of urban emissions from buildings and transport will originate from both the construction and the usage of new cities, particularly in Asia, the Middle East, and Africa (Creutzig et al. 2016). It is hence of high importance to properly understand the energy and emission implications of urbanization in these world regions, explored in detail in the data spotlight on Asian cities (Box 1). Prospective studies of urban greenhouse gas emissions with a full infrastructure and life-cycle perspective however remain scarce and require further investigation.

Figure 1. Production vs consumption based emissions in UK, US, and Chinese cities. Chinese cities have typically higher production-based footprints, corresponding to their industrial export-oriented activities, whereas US and UK cities have
higher consumption-based footprints, corresponding to imports of goods. Exceptions are Beijing and Shanghai, which also have a disproportionately high consumption footprint, and Houston dominated by oil refineries and hence production-based emissions. Source: (Sudmant et al 2018). See (Minx et al 2013) for similar results displaying full diversity of footprints of English districts.

Figure 2. London administrative units are characterized by patterns of residential emission drivers. Each node corresponds to a statistically distinct combination of housing, climate, urban form and socio-economic characteristics explaining location-specific residential GHG emissions. Source: (Baiocchi et al 2015)

Box 1. Spotlight on Asian cities.
As the quantitative scale of urbanization is nowhere as prominent as in Asia in the first decades of the 21st century, a focus on Asian cities is warranted. In fact, Asian cities are growing at a historically unprecedented scale and speed. Unfortunately, the scale of the phenomenon is not adequately matched with data. City scale data for Asian cities, especially for South and South East Asian remain precarious. Even cities seen as active contributors in climate change solutions have only unreliable and basic data on GHG emission inventories and underlying social-economic drivers. In CDP’s global data bank of 187 city data only 18 are from Asia, mostly from Japan, South Korea and Taiwan with the exception of two from the Philippines and two from Indonesia. However, existing city emissions and database systems in as Japan, South Korea, Taiwan, and Hong Kong provide high detail-resolution, but remain mostly undisclosed, as data are coded in local languages and few efforts have been made to bring them to English except for big cities like Tokyo, Kyoto, and Hiroshima. However, 50 mid-size Japanese cities’ emissions data have been reported and analyzed (Makido et al 2012).
In China, the epicenter of current urbanization, there has been considerable increase in city data and studies related to city emissions. A recent study showed existence of 177 studies, 80 of them in English and 97 related to cities and emissions (Chen et al 2017). 122 (or about 45%) of 283 prefecture-level cities have some emission estimates with various level of details and methodologies-used, mostly based on city energy consumption data only (Figure 3). Other studies have accounted for emissions in 18 (Xu et al 2018) and 24 Chinese cities (Shan et al 2017a), developed consumption-based emission accounts for 13 Chinese cities (Mi et al 2016), and have specifically highlighted emission accounts for Tibet (Shan et al 2017b).
Figure 3. Per capita emissions and GDP in 122 Chinese cities between 2009 and 2012. Total city emissions vary 1000 fold (from 0.23-305 million tCO\textsubscript{2} per year). Source: (Chen et al 2017). The time gap between data and publication is significant: the average gap between the year of the city carbon emissions data and the year of publication is reported as 3.9 years.

In India, urbanization is slower than in China, but urban areas are likely to hold 40% of India’s population and could contribute 75% of India’s GDP by 2030 (compared to 31% and 63% respectively in 2010). Only a few emissions data were reported. Selected cities where emission are estimated include Ahmadabad (Shukla et al 2009) and Delhi (Chavez et al 2012), followed by a comparative study of six major Indian cities among which Chennai has the highest emissions, at 4.79 tCO\textsubscript{2}e per capita (Ramachandra et al 2015). Another recent study uses survey data to estimate urban household based GHG emissions for the largest 60 municipalities finding efficiency gains in larger cities (Ahmad et al 2015).

Urban data scarcity is also pervasive in other South Asian countries. The Japanese Low Carbon Society supported emissions estimates in a number of smaller cities in Malaysia and Vietnam (National Institute for Environmental Studies 2018), and a few other studies on specific cities, such as Kathmandu, have been reported (Shrestha and Rajbhandari 2010). The key data issues for Asian cities are similar to other world regions, and include a lack of reliable key socio-economic and activity data at city scale; data boundary mismatches; inconsistent methods application; and a lack of realistic assessments of climate solutions and potentials. Consistent evaluation, e.g., under the Global Protocol for Communities (GPC) and support for database development and analysis are crucial to upscale urban data-based climate solutions.

Remote sensing in the assessment of cities
Measuring CO\textsubscript{2} concentrations with very high spatial and temporal resolutions, and hence enabling emissions source allocation, would enable unforeseen opportunities in urban GHG emission accounting, and the corresponding identification of climate solutions. First steps towards such measurements have already been taken. Already in the mid-2000s, remote sensing of huge emitters like power plants was technologically feasible (Bovensmann et al 2010), and is becoming increasingly possible for lower emissions and at higher resolutions with instruments like NASA’s Orbiting Carbon Observatory ( OCE-2 and OCO-3) (Eldering et al 2017) or ESA’s Sentinel-5 for Methane emissions (Gurney 2015). Los Angeles emerges as testbed for these advanced monitoring technologies and is following ambitious efforts to provide monitoring and controlling of its carbon emissions (Newman et al 2016). Urban CO\textsubscript{2} emissions accounting based on remote sensing will be of increasing relevance in future climate-policies, and has the potential to alleviate data scarcity in global south cities. However, so far, remote sensing can best serve combined with other data.

Remote sensing data offers extensive spatial coverage and (potentially) temporal consistency. Simple satellite imagery in the visual spectral lengths – \textit{photography} – allows the inference of structural characteristics such as street networks and the historical development of the built-up areas (Yuan et al 2005, Xiao et al 2006). Multispectral analysis, aided by established indicators such as the
normalized density vegetation index (NDVI) (Tucker et al 2005, Gallo et al 1993) enables the
determination of material properties or vegetation assessment (Ridd 1995), urban land-use structure
(Herold et al 2002), temperature measurements (Imhoff et al 2010, Tran et al 2006) and pollution
(Gupta et al 2006). A global analysis of the vertical shapes of cities has revealed how cities in
different world regions grow differently: from 1999 – 2009, growth in Tokyo or New York has been
into the vertical dimension, Indian cities have mostly expanded horizontally, while big Chinese cities
have grown in both dimensions (Frolking et al 2013) (Figure 4). Combining satellite data with other
datasets and analysing it via state-of-the-art machine learning even allows the estimation of poverty
levels from satellite data (Jean et al 2016) (see also Box 2).

Figure 4: Changes from 1999 - 2000 in backscatter power (PR) and nightlight intensity (NL) for 12 global cities. Large
differences in growth characteristics are revealed. From Frolking et al (2013).

Remote Sensing is a powerful tool, but only if combined with other sources (Seto and Christensen
2013). The combination with weather data is an obvious candidate for a global analysis of urban
areas and climate change. While weather models or climate reanalyses are able to provide live
assessments and temporally highly resolved meteorological parameters, remote sensing and global
climatological climate data sets can inform the long-term climatological characteristics of a city,
particularly interesting are temporal and spatial high-resolution microclimate data, including land
surface air temperature and humidity, available for Europe (Haylock et al 2008) and worldwide
(Kearney et al 2014), and urban heat island climatologies (CIESIN 2016).

