Comparing sources of uncertainty in community greenhouse gas estimation techniques

Michael Blackhurst* and H Scott Matthews

1 University Center for Social and Urban Research, University of Pittsburgh, Pittsburgh, PA, 15213, United States of America
2 Avenue C Advisors, LLC, Pittsburgh, PA, 15208, United States of America

* Author to whom any correspondence should be addressed.
E-mail: mfb30@pitt.edu

Keywords: greenhouse gas emissions, uncertainty, remote sensing, atmospheric CO₂, regression, cities

Abstract

Independent methods for estimating local greenhouse gas emissions have been developed utilizing different instrumentation, sampling, and estimation techniques. Comparing independent estimates theoretically improves understanding of emission sources. However, each method estimates emissions with varying fidelity, complicating comparisons across methods, cities, and over time. It is thus difficult for decision-makers to judge how to use novel estimation methods, particularly when the literature implies a singular method is best. We review 650 articles to define the scope and contours of estimation methods, develop and apply an uncertainty typology, and describe the strengths and weaknesses of different approaches. We identify two prominent process-based estimation techniques (summing of utility bills and theoretical modeling), three techniques that attribute observed atmospheric CO₂ to source locations (eddy covariance footprinting, dispersion models, and regression), and methods that spatiotemporally distribute aggregate emissions using source proxies. We find that ‘ground truth’ observations for process-based method validation are available only at the aggregate scale and emphasize that validation at the aggregate scale does not imply a valid underlying spatiotemporal distribution. ‘Ground truth’ observations are also available post-combustion as atmospheric CO₂ concentrations. While dispersion models can spatially and temporally estimate upwind source locations, missing validation data by source introduces unknowable uncertainty. We find that many comparisons in the literature are made across methods with unknowable uncertainty, making it infeasible to rank methods empirically. We see promise in the use of regression for source attribution owing to its controlling for confounding emissions, flexibly accommodating different source proxies, explicitly quantifying uncertainty, and growing availability of CO₂ samples for modeling. We see developing cross-walks between land use and end-use sectors as an important step to comparing process-based methods with those attributing atmospheric CO₂ to sources. We suggest pooling data streams can produce better decision support resources for cities with proper attribution of empirical fidelity.

1. Background

Over the last several decades, hundreds of cities globally have estimated their greenhouse gas emissions, many for the purpose of planning subsequent reductions [1, 2]. During this time, local emission estimation methods have changed and diversified considerably. Methods can largely be categorized according to whether they measure carbon pre- or post-combustion. Process-based methods estimate emissions as the product of fossil fuel consumption and combustion emission factors. Some process-based methods utilize high-fidelity consumer billing data to estimate emissions. These data are often summed by end-use sector, city, or state to protect sensitive information. In doing so, the temporal and spatial distributions of these data are lost. As a result, researchers use myriad proxies...
to spatially and temporally distribute aggregated emissions. Researchers have also developed methods to estimate emissions post-combustion, attributing observed atmospheric CO$_2$ to upwind source locations using regression or theoretical dispersion modeling. The literature often uses the term ‘bottom up’ to refer to a sum of individual source data. In contrast, methods that attribute aggregated process-based estimates or atmospheric carbon to finer scale source locations are referred to as ‘top down’ or ‘downscaling.’

Within these broad categories, estimation methods utilize fundamentally different instrumentation, sampling, and estimation techniques. Researchers have used both satellite- and ground-mounted (a.k.a. in situ) instruments to observe atmospheric CO$_2$. Energy flows may be observed using billing records or estimated using theoretical models. Estimation methods increasingly leverage both pre- and post-combustion observations by modeling the transport of emissions from estimated source locations to downwind observations of atmospheric CO$_2$.

In parallel to research, cities have produced estimates that mostly follow the ICLEI protocol. While the ICLEI protocol includes only process-based methods, these methods vary in their empirical fidelity. Some researchers refer to emissions estimates prepared by cities as ‘self-reported inventories’ and adopt these as the incumbent method to which more sophisticated techniques are compared.