Contextualizing remote sensing data with spatialized socio-economic data emerges as an increasingly
relevant area of study. In 2010, 4231 cities had a population of more than 100,000 (Atlas of Urban
Expansion 2018). Remote Sensing offers the opportunity to assess these cities in a consistent manner
and analyse the impacts that these settlements have on land-use, greenhouse gases, and how they
will be impacted by climate change. Night-time imagery has been demonstrated to be a useful proxy
for urban extent and economic affluence (Doll et al 2000), and can be used to estimate spatialized
population density data (Bagan and Yamagata 2015). Visual observations of urban area can be
combined with forecasts of economic growth to create spatially-explicit projections of future urban

Remote Sensing information is widely used for climate-change related risk assessment and disaster risk reduction. In particular, understanding flood risks requires a combination of spatially resolved data on physical flood exposure – containing data on elevation, hydrology and built-up area – as well as socio-economic data highlighting economic vulnerabilities to floods. Such a framework has been presented and applied to case studies, e.g. in Copenhagen (Ranger et al 2011, Hallegatte et al 2011) and in an overview of future flood risks in 136 major coastal cities (Hallegatte et al 2013).

Downscaled models of climate impacts are crucial to map urban adaptation challenges worldwide. Scheurer et al. (2017) propose using the Theil-San estimate and Euclidian distance as a measure of magnitude of climate change, both in temperature and humidity, including long-term average changes as well as weather extremes. This method can be applied for any city and hence enables global comparisons and rankings of climate change impacts (Figure 5). A US-specific study combines Landsat and MODIS data in a land model to assess the impact of urbanization on US surface climate, finding relevant warming and increased surface run-off in built-up areas, but with varying patterns across cities (Bounoua et al 2015).

Remote sensing can be used to help determine the urban heat island effect and properties of urban climates, enabling the application of a consistent methodology across many cities (Peng et al 2012). Specifically, quantities from satellite remote sensing, such as the normalized difference vegetation index NDVI (Grover and Singh 2015), can partially predict surface UHI (Zhou et al 2017, Peng et al 2012). Remote sensing can reveal the growth of the urban heat island over time with growth of the

Figure 5. Mapping direction of change in temperature and precipitation in cities. The direction of change is indicated by colour, where I equals warmer and wetter; II colder and wetter; III colder and drier; and IV warmer and drier conditions.

Source: (Scheuer et al 2017)
urban area extent (Streutker 2003). Urban climates will most significantly be impacted by more severe heat waves (Ganguly et al 2009) (Li and Bou-Zeid 2013), which can be explicitly scaled down to specific urban locations in climate models. This is of particular importance as heat waves lead to deadly outcomes (Mora et al 2017, Nissan et al 2017), with currently 30% of the global population living in areas experiencing deadly heat (Mora et al 2017). Such spatially-explicit data can help to assess and ease communications of climate impact and enable the identification of cities with similar adaptation challenges (Creutzig 2015).

High-resolution maps containing population, settlements and urban footprints form the basis for an integrated assessment of global settlement patterns. Rapid Advances have been made in the past decade. Both new satellite technology – such as the Tandem-X radar satellites – and improved data processing via machine learning have facilitated rapid advances in accuracy and resolution of such maps. Until recently, the MODIS 500 urban land cover (Schneider et al 2010) represented the state of the art in urban land cover data sets (Potere et al 2009). It is now outperformed by both the Global Urban Footprint (GUF) data which provides higher resolution and accuracy than any other urban land cover data set (Esch et al 2017), even including the high quality Global Human Settlement Layer (GHSL) (Pesaresi et al 2013, 2016) or GlobeLand 30 (Chen et al 2015). The GUF features a binary urban footprint at a resolution as high as 0.4” (approximately 12 m) at the equator and 0.6” in the mid-latitudes on a global coverage, freely available for scientific use. This high resolution constitutes a paradigm shift in studying urban extent in cities worldwide.

**Box 2. Machine learning approaches to process remote sensing data**

Machine learning techniques, such as neural networks, are powerful tools for the analysis of big, multi-dimensional and often complex “Big Data”, where complexity needs to be reduced to understand its main drivers (Hinton and Salakhutdinov 2006). Convolutional neural networks (CNN) serve well to classify images (Krizhevsky et al 2012), and are increasingly used to assess land use patterns (Castelluccio et al 2015). Some research takes this approach further and combines it with the analysis of socio-economic data (Tapiador et al 2011, Jean et al 2016). Jean et al. (2016) are particularly instructive. They predict poverty in five different African countries - Nigeria, Tanzania, Uganda, Malawi and Rwanda – at a ward/village scale using a combination of CNN, daytime satellite imagery and nightlight data. They use a three-step approach. First the CNN is trained on ImageNet (Deng et al 2009) to learn how to distinguish visual properties like edges and corners. In the second step it is fine-tuned so that it is able to predict night-time intensities from daytime images. Nightlights are a globally consistent predictor of poverty, thus the model is trained to focus on the aspects in daytime imagery that are relevant to poverty estimation. In the third and final step, socioeconomic survey data is added to the analysis. It is used to train ridge-regression models on both household surveys and the image features from steps 1 and 2. Their approach exploits night-time data as a globally consistent, but highly noisy proxy for poverty in an intermediate step and ultimately explains 37% to 55% of average household consumption, and 55% to 75% of the variation in average household asset wealth. While using publicly available data, it provides better results than mobile-phone based studies and far outperforms products that rely solely on nightlights. Another recent study uses data taken from Google Street View images and machine learning techniques (feature extraction and v-support regression) to successfully estimate high income areas in U.S. cities, using (Glaeser et al 2018). The usage of phone records can reveal detailed mobility patterns for improving both the understanding of travel behaviour and traffic management (Toole et al 2015).
Big Data approaches at city-scale

New forms of data, including “Big Data,” have the potential to generate new hypotheses and develop new methods for understanding interactions between social, biophysical, and infrastructure domains of complex urban systems under climate change (Herold et al 2002, McPhearson et al 2016). Crowd-sourced and big data, such as the movement of people tracked by cell phones, offer manifold new possibilities for assessing the inner workings of a city. Crowd-sourced information can serve as a reliable proxy, with vastly improved resolution and replication, for more traditional empirical social survey methods (Wood et al 2013). This and similar social media analyses (Keeler et al 2015) offer the opportunity to complement existing traditional approaches to collecting information on especially human behavior in cities, that can be brought together with other sources of biophysical and infrastructural data, especially in spatial formats through GIS. Big data can also emerge from municipal hotlines, city planning offices, utility use and repair records, tax assessor databases, and the rapid emergence of sensors and instrumented buildings, roads, and even ecological spaces. The utility of big data for understanding urban systems including climate impacts and efficacy of climate solutions will only increase with time (Knox 2014).

Urban decision-makers need improved and regularly updated information on human behavior and perceptions and how they relate to global and local climate change. Linking human behavior in cities to downscaled climate projections and remotely sensed observations of urban form, land use patterns, land cover, and social-demographic information from national and international databases has the potential to drive a much more nuanced, and high spatial resolution platform for improved decision-making. Over the past decade, with the advance of Big Data and the digital social sciences, as well as the growing use of social media data (SMD) in geography studies, a host of new opportunities to augment and expand urban systems and climate impacts research have emerged.

Geocoded social media data (SMD) from social media users (e.g., Flickr, Twitter, Foursquare, Facebook, Instagram) may offer one of the most important new opportunities to fill data gaps and tackle many of the barriers that prevent researchers and practitioners from understanding the human behavior component of urban system dynamics and climate change. SMD and other “Big Data” enable researchers to ask a wide range of spatially explicit questions at an unprecedented scale. Most social media provide the possibility of either to manually select the location from where one posts a message, or have it automatically added through geolocation tracking services. Even though, at present, geolocated tweets and Flickr photographs represent a tiny fraction of the overall volume of SMD (e.g., tweets geocoded via GPS constitute only 1% of all tweets) (Crampton et al 2013) the sheer quantity of these data makes them worth examining. Geotagged tweets can augment traditional control data (e.g., remotely sensed images, roads, parcels). For example, geolocated Twitter messages can serve as control data for modeling population distribution (Lin and Cromley 2015).

Research using geolocated SMD to study socioeconomic disparities and its relationship to climate impacts in cities is also starting to take shape. Crowd-sourced data from Foursquare users in London, for instance, has been shown to be a reliable proxy for the localization of income variability and highlighting where more at risk neighborhoods are across the city (Quercia and Saez 2014). Yet, mapping based on demographically unrepresentative data can also reproduce spatial segregation and provide an unfair picture of the places that matter citywide (Cranshaw et al 2012). This holds true for global-scale analyses as well. The volume of geocoded tweets greatly differs across nations worldwide. The U.S. and Brazil are some of the countries where the ratio between geocoded and non-geocoded tweets is the highest, while countries like Denmark and Norway register significantly lower values (Graham et al 2013). The emergence of multiple forms of big data creates exciting
alternatives to assess how people use and respond to urban events, plans, policies, and designs for climate change adaptation and mitigation. New forms of data may be a crucial resource in examining the use, value, and social equity in access of particular spaces in the city that provide refuge during climate driven extreme events, such as the parks, vacant areas, and nonpark open spaces that can provide, for example, cooling during heat waves. Working with big data offers opportunities with multiyear to decadal data sets to understand human–nature interactions in the city as never before and could prove crucial to assessing progress on examining impacts of climate change and mitigation options in cities.

Various sources of big data have already proving useful for informing disaster risk management and climate adaptation planning. Kusumo et al (2017) used volunteered geographic information through SMD as source for planning flood evacuation shelters, while Brouwer et al (2017) used SMD sourced observations of flooding to develop a method for estimating flood extent in Jakarta, Indonesia. In New York City following the devastating impact of Hurricane Sandy researchers used Twitter SMD to reveal the geographies of a range of social processes and practices that occurred immediately after the event (Shelton et al. 2014). Stefanidis et al (2013) used Twitter data collected during the devastating Sendai (Tohoku) earthquake in Japan (3/11/11) to examine social networks and build a database for studying the human landscape of post-disaster impacts. Understanding interactions between climate change and fire prone landscapes is another important area of concern for both climate change adaptation and disaster risk reduction. Kent et al. (2013) were able to use SMD (included Twitter, Flickr, Picassa, and Instagram data) to examine spatial patterns of situational awareness during the Horsethief Canyon Fire (2012) in Wyoming, USA and demonstrated that the utility of SMD for actionable content during a crisis.

From big data application for specific analysis, big data can become a central tool for urban risk monitoring and climate policies, enabled by sensor-based cities and the vast amounts of data routinely generated by their inhabitants via social media (Box 3 illustrates the Newcastle case). Applications include:

- Using real-time data feeds from local weather stations, rainfall gauges and sewer gauging to assimilate real-time data within hydrodynamic models for improved flood prediction;
- Combining local high frequency observations, with regional/national monitoring and predictions, along with tracking of geospatial social messaging (e.g., tweets of incidents as they occur) to provide improved early warning of potential impacts;
- Employing image processed CCTV feeds to understand hazards e.g. of surface water locations and social media feeds to validate in real-time the emergent patterns of flooding;
- Integrating spatially heterogeneous sensor data feeds on flows and movements (e.g., traffic) with geospatial social messaging, CCTV and other data for improved understanding of the temporal dynamics of impacts;
- Coupling CCTV monitoring with social media data feeds to understand better citizen reaction and response to emerging impacts for improved future hazard mitigation; and,
- Using knowledge from previous events, including modelling result-sets of both hazards and impacts, to improve ‘ahead of event’ response from the site to the city-scale for future ‘events’.

However, in order to realise the full potential of the ‘big data’ sensor networks and social media feeds that are continuously capturing the dynamics of cities in real-time, new approaches are required. These must rigorously evaluate and integrate real-time data and information from traditional and new ‘big data’ sources. The sheer volume of information requires advanced analytics
and methods for uncertainty handling to process and assimilate this data for real-time decision-making and long term planning.

Box 3. The Newcastle Urban Observatory

The Newcastle Urban Observatory records 1 million city observations a day, of over fifty social, environmental and technical processes, across the city. These include transport emissions, precipitation, surface water and river flows, air and water quality, biodiversity health indices such as beehive weight. The data are openly available through an API and can inform adaptation and mitigation activities in sectors such as transport, building energy, urban greening and flood management. The City Council are already using this high resolution dataset to inform environment and transport strategies.

Data-based approaches for urban climate policies

The greatest challenges in achieving sustainable cities seem to mould insights from studies and assessments into actual policies and urban planning practices. Three data-based approaches will help to directly address appropriate climate policies and their mitigation and adaptation potential. The first approach makes use of GHG accounting data and ancillary variables to identify drivers of emissions and climate risks, and infer policy levers. The second approach combines urban economic and urban planning insights to properly explain observed data and predict consequences of policy implementations. The third approach performs systematic policy reviews of urban municipalities, providing an ex-post analysis of climate actions. All three approaches are required to upscale urban climate solutions beyond individual cities.

Driver analysis of urban climate data. Data of urban GHG emissions and energy use, combined with ancillary socio-economic, geographical, and urban form parameters, allow the identifications of policy levers stratified across types of cities. Globally, urban typologies demonstrate how climate mitigation efforts require different policies for different city types and how advanced statistical analysis of large data sets (decision tree methods) reveal the uniting and dividing drivers of urban energy use across different cities (Creutzig et al 2015a). Results demonstrate that transport fuel taxes are a main lever to not only influence urban transport emissions, but also urban GHG emission en large (Creutzig et al 2015a). The same method has also been successful applied for countries like the UK, enabling a more refined analysis of drivers, such as heating systems and building vintages (Baiocchi et al 2015). Intertemporally consistent household surveys similarly enable the spatially complete analysis of household energy demand and GHG emissions, stratified along individual social
and economic characteristics, as shown for India (Ahmad et al 2015). Arguably, similar typologies would also be suitable for identifying climate adaptation strategies (Creutzig 2015).

The advantage of this approach is a globally consistent method identifying options stratified across city-specific characteristics. This inter alia allows mayors to learn from best-in-class cities of the same type. The disadvantage is that this method neither provides city-specific resolution on implementation, nor does it identify barriers and obstacles in the political economy.

**Urban economics and urban planning.** Theory and conceptual insights are highly relevant for urban policy decisions, especially when empirically founded. A highly impactful urban planning empirical analysis pointed to the influential role of urban density for transport energy demand (Newman and Kenworthy 1989). However, population density was shown to be a proxy variable for more specific urban form parameters, such as street connectivity and commuting distance (Mindali et al 2004). More general meta-analysis highlights the relevance of the 5Ds (density, diversity, design, destination accessibility, and distance to transit) as key variables of urban planning for reducing GHG emissions (Ewing and Cervero 2010, Stevens 2017). And population density reveals to be useful proxy nonetheless: thresholds for energy-efficient public transportation of around 50-150 persons/ha point to sustainable urban development windows, enabling low-carbon mobility (Lohrey and Creutzig 2016) and low-carbon residential energy use (Baiocchi et al 2015).