While developing multiple, independent emission estimation methods advances the field, it remains difficult to fairly compare estimates across methods and cities and over time. Comparisons are obscured because methods vary in their treatment of variability and the resulting estimation of uncertainty. One common approach to quantifying uncertainty is to compare estimates across methods. For example, Hutchins et al. compares estimates for five process-based databases (one ‘bottom up’ and four ‘top down’) that estimate spatially gridded emissions, finding estimates differed significantly, mostly due to the proxies used for spatial allocation and that such differences increased with finer spatial scales. Zeng et al. found that satellite-based emission estimates for 60 cities poorly correlated ($R^2 = 0.18$) with ‘top down’ process-based estimates but that the sums over all cities differed by only 10%. Watham et al. found that satellite-based estimates were higher on average, but that both methods capture similar monthly seasonality. Fu et al. found moderate correlation ($R^2 = 0.22$) between monthly emissions estimated using ‘top down’ estimates compared to satellite based emissions for 48 cities. Gurney et al. found that self-reported inventories are, on average, 18% lower than ‘bottom up’ process-based estimates, with percent differences ranging from 150% to 64% higher.

While ranges from different estimation methods are helpful, true uncertainty is best quantified by comparing model estimates with ground-truth observations. Utility billing records provide one ground-truth estimate but do not cover all fossil-based sources and are only available at aggregated scales to protect customer data. Atmospheric CO$_2$ provides a separate ground-truth observation but reflects mixed emissions flows, including confounding flows from transboundary, biogenic, and other non-fossil sources. While the literature leaves the impression that estimation methods can be validated by aggregate emission measures, agreement by aggregate does not imply agreement by source. In other words, the spatiotemporal distribution of estimated versus observed emission sources may be materially different, yet still produce similar totals (the integral of the distribution). Where observations are unavailable for comparisons with predictions, uncertainty can be estimated by error propagation. However, error propagation presumes correct model structure, input distributions, and input covariance, assumptions that are currently untestable in current methods. As a result, reported uncertainty estimates may be misleading.

The literature further suggests that there is or will be a ‘one size fits all’ optimal estimation method. In reality, methods vary not only in their empirical fidelity but also in their typical source, sector, and spatial resolution and coverage (from individual sources to cities). While robust mitigation decisions will need to draw on various strengths from different methods, the trade-offs inherent in these methods are currently unclear.

The objective of this study is to improve upon understanding of sources of uncertainty and variability expressed in the data and models used to estimate greenhouse gas emissions at the local scale. We review related literature to document the scope and contours of greenhouse gas estimation methods in use, summarize the sources of uncertainty expressed in common techniques, and compare uncertainty across methods. An improved understanding of emissions uncertainty and variability is important for coordinating research and informing myriad climate change mitigation decisions. We use our results to suggest next steps aimed at improving emissions uncertainty quantification and integrating better the myriad emission estimates into policy decisions.

2. Materials and methods

We searched the Web of Science and SCOPUS databases for methodological papers relevant to city level inventories and planning. After various iterations and reviews of keyword search results, we finalized the following criteria applied to the title field:

(greenhouse OR carbon OR co2 OR ghg)
AND (city OR cities OR urban OR community OR metropolitan OR local)

NOT (soil ∗ OR organic OR black OR nutrient ∗ OR land ∗ OR forest OR agriculture ∗ OR waste OR water OR storage OR food)

The first and second lines of logic above returned articles titled as studying greenhouse gas emissions at the local level. The third line excludes articles that focus on emissions primarily out of the scope of this study, such as black carbon and emissions from biomass.

The above search returned 2287 articles published between 1964 and January 2020, and we identified a few additional articles published subsequently in preparing the manuscript. Using only the titles and abstracts, we excluded the articles that either:

- Did not explicitly or implicitly refer to quantitative GHG estimates;
- Focused on fugitive methane or emissions from sewers or decomposing organic matter;
- Used synthetic estimates of energy or emissions;
- Estimated emissions for a narrow set of activities (e.g. a single transit line);
- Focused on an evaluation of a narrow intervention (e.g. a technology assessment);
- Estimated emissions at spatial scales larger than a county; or
- Were corrections of previous manuscripts.

We use the collected articles to define the scope and contours of uncertainty sources in local greenhouse gas estimation methods. We develop and apply an uncertainty typology that differentiates methods based on their use of primary and secondary data, sampling protocols, and their use of statistical and theoretical models. We then pair this uncertainty typology with the typical source, sector, and spatial resolution and coverage (from individual sources to cities) associated with each method to identify trade-offs in existing estimation methods.