Urban economic theory explains the core principle: land prices, land availability and transportation cost largely determine the density of a city, with lower costs for transport driving suburbanization and longer commuting distances, hence leading to a higher transport energy demand. At the intersection between urban planning and computational urban economics, detailed land-use and transport models (e.g. agent-based models) can capture system-wide emission effects of planning decisions over time (Echenique et al 2012). Such models are extremely promising but depend on rich data input from municipalities not everywhere available, and are not easily tractable in terms of identifying key drivers. Urban econometric models aim to identify key drivers of emissions that have complex repercussions (Creutzig 2014, Borck and Brueckner 2016). This theory supports the strong and non-linear effect of fuel prices (Creutzig 2014, see also Figure 6). Urban econometric analysis can reveal unexpected effects. In the U.S., urban planning regulations correlate lead to counterfactually higher emissions from non-regulated cities, as new city construction takes place in low density development and hot climates that demand energy for cooling (Glaeser and Kahn 2010).
Figure 6. An urban economic model that explains the relationship between the marginal costs of car transport and public transit, and their interaction via urban form. With higher fuel prices and marginal transport costs for cars (coded as $m_c$), transport distance $r$ is reduced, more competition for land reduces average plot size $s$, urban sprawl is limited and urban form becomes more compact. This in turn makes public transit financially more viable, as ridership is improved, per unit of infrastructure investments $C$ total passenger km traveled $D$ increase, thus making the compact inner city even more attractive (marginal user fees of public transit to refinance infrastructure $m_p$ are decreasing). With lower fuel prices, the dynamics are inversed, urban sprawl is incentivized and public transit becomes financially unattractive. Source: (Creutzig 2014).

Ex-post analysis of urban policies. Triggered by the slow progress in international climate policy over the last two decades, cities and local governments increasingly cooperate to combat climate change from the bottom up: thousands have developed and implemented local climate action plans. Yet, little is known about the impact those measures had on reducing emissions (Seto et al., 2014; Minx, 2017). This lack of knowledge is currently a direct barrier to learning about local climate solutions. Hence, developing a body of literature that aims to understand what solutions work for whom under what conditions and why would meet an important information demand by policymakers and practitioners. Yet, conducting such analysis at the local scale is complex and littered by data challenges.

Qualitative data is important for improving the understanding of adaptation and mitigation in cities. It is typically descriptive and often concerned with understanding behaviour, institutions, and context. These have a strong impact upon climate impacts, vulnerability, and the effectiveness of adaptation and mitigation options (Bulkeley and Betsill 2005). Typically, qualitative data provide a depth of information not possible with the big data approaches described previously, although is usually more time consuming to collect and process.

Analysis of urban adaptation and mitigation strategies in EU cities (Reckien et al 2014) and specifically in the UK (Heidrich et al 2013) highlighted that cities often use different baseline years, include different emissions in their accounts, set reduction targets differently, cover different sectors, and describe policies or approaches to influence demand in quite different ways. Analysis of 627 climate actions in 100 cities revealed the heterogeneous mix of actors, settings, governance arrangements and technologies involved in the governance of climate change in cities in different parts of the world (Broto and Bulkeley 2013). Along the same line, C40 highlighted a diversity of
governance structures in just 66 cities, which influence the levers available to city actors or other non-city actors to implement change (C40 2015).

This heterogeneity and inconsistency in data makes it extremely difficult to answer questions such as, what effect will the combined mitigation plans of the world’s cities have on global greenhouse gas emissions? How well protected are city inhabitants against a flood of any given return period? What actions have been taken, and how effective have they been, to enhance adaptive capacity and reduce risks? Which combination of city and non-city organisations need to be involved to deliver a particular climate action in a given city?

There are a wide range of approaches to reporting governance structure, policies, regulation, culture, institutions, adaptation and mitigation actions, and other factors relevant to climate action. The intra- and inter-urban variability of the quality, reliability and completeness of information is huge. The current lack of standardisation of records, quality assurance, and incomplete availability, poses a significant challenge for understanding climate risks, diagnosing greenhouse gas emissions, and assessing the effectiveness of climate action.

**Building the data foundation for a Global Urban Sustainability Science**

Urban data science has been compartmentalized for too long, even as data and observation is identified as one of the six research priorities in cities and climate change (Bai et al 2018) All of geography, geomatics, political science, urban planning, industrial ecology, urban economics, and most recently computer science (with machine learning approaches on big urban data) are tackling various aspects of building and transport energy demand with data analysis and modelling. Industrial ecology provides frameworks for analysis that take into account indicators across disciplines and scales (Bai 2007a). Geomatics provide data on urban extent and similar variables that are slowly taken up for global urbanization research (Seto et al 2012, Bren d’Amour et al 2017, Güneralp et al 2017a). But urban research remains disparate, marginalized and ill-prepared to interact effectively with global policy (McPhearson et al 2016), leading inter alia the inaugural issue of Natural Sustainability to call for a Global Urban Science (Acuto et al 2018). But vocabulary is however inconsistent across disciplines, and cross-citation, and more importantly, cross-fertilization remain at low levels.

Here we attempted to bring the insight and approaches from different urban data sciences together, demonstrating that the interdisciplinary research conveys the promise of extraordinary synergies, forming the quantitative foundation of Global Urban Sustainability Science (Figure 7).
Interdisciplinary efforts bringing data approaches from various epistemic communities together will provide the quantitative foundations for a globally applicable and consistent urban sustainability science. A multitude of disciplinary efforts tackles empirical and theoretical foundations relevant to quantify climate change issues in cities. Cross-disciplinary research becomes increasingly common, but only a strong joint effort will provide the quantitative foundations of GUSS. Some disciplines, such as geomatics and remote sensing studies provide globally consistent data, but not necessarily in metrics directly applicable to climate change issues. Other disciplines, such as industrial ecology, provide relevant metrics, but data gathering is insufficiently harmonized. Even other disciplines, such as theoretical economics developed theoretical explanations of how quantitative variables interact dynamically. Theoretical efforts can help the transformation of empirical variables into climate change relevant metrics; and in turn, empirical data can gauge the models and support the identification of relevant dynamics.

We envision a platform for comprehensive and harmonized urban data and methods that is available to urban decision makers worldwide. This platform should have low transaction costs in knowledge exchange. Incentives should guarantee high level of contributions by scientists and decision-makers alike. We suggest three specific goals:

1) Mainstreaming and harmonizing data collection in cities worldwide;
2) Exploiting big data to upscale solutions on agent and/or building scale while maintaining privacy; using high-resolution remote sensing data, and social media data to automatize computation of first-order climate effects and solutions;
3) Performing synthetic research on city case studies to aggregate qualitative information on cities and urban solution strategies.
Mainstreaming and harmonizing urban data
Several attempts have been made so far to collect and present comparable data across cities to foster climate solutions. We shortly discuss a few relevant ones, and identify their strengths and weaknesses.

**Metabolism of cities (metabolismofcities.org).** Recent efforts towards realizing unified indicators for the study of urban metabolism have led to the creation of the “Metabolism of Cities”, a collaborative platform and dataset that aims to connect researchers and practitioners (Metabolism of Cities 2018). The *Metabolism of Cities* project provides a data portal to assess the metabolism of cities, including mainly electricity consumption, waste numbers and GHG emissions. The current data set contains 148 cities, with 465 total indicators and 8973 data points. Many of these data stem from Hoornweg et al. (2011). Vienna and Brussels have an exceptional abundance of data points (above 1000), with 36 further cities featuring more than 10 data points each.

**CDP (cdp.net/en/cities).** The CDP Open Data Portal collects self-reported urban greenhouse gas emission data, and currently presents a global dataset based on city self-reported CO₂ emissions for the period 1990–2016 for 187 cities in five geographic regions for 2016 and 229 cities for 2017 (153 overlap between the two years) (CDP 2017a). CDP reports units of CO₂ equivalents correspond to different greenhouse gases with different warming potentials for a standard 100-year time horizon. CDP issues consistent guidelines for reporting(CDP 2017b). Of reported emissions, 89% are from the period 2010–2015, including Scope 1 (direct fossil CO₂ emissions from residential and industrial heating, transport, industrial sectors), Scope 2 (indirect CO₂ emissions from the consumption of electricity and steam generated upstream from the city), and Total Emissions, nominally equal to the sum of Scope 1 and Scope 2. Cities also reported the change in emissions between the most recent and the preceding reporting periods, explanations for this change, methodology details, and the gases included in the total emissions in CO₂ equivalents. A detailed analysis of this data is in preparation (Nangini et al under review).

**UCCRN (uccrn.org/resources/case-study).** The urban climate change research network (UCCRN) built up a case study database of more than 100 cities worldwide, covering topics such as vulnerability, hazards and impacts, mitigation and adaptation actions, and sector-specific themes, such as waste water and flood management.

**CEADS (ceads.net/).** China Emission Accounts & Datasets (CEADS) provides up-to-date energy, emission and socioeconomic accounting inventories for China. Datasets published by CEADs are the results of research projects and data is free to download for academic usages. While most data are national or province level, recent contributions add detail on China city-level emissions (e.g., Shan et al 2017a, see also Box 1).

**The Atlas of Urban Expansion (http://atlasofurbanexpansion.org/).** The Atlas of Urban Expansion (Atlas of Urban Expansion 2018) is a joint project by UNHABITAT, New York University (NYU) and the Lincoln Institute. It is the output of the first two phases of the Monitoring Global Expansion Program, an initiative that gathers data and evidence on cities worldwide, analysing quantitative and qualitative data on urban expansion in a stratified sample of 200 global cities. Their data is freely accessible on the website above, and an impressive collection of high-quality data that showcases urban growth dynamics and developments based on remote sensing and is one of the most comprehensive data bases of urban spatial data. Among others, the Atlas provides indicators representing the actual urban extent of a city, rather than describing its political or administrative boundaries. Data that applies to the actual urban extent overcomes one of the biggest hurdles in the analysis of urban spaces, but requires a detailed analysis of each individual city. For 30 cities, the
historical growth has been mapped and visualized by use of both historical maps and satellite observations.

All these efforts are extremely valuable and enrich our understanding of cities and climate change. The advantage of the CDP, CEADS, and Metabolism of Cities projects are that they directly aim to gather GHG emission data. Their disadvantage is that data are obtained from diverse sources, and inconsistencies in methodologies remain unavoidable, often rendering interpretation difficult. In contrast, the Atlas of Urban Expansion gathers consistent satellite and map-based data. However, their data products are only indirectly relevant for climate change related issues.

Data gathering, building on initiatives from CDP, Metabolism of Cities, and the Atlas of Urban Expansion, should further intensify, and based on common protocols and standards, such as the Global Protocol for Communities. This requires data verification and gauging, as done for the CDP data (Nangini et al under review).

Building a data platform for global urban data requires combining existing data sources. A standardised tagging system for data sources can help to unify data across sources. One way to build a global data platform is hence to link existing data via a unified, standardized tagging design, rather than doubling efforts by duplicating data sets. Machine learning of meta-information, as outlined above, may be an efficient way to proceed in such a direction.

It also requires financial efforts and leadership, and willingness from municipal employees worldwide to invest time and resources into improving urban data availability. International aid should support urban agencies from cities in Least Developing Countries, especially in smaller (<1million) municipalities and where informal settlements are abundant to overcome current bias in research (even as it remains important to properly estimate this bias). Importantly, locally generated knowledge about climate issues is not only relevant for data gathering efforts, but at least as relevant as precondition for locally motivated action. Even as remote sensing and other big data can help to generate consistent and harmonized data for all cities worldwide, local data competencies are crucial for urban planning and policy design (Brković-Bajić 2008).
**Table 1. Key urban climate issues and their state of data.**

<table>
<thead>
<tr>
<th>Urban climate issue</th>
<th>State-of-the art knowledge / insights and key papers</th>
<th>Relevant remote sensing data</th>
<th>Relevant city-specific big data</th>
<th>Degrees of data coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Floods and storm surge</td>
<td>Detailed understanding of individual flood maps, e.g. (Winsemius et al. 2013) A general framework for assessing flood risk (Hallegatte et al. 2013) Numerous case studies, e.g. (Ranger et al. 2011, Hallegatte et al. 2011) Understanding the impact of urbanization and air quality in increased heavy rain (Shi et al. 2017, Fan et al. 2018)</td>
<td>Flood maps are aggregated products of elevation and hydrology models, including: GLoFAS (Winsemius et al. 2013), JBA flood risk maps (JBA 2018), and Digital Elevation Models such as SRTM 1 arc-second topography (Farr et al. 2007)</td>
<td>Many cities employ individual flood maps that are locally constrained. Global models and data are mostly commercial (e.g. JBA flood risk model).</td>
<td>high</td>
</tr>
<tr>
<td>Landslides</td>
<td>Global landslide risks (Dilley 2005, Glade et al. 2006, Petley et al. 2005)</td>
<td>Digital Elevation Models such as SRTM 1 arc-second topography (Farr et al. 2007) and Global Risk Data Platform (UNEP / UNISDR)</td>
<td>Past disaster risk data, e.g. DesInventar or Munich Re NatCatServices</td>
<td>medium</td>
</tr>
<tr>
<td>Consumption-based emissions</td>
<td>Urban lifestyles have different emission patterns from rural ones (Minx 2017), particularly in aviation-use (Heinonen et al. 2011, Ottelin et al. 2014) Large disparities in household consumption based emission (Lin et al. 2013, Wiedenhofer et al. 2017)</td>
<td>Hybridization of multi-region input-out tables (MRIO) with spatially gridded population data and GDP data allows crude gridded estimates of consumption-based footprints (Moran et al. 2017)</td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Urban equity</td>
<td>Climate change will impact very differently on various population groups (Reckien et al. 2017)</td>
<td>Visual satellite imagery, particularly from AVHRR instruments. Nightlight data (NOAA 2018) Google Street View, Mapillary, Determine poverty from visual and nighttime imagery (Jean et al. 2016)</td>
<td>UN-Habitat equity indicators Street level imagery (Glaeser et al. 2018)</td>
<td>Low</td>
</tr>
</tbody>
</table>
Exploiting big data

Big data is already used widely, but neither remote sensing data nor social media and other geolocalized data are fully utilized to their full potential. Remote sensing data remain insufficiently integrated with other heterogeneous data sources, such as OpenStreetMaps, and SMD remain often constraint to specific application and cities. Table 1 lists the availability of data and approaches for different urban climate issues. It reveals that a high amount of data is already available, and full integration is likely to lead to relevant synergies. Several approaches promise high potential.