We define primary data as those describing the population studied. For example, energy use in buildings in Pittsburgh, PA would constitute primary data for estimating emissions in Pittsburgh. Secondary data describe related phenomena in other populations. For example, national samples of building energy use would constitute secondary data, if applied to Pittsburgh, PA [12].

We include only methods that estimate in full or in part anthropogenic scope 1 or scope 2 emissions. Scope 1 refers to emissions from combustion processes that occur within a city’s geographic boundary. Scope 2 refers to emissions resulting from energy consumed within a city’s boundary but generated elsewhere, such as most electricity consumption. We treat other emissions sources (transboundary, biogenic, and other non-fossil emissions) as confounding in relation to estimating emissions scopes 1 and 2. We recognize that some researchers estimate local emissions from carbon stored in biomass but exclude these studies due to low publication counts.

We identify three types of uncertainty demonstrated in methods used to estimate local emissions: instrumental uncertainty, sampling uncertainty, and model uncertainty, each of which is described briefly below.

2.1. Instrumental error (I)

A strict definition of ‘empirical uncertainty’ refers to inaccuracy, bias, or imprecision in raw measurements. All emissions estimates include some instrumental error (IE) (e.g. ‘measurement error,’ ‘aleatory uncertainty’) resulting from errors in instruments used for measurement. IE can be estimated by repeatedly sampling from a known quantity (e.g. a standardized concentration), resulting in a probability distribution describing the accuracy and potential bias of the instrument. In many applications, instrumental uncertainty is sufficiently low to treat observations as deterministic.

Once instruments are properly calibrated, they are used for observations. If a population of interest can be completely observed, then uncertainty will be limited to the propagation of IE over the population. However, such observational completeness is rare owing to cost and other practical limitations. Therefore, samples of the populations are taken. In some cases, samples alone provide sufficient estimation. In other cases, modeling may be required, which may or may not include samples of primary data. Both sampling and modeling introduce uncertainties.

2.2. Sampling uncertainty (S)

Sampling can introduce unobserved variability that can bias estimates. We call these errors ‘sampling uncertainty’ and use the label ‘S.’ For example, models built using a sample of residential electricity use in high-income homes would be biased in home size, end-use efficiency, and energy affordability, all of which introduce uncertainty when extended to other demographics. Samples can also be biased in their inclusion of confounding sources, transboundary, biogenic, and other non-fossil emissions. We use the sublabel ‘B’ where sampling of primary data could be biased. We further distinguish uncertainty from samples of secondary data with the label ‘SS’ given the extensive use of secondary data in theoretical models.

2.3. Model uncertainty

Models make simplifying structural or empirical assumptions about the relationships between emissions and their predictors. Model uncertainty is due to missing empirical information or related model dynamics that could theoretically be included, but are otherwise not. While definitions vary, some analysts use the terms ‘epistemic uncertainty’ or ‘model bias’ to describe model uncertainty. There are three
broad classes of models used to estimate local emissions: theoretical (inductive), statistical (deductive), and proxy-based models. Each method has different qualities that influence the degree to which uncertainty can be estimated.

Theoretical models make \textit{a priori} assumptions about the structure of relationships between variables, often informed by accepted theory. A simple example of a theoretical model is the energy used by building space conditioning equipment. Equation (1) leverages accepted theory of heat transfer to predict the energy consumed for space conditioning [13]:

$$E_{i,t} = A_i \times e_i \times (S_i - T_i)$$

where $E_i$ = energy used to condition building $i$ at time $t$  
$A_i$ = envelope area for building $i$  
$e_i$ = envelope thermal efficiency for building $i$  
$S_i$ = thermostat set point temperature for building $i$  
$T_i$ = outdoor temperature at time $t$

Some potential sources of model uncertainty in equation (1) are incomplete empirical distributions for model inputs and missing dynamics related to air infiltration and occupant behavior.

In contrast, statistical models infer relationships between emissions and their predictors. Taking the log of both sides, an example of a statistical model that complements equation (1) is:

$$\ln(E_{i,t}) = \beta_0 + \beta_1 \times \ln(A_i) + \beta_2 \times \ln(e_i) + \beta_3 \times \ln(S_i - T_i) + \epsilon.$$  

Statistical modelers would determine mean values for coefficients ($\beta$s) that best fit the data. If all of the coefficients were estimated to be one (within acceptable levels of significance), then equation (2) is reduced to equation (1), confirming the theory behind equation (1). However, fitted models may produce coefficients that are different from one. Statistical models need not use theoretical predictors, but could use proxies. For example, the energy used for space conditioning can be predicted using floorspace (a proxy for the envelope area) and building vintage (a proxy for envelope efficiency).