A first approach involves systematic city-wide data gathering for the consistent application for multiple climate and sustainability related purposes, as tentatively attempted in Newcastle (Box 3). Making big data part of urban policies will produce synergies across reaching different policy goals.

A second approach involve the more sophisticated hybridization of data sources and collaboration across disciplines. The need for interoperability between SMD-based research and the work of many public and private organizations by means of intelligible units of analysis (Housley et al. 2014) and interdisciplinary approaches (Young 2014) are common hurdles to big data scholarship. To pool the necessary competencies for a strong SMD research agenda and its execution it would be key that researchers from different walks of science – such as computer scientists, computational social scientists, linguists, geographers, and urban climate modelers, and urban ecologists, among others – join forces and work side by side as part of the same Big Data research collective (Stefanidis et al. 2013). To overcome the relatively slow-pace of urban systems research for addressing pressing, and in many cases, urgent problems facing cities and urbanized regions (Harris 2012, McPhearson et al. 2016), linking more traditional remotely sensed and local data with new emerging, crowd-sourced data is necessary. This data linkage also requires creating new interdisciplinary institutional bridges for scientists in different disciplines to work across a range of big data sets applied to advancing climate change adaptation and urban sustainability. Promising examples include the training of high-resolution gridded and/or remote sensing data for predicting building or transport energy use in the city of Porto (Silva et al. 2017, 2018). Approaches can also be applied to the transport sector. Street networks can be classified with simple hierarchical clustering, producing distinct street block fingerprints, which can in turn be used to produce a typology of cities (Louf and Barthelemy 2014).

A third approach includes the upscaling of a specific method from one specific city to national areas and beyond. The example of synthetic building stocks to investigate high-resolution strategies for energy conversation in buildings is summarized in Box 4. In each case, however, privacy is a concern, and strategies that prevent the profiling of individuals should have priority.

### Box 4. Merging big spatial data with models to build synthetic building stocks

Datasets capturing the building stock of cities are becoming available worldwide, either through crowd-sourced data (e.g. OpenStreetMap) or through the combination of official property (tax records) and cadastral data (geometrical information). Many cities are starting to develop 3D digital cadastre of their building stock at different levels of detail, inter alia to compute heat/cooling demands of individual buildings (Nouvel et al. 2015). However, the precise geometry is mostly irrelevant for estimating energy demand (Kim et al. 2014), and instead it is more important to hybridize spatial data with other data to develop more accurate and complete energy demand models.

A novel modelling technique to achieve this goal involves the construction of a synthetic building stock, benchmarked to available small area census statistics allows to match the synthetic building stock data to available cadastre data (Muñoz Hidalgo. et al. 2016). This method maintains relevant geometrical information of cadastre data, but additionally imports information from the census or other national surveys (Muñoz Hidalgo. et al 2016). The approach allocates construction materials to the individual buildings, allowing for the estimation of indirect GHG emissions in addition to
direct GHG emissions. The main advantage of this method is its transferability and scalability. The model does not depend on local data availability or specific data formats, and a synthetic building stock can be generated at national level (Muñoz Hidalgo 2016).

Figure 8: Estimated heat demand with a synthetic building stock based on 3D cadastre data and synthetic families based on census data. Source: (Muñoz Hidalgo. et al 2016).

As a demonstration of the method, the synthetic building stock for the city of Hamburg has been constructed with the available cadastral data and populated with synthetic families (Figure 8) (Muñoz Hidalgo. et al 2016), relying on a spatial microsimulation method (Orcutt 1957, Clarke 1996, Tanton et al 2011). The World Bank has successfully implemented this methodology for the estimation and mapping of poverty levels (Hentschel et al 2000, Elbers et al 2003) and UN Environment is currently working on the development of a Spatial Microsimulation Urban Metabolism (SMUM) model for the simulation and projection of resource flows of cities (Muñoz Hidalgo 2018). The SMUM model creates a sample survey via a Markov-Chain Monte-Carlo (MCMC) sampling procedure and reweights this sample via a GREGWT algorithm (Generalized Regression Weighting procedure) to known small area aggregates. The SMUM model projects resource consumption based on official population projections, which is a standard practice for most statistical offices around the world and an important urban planning tool. Data sets required for the projection of populations (population census or registers) are available for 83 percent of the global population (UN-DESA 2016), making these the ideal based-input-data for cities around the globe.

Projections of the buildings stock serve the estimation of future resource consumption of cities. The projections are simulated under a business-as-usual scenario, as well as under transition scenarios. Simulation models need to be able to compute the effect of policies at a household/individual level in order to identify vulnerable sections of the populations that might be disproportionately affected under these scenarios. The construction of synthetic populations is the only alternative to perform this type of analysis and maintain at the same time the anonymity of its population.
Synthetic literature reviews and the integration of qualitative data

Triggered by the slow progress in international climate policy over the last two decades, cities and local governments have teamed-up to combat climate change from the bottom up: thousands have developed and implemented local climate action plans. Yet, little is known about the impact those measures had on reducing emissions (Seto et al., 2014; Minx, 2017). This lack of knowledge is currently a direct barrier to learning about local climate solutions. Hence, developing a body of literature that aims to understand what solutions work for whom under what conditions and why would meet an important information demand by policymakers and practitioners. Yet, conducting such analysis at the local scale is complex and littered by data challenges.

Data-science should also make use of the data and insights published in the scientific literature. The literature on climate change is growing exponentially (Minx et al. 2017) and potentially thousands of new articles on urban climate solutions appear annually (Lamb et al. 2018). As this volume of work is quickly becoming unmanageable for individuals to track, bibliometric methods and systematic review techniques are needed to identify, extract and synthesize relevant information. In principle, this may follow similar procedures to those found in data science, i.e. gathering, validation, extraction and consolidation (Fig 9).

Systematic reviews begin with a literature database search, on platforms such as the Web of Science, Scopus, or Google Scholar. Depending on the scope of the intended review, keyword searches will return hundreds, or thousands of results. However, the attained bibliometric data – titles, abstracts, keywords, and references – can already provide significant insight into literature structure and content: (1) epistemic communities can be inferred from citation patterns; (2) thematic content can be identified using natural language processing; (3) case study locations can be extracted from titles and abstracts using a database of urban location names. Lamb et al (2018) demonstrate (1) and (2) for the literature on urban demand-side mitigation policies. They identify, for instance, a considerable body of work on concrete multi-objective policies that consider climate mitigation only as a secondary issue, such as parking management and congestion charging (Figure 10).

As prior steps to an assessment of the urban literature, these analyses put forth foundational questions: who is researching what topics, on which cities? But arguably, the full potential of such an approach is found in combination with data-science approaches (Figure XY). Data-science can reveal the structural similarities between cities, generate typologies, and identify salient issues that are shared across contexts. By contrast, the urban literature is often based in highly contextualised and difficult to generalise case studies. A combination of approaches – e.g. systematic reviews of the case study literature across cities of a particular type, focused reviews on specific policies across multiple (quantitatively expressed) contexts – suggests routes out of the current impasse in comparative urban research and case study review methods (Scott and Storper 2015, Steinberg 2015), while bearing the promise of systematic learning and horizontal knowledge exchange between cities.

With state-of-the computational tools, even contextual analysis of qualitative data can be automatized. For example, a textual analysis of CDP qualitative data with machine learning methods identifies the transport sector as focal point for emission reduction policies (Madu et al 2017).

Local non-textual knowledge and narratives provide strong complementary information to scientific study both on climate impact and social response dynamics (Alexander et al 2011). Following (Corburn 2003), local knowledge can improve urban planning in four ways (1) epistemology, adding to the knowledge base of climate policy; (2) procedural democracy, including new and previously silenced voices; (3) effectiveness, providing low-cost policy solutions; and (4) distributive justice, highlighting inequitable distributions of climate impact.
Non-textual traditional knowledge is inherently challenging to capture, and can at best be understood by dedicated and case-specific ethnographic research (e.g., Boyd et al. 2014). Comparative analysis of case studies enables a more systematic understanding of non-quantitative outcomes of policies and power dynamics, tentatively bridging the gap between place-specific insights and global dynamics (Creutzig et al. 2013). Meta-analysis and systematic reviews of ethnographic and human geography research has a larger role to play to better represent at least some dimensions of local knowledge and narratives.

Conferences and scientific events will profit from joint events with local and indigenous on climate challenges, as for example aspired in the IPCC Cities conference in Edmonton, 2018, where the pre-conference afternoon is dedicated to “A Village of Hope”, sponsored by indigenous knowledge holders. Such events will not only directly help to better integrate local knowledge into scientific discourses, but also serve the purpose by bilateral inspiration for research and action.

![Figure 9](image.png)

*Figure 9. A parallelized and synergetic processing of data and case study literature, for example with bibliometric methods and systematic reviews, has potential to synthesize available information at large scale and help upscaling urban data to a global urban science. Source: own design.*
Designing the global urban data platform

The architecture of the global urban data platform takes shape, comprising harmonized and upscaled data gathering and comparison efforts, the systematic application of remote sensing, social media and other big data, and the performance of systematic reviews and meta-analysis, integrating qualitative data, narratives and local knowledge as far as possible. However, even as teams of researchers put increasing effort into these endeavors, a common global urban data platform (GUDP) would dramatically facilitate exchange of information and accelerate knowledge accumulation.

The design of such a platform is unclear. Here, we suggest that institutions such as Future Earth, the Global Carbon Project, ICLEI, CDP, and C40 are all well positioned to play a role. Joining forces would lead to synergies. These organizations and others could create a joint platform by creating a Scientific Committee that is in charge of creating the platform, by organizing annual meetings, and by financing a data analyst that brings data together and analyzes them, in addition to efforts of research teams. Annual data publications, inspired by the annual Global Carbon Budget publications of the Global Carbon Project, could serve as academic focal point, incentivizing the contribution of researchers.

Municipal policy makers and administrations could be incentivized by high visibility of their cities on the GUDP. Prizes for both urban data gathering and commendable urban climate policies could further motivate urban participation and also increase the profile of climate staff within municipal administrations.

The time is ripe for a global urban sustainability science will develop rapidly and a global urban data platform will be established. The urgency of urban contributions to climate mitigation and adaptation becomes increasingly clear. And the increasing data availability enables the common quantitative foundations. We are optimistic, not only for Manhattan and Berlin, but also for everywhere else.
References

Acuto M, Parnell S and Seto K C 2018 Building a global urban science Nature Sustainability 1 2

Ahmad S, Baiocchi G and Creutzig F 2015 CO2 Emissions from Direct Energy Use of Urban Households in India Environmental Science & Technology 49 11312–20


Arnfield A J 2003 Two decades of urban climate research: a review of turbulence, exchanges of energy and water, and the urban heat island International Journal of Climatology 23 1–26


Bagan H and Yamagata Y 2015 Analysis of urban growth and estimating population density using satellite images of nighttime lights and land-use and population data GIScience & Remote Sensing 52 765–80

Bai X 2016 Eight energy and material flow characteristics of urban ecosystems Ambio 45 819–830

Bai X 2007a Industrial ecology and the global impacts of cities Journal of Industrial Ecology 11 1–6

Bai X 2007b Integrating Global Environmental Concerns into Urban Management: The Scale and Readiness Arguments Journal of Industrial Ecology 11 15–29


Bai X, Roberts B and Chen J 2010 Urban sustainability experiments in Asia: patterns and pathways environmental science & policy 13 312–325


Borck R and Brueckner J 2016 Optimal energy taxation in cities


plant CO₂ emissions from space and related applications *Atmospheric Measurement Techniques* 3 781–811

Boyd E, Ensor J, Broto V C and Juhola S 2014 Environmentalities of urban climate governance in Maputo, Mozambique *Global Environmental Change* 26 140–51


Brković-Bajić M 2008 Digital realm: Implications on urban development and planning *Spatium* 13–20

Broto V C and Bulkeley H 2013 A survey of urban climate change experiments in 100 cities *Global Environmental Change* 23 92–102


Bulkeley H and Betsill M M 2005 *Cities and climate change: urban sustainability and global environmental governance* vol 4 (Psychology Press)

C40 2015 Powering Climate Action: Cities as Global Changemakers *C40 CITIES*


CDP 2017a Cities Community Wide Emissions Online: https://data.cdp.net/Cities/2017-Cities-Community-Wide-Emissions/kyi6-dk5h/data

CDP 2017b Guidance for responding city governments, v1.0 Online: https://b8f65cb373b1b7b15feb-c70d8ead6ced550b4d987d7c03fcdd1d.ssl.cf3.rackcdn.com/cms/guidance_docs/pdfs/000/000/507/original/CDP-Cities-Guidance-2017.pdf?1484751625

Chavez A and Ramaswami A 2011 Progress toward low carbon cities: approaches for transboundary GHG emissions’ footprinting *Carbon Management* 2 471–82


CIESIN 2016 Global Urban Heat Island (UHI) Data Set, 2013 Online: http://dx.doi.org/10.7927/H4H70CRF

Clarke G P 1996 *Microsimulation for urban and regional policy analysis* vol 6
(https://katalog.b.tuhh.de/DB=22/XMLPRS=N/PPN?PPN=220691223: Pion)


Corburn J 2003 Bringing Local Knowledge into Environmental Decision Making: Improving Urban Planning for Communities at Risk *Journal of Planning Education and Research* 22 420–33

Crampton J W, Graham M, Poothuis A, Shelton T, Stephens M, Wilson M W and Zook M 2013 Beyond the geotag: situating ‘big data’ and leveraging the potential of the geoweb *Cartography and geographic information science* 40 130–9

Cranshaw J, Schwartz R, Hong J and Sadeh N 2012 The livehoods project: Utilizing social media to understand the dynamics of a city

Creutzig F 2016 Evolving Narratives of Low-Carbon Futures in Transportation *Transport Reviews* 36 341–60

Creutzig F 2014 How fuel prices determine public transport infrastructure, modal shares and urban form *Urban Climate* 10 63–76

Creutzig F 2015 Towards typologies of urban climate and global environmental change *Environmental Research Letters* 10 101001


Creutzig F, Corbera E, Bolwig S and Hunsberger C 2013 Integrating place-specific livelihood and equity outcomes into global assessments of bioenergy deployment *Environmental Research Letters* 8 035047

Creutzig F and He D 2009 Climate change mitigation and co-benefits of feasible transport demand policies in Beijing *Transportation Research Part D: Transport and Environment* 14 120–31


Creutzig F, Mühlhoff R and Römer J 2012 Decarbonizing urban transport in European cities: four cases show possibly high co-benefits *Environmental Research Letters* 7 044042

De Rosa M, Bianco V, Scarpa F and Tagliafico L A 2014 Heating and cooling building energy demand evaluation; a simplified model and a modified degree days approach Applied Energy 128 217–29


Dilley M 2005 Natural disaster hotspots: a global risk analysis vol 5 (World Bank Publications)