While both theoretical and statistical models simplify structural assumptions that introduce uncertainty, they differ significantly in how they quantify uncertainty. Statistical models are fit exclusively to primary data by testing different structural relationships between and among the dependent and independent variables. Fit quality is determined by the predictive power of the model and the degree to which model assumptions are met. Properly fit, statistical models estimate posterior distributions of both prediction errors (reflected in the error term $\epsilon$) and confidence intervals for individual predictors.

A benefit of statistical models is that they can test relationships between only a few variables without having to assume values for important but missing variables. For example, data can be properly fitted to equation (2) without having to measure confounding information (e.g. occupant behavior and air infiltration). Conversely, sample sizes often constrain the number of independent variables that are appropriate for modeling.

A primary benefit of theoretical models is that they can be simulated with limited to no primary data. Unlike statistical models, theoretical model predictions require \textit{a priori} empirical assumptions for all inputs. Given the number of inputs required for theoretical models, it is common to utilize secondary data in theoretical models. Most related software reflects default assumptions derived using secondary data such that limited to no primary data are needed for simulation [14, 15].

Where primary observations are available, theoretical model validation involves quantifying uncertainty through error propagation (e.g. Monte Carlo simulations), although quantification can be problematic in practice. Potential dependencies in independent variables are often unknown and, thus, are not modeled correctly. Ideally, theoretical modelers should be disciplined about differentiating input variability from uncertainty, estimating uncertainty in output distribution conditional on variable inputs [16]. However, many modelers treat input variability as uncertainty, resulting in output distributions that do not represent uncertainty exclusively, but instead include variability. Methods have been proposed to quantify empirical epistemic uncertainty [16], although uncertainty in model structure remains harder to quantify. Confidence in model uncertainty is clouded by the mixed use of primary and secondary data.

Beyond validating theoretical models adjusted for local conditions, calibration involves further adjusting model inputs to best match observations. Theoretical model calibration requires judgment in that it is often unclear which variables to adjust. For example, modelers using equation (1) may find that multiple sets of assumptions about $A$, $e$, and $S$ calibrate equally well, without knowing which set is correct. This is particularly problematic when models are over-parameterized. For example, building energy models used to estimate emissions include hundreds of variables [17]. Finally, some local emissions estimates use uncalibrated or unverified theoretical models due to missing predictions.

Proxy-based estimation methods (a.k.a., ‘down-scaling’) aggregated estimators to finer spatial scales using proxies such as population or nighttime lights [18–20]. In theory, proxy-based methods are similar to statistical models in that they assume non-physical correlations with emissions. However, proxy-based...
methods are not fit to data and therefore can neither quantify uncertainty nor capture the independent contributions of multiple variables to predictions.

We use the above variation in uncertainty quantification to further categorize model uncertainty as observed (MO), estimated from input variability (ME), and unknowable (MU). We reserve the unknowable label (MU) for theoretical models that use only secondary data, that is, they are neither validated nor calibrated. Figure 1 summarizes the overall uncertainty typology used to label emissions estimation methods.

Figure 2 applies the uncertainty and variability concepts discussed above to a hypothetical distribution of local emissions. Consider the case of estimating building-level emissions using models calibrated on a national scale. The solid red distribution in figure 2 (drawn as a hypothetically normal distribution) shows a representative sample of observed emissions. Independent of the estimation method, the estimated national distribution (shown as a dotted red line near the observed national distribution) differs from the actual, unobserved national distribution. Unless some observed samples are available for
fitting models, uncertainties in the resulting estimates are unknown.

While the solid red line represents primary data at the national level, most of the observations are secondary to a specific community, whose distribution is shown as shaded in gray. Local distribution estimates (shown as a dashed red line) lead to errors driven by both sampling uncertainty (unrepresentativeness of the secondary sample used for validation or calibration) and model uncertainty. Importantly, we consider validation or calibration proper when comparing full distributions such as the hypothetical ones shown in figure 2, not just when comparing mean, medians, or totals.