Frank T 2005 Climate change impacts on building heating and cooling energy demand in Switzerland Energy and Buildings 37 1175–85

Frolking S, Milliman T, Seto K C and Friedl M A 2013 A global fingerprint of macro-scale changes in urban structure from 1999 to 2009 Environmental Research Letters 8 024004


temperature and heat waves *Proceedings of the National Academy of Sciences* **106** 15555–15559

Ghiaus C 2006 Equivalence between the load curve and the free-running temperature in energy estimating methods *Energy and Buildings* **38** 429–35


Grübler A and Fisk D J 2013 *Energizing sustainable cities: assessing urban energy* (Routledge)


Gurney K R 2015 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** 179–182


Harris M 2012 Oh man, the crowd is getting an F in social science *Daily Crowdsourced* Online: http://dailycrowdsourced.com/crowdsourcing/articles/opinions-discussion/1158-oh-man-the-crowd-is-getting-an-f-in-social-science

Heidrich O, Dawson R J, Reckien D and Walsh C L 2013 Assessment of the climate preparedness of 30 urban areas in the UK. *Climatic Change* 120 771–784


Hinton G E and Salakhutdinov R R 2006 Reducing the Dimensionality of Data with Neural Networks. *Science* 313 504–7


JBA 2018 JBA Hazard Maps Online: http://www.jbarisk.com/


Kearney M R, Isaac A P and Porter W P 2014 microclim: Global estimates of hourly microclimate based on long-term monthly climate averages Scientific Data 1 140006


Kennedy C and Hoornweg D 2012 Mainstreaming Urban Metabolism Journal of Industrial Ecology 16 780–2


Kent J D and Capello Jr H T 2013 Spatial patterns and demographic indicators of effective social media content during the Horsethief Canyon fire of 2012 Cartography and Geographic Information Science 40 78–89


Kusumo A, Reckien D and Verplanke J 2017 Utilising volunteered geographic information to assess resident’s flood evacuation shelters. Case study: Jakarta Applied Geography 88 174–85

Lamb W F, Callaghan M, Creutzig F, Khosla R and Minx J 2018 The literature landscape on 1.5°C Climate Change and Cities Current Opinion in Environmental Sustainability

Li D and Bou-Zeid E 2013 Synergistic interactions between urban heat islands and heat waves: the impact in cities is larger than the sum of its parts Journal of Applied Meteorology and Climatology 52 2051–2064

Lin J and Cromley R G 2015 Evaluating geo-located Twitter data as a control layer for areal interpolation of population Applied Geography 58 41–7


Lohrey S and Creutzig F 2016 A ‘sustainability window’ of urban form Transportation Research Part D: Transport and Environment 45 96–111


Madu C N, Kuei C and Lee P 2017 Urban sustainability management: A deep learning perspective *Sustainable Cities and Society* **30** 1–17


McPhearson T, Parnell S, Simon D, Gaffney O, Elmqvist T, Bai X, Roberts D and Revi A 2016 Scientists must have a say in the future of cities *Nature News* **538** 165


Metabolism of Cities 2018 Data | Metabolism of Cities Online: https://metabolismofcities.org/data


Moran D D, Kanemoto K, Jiborn M, Wood R and Seto K C 2017 Carbon Footprints Concentrated in Few Global Cities


Muñoz Hidalgo M E 2018 (SMUM) Spatial Microsimulation Urban Metabolism Model.


National Institute for Environmental Studies 2018 2050 Low-Carbon Society Scenarios (LCS) - Scenario list Online: http://2050.nies.go.jp/LCS/eng/list.html


Nissan H, Burkart K, Coughlan de Perez E, Van Aalst M and Mason S 2017 Defining and Predicting Heat Waves in Bangladesh Journal of Applied Meteorology and Climatology 56 2653–70


Oke T R 1973 City size and the urban heat island Atmospheric Environment (1967) 7 769–779

Oke T R 1982 The energetic basis of the urban heat island Quarterly Journal of the Royal Meteorological Society 108 1–24

Openstreetmap 2018 OpenStreetMap OpenStreetMap Online: http://www.openstreetmap.org/


Ottelin J, Heinonen J and Junnila S 2014 Greenhouse gas emissions from flying can offset the gain from reduced driving in dense urban areas Journal of Transport Geography 41 1–9


Quercia D and Saez D 2014 Mining urban deprivation from foursquare: Implicit crowdsourcing of city land use *IEEE Pervasive Computing* **13** 30–6

Ramachandra T V, Aithal B H and Sreejith K 2015 GHG footprint of major cities in India *Renewable and Sustainable Energy Reviews* **44** 473–495


Scheuer S, Haase D and Volk M 2017 Integrative assessment of climate change for fast-growing urban areas: Measurement and recommendations for future research *PloS one* 12 e0189451

Schneider A, Friedl M A and Potere D 2009 A new map of global urban extent from MODIS satellite data *Environmental Research Letters* 4 044003

Schneider A, Friedl M A and Potere D 2010 Mapping global urban areas using MODIS 500-m data: New methods and datasets based on ‘urban ecoregions’ *Remote Sensing of Environment* 114 1733–46


Seto K C and Christensen P 2013 Remote sensing science to inform urban climate change mitigation strategies *Urban Climate* 3 1–6


Seto K C, Güneralp B and Hutyra L R 2012 Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools *Proceedings of the National Academy of Sciences* 109 16083–16088


Shan Y, Zheng H, Guan D, Li C, Mi Z, Meng J, Schroeder H, Ma J and Ma Z 2017b Energy consumption and CO2 emissions in Tibet and its cities in 2014 *Earth’s Future*

Shelton T, Poorthuis A, Graham M and Zook M 2014 Mapping the data shadows of Hurricane Sandy: Uncovering the sociospatial dimensions of ‘big data’ *Geoforum* 52 167–79


Shrestha R M and Rajbhandari S 2010 Energy and environmental implications of carbon emission reduction targets: Case of Kathmandu Valley, Nepal *Energy Policy* 38 4818–4827


Solecki W, Seto K C and Marcotullio P J 2013 It’s Time for an Urbanization Science Environment: Science and Policy for Sustainable Development 55 12–7

Stefanidis A, Crooks A and Radzikowski J 2013 Harvesting ambient geospatial information from social media feeds GeoJournal 78 319–38

Steinberg P F 2015 Can we generalize from case studies? Global Environmental Politics 15 152–175

Stevens M R 2017 Does compact development make people drive less? Journal of the American Planning Association 83 7–18


Tran H, Uchihama D, Ochi S and Yasuoka Y 2006 Assessment with satellite data of the urban heat island effects in Asian mega cities International Journal of Applied Earth Observation and Geoinformation 8 34–48


Wood S A, Guerry A D, Silver J M and Lacayo M 2013 Using social media to quantify nature-based tourism and recreation *Scientific reports* 3 2976


Young S D 2014 Behavioral insights on big data: using social media for predicting biomedical outcomes *Trends in microbiology* 22 601–2


Zangheri P, Armani R, Pietrobon M, Pagliano L, Boneta M F and Müller A 2014 Heating and cooling energy demand and loads for building types in different countries of the EU *Polytechnic University of Turin, end-use Efficiency Research Group* 3

Zhao H and Magoulès F 2012 A review on the prediction of building energy consumption *Renewable and Sustainable Energy Reviews* 16 3586–92

Zhou B, Rybski D and Kropp J P 2017 The role of city size and urban form in the surface urban heat island *Scientific Reports* 7 Online: http://www.nature.com/articles/s41598-017-04242-2