3. Results

3.1. Emissions measurement
All emission estimation methods rely on observations of either energy used in combustion equipment, estimates of ambient atmospheric carbon, demand for energy services (as in travel demands), or the technical efficiency of end-use equipment (as in vehicle fuel economies). Table 1 summarizes the instruments and measurands used to observe fossil energy use and CO₂ emissions in support of the local emissions estimation methods used in the literature. Instruments range from measuring pre-combustion energy flows (utility meters and gasoline pumps), spectrometers to estimate atmospheric CO₂ concentrations, and surveys used to estimate transportation emissions. Pre-combustion instruments can distinguish between end-use sectors. In contrast, measures of atmospheric CO₂ include mixed sources, which are indistinguishable from trans-boundary and non-fossil sources without subsequent analyses.

IE is expected to be relatively low for each measurement. The precision of energy utility meters is generally regulated by states to protect customers. Typical natural gas and electric meter precisions are less than 1% and 2% (98%–99% accurate), respectively, with some technologies demonstrating much lower precisions [21]. In practice, emission factors are often treated deterministically, reflecting sample averages across different combustion processes and fuel loads. However, there is some variation in emissions measured across different combustion processes. For example, Quilcaille et al [22] demonstrated a minimum variation of approximately 3% in combustion emission factors for anthracite coal, conventional natural gas, and crude oil, and approximately 9%, 4%, and 1% for oxidation rates of anthracite coal, conventional natural gas, and crude oil, respectively. It is not clear how much of this variation stems from instrumental uncertainty versus variation in instrument quality, natural variation in fuels, and combustion conditions. It is also unclear whether these uncertainties were independent. Nevertheless, these data provide a minimum uncertainty bound for the observed process-based estimates.

Process-based estimates for scope 2 emissions include empirical estimates for energy used to transmit and distribute (T&D) electricity. Respective estimates are often referred to as ‘transmission and distribution losses’ and are expressed as the fraction of electricity lost relative to generation. While theoretical estimation methods are available, losses are most often estimated by comparing demand (billed electricity use) to generation less electricity used at the generation facility. As a result, line loss estimates are subject to uncertainties in power flow meters for generation and billing. Losses are generally estimated to be less than 10% in the United States but vary depending on the quality of and stress on the T&D assets and the length of T&D [23, 24]. Unknown city-level variation adds to minimum uncertainty bounds for scope 2 process-based estimates.

Precision for in situ infrared gas analyzers is around or less than 5 parts per million (ppm) at a mean global CO₂ concentration of 409.8 ppm 1% [25]. Remote spectrometers on the satellite OCO-2 were designed to have IEs less than 1 ppm or 0.25% of global average of concentrations. Field tests of OCO-2 observations estimate accuracies ranging from 0.4 ppm to 2 ppm (0.1% to 0.5%) per reading [26–29]. Instrumental uncertainty can, however, make it challenging to distinguish small deviations. At an IE of 1 ppm, relatively large sample sizes would be required to distinguish a typical urban enhancement of 5 ppm from random IEs [30, 31].

3.2. Estimation methods
Primary, unbiased samples of sufficient size are often unavailable to support related emissions planning and mitigation decisions. Therefore, researchers have developed various estimation techniques to adjust and attribute emission observations to different temporal, spatial, or source/sector scales. This section describes the predominant sampling and modeling methods demonstrated in the literature, with a focus on attributing the uncertainties inherent in each method. Herein, we use the terms ‘metered’ to refer to methods that measure emissions in-situ (Eddy covariance and dispersion models) and ‘sensed’ to refer to remotely sensed emissions.

3.2.1. Process-based estimates
Process-based methods estimate emissions as the product of emissions factors, energy used in a conversion process (e.g. a boiler, a gasoline-powered vehicle), adjusting scope 2 emissions for electricity line losses. There are three broad types of scaling methods demonstrated in the literature: summing of billing records, modeled estimates, and attribution by proxy (e.g. downsampling).
Table 1. The measurands and instruments used to estimate local greenhouse gas emissions. Pre-combustion instruments, those used to measure energy flows before combustion, are bolded.

| Measurand(s)          | Instrument(s)                              | Unit           | Frequency        | Sectors             |
|-----------------------|--------------------------------------------|----------------|------------------|---------------------|
| Billed energy use     | Power flow instruments (utility meters, gasoline pumps, billing records) | Customer, Vehicle | Month, Refueling cycle | All, Transportation |
| Line losses           | Generated or consumed electricity          | Reported monthly or yearly but measured more frequently | All               |
| Travel demands        | Surveys, Odometers                         | Traveler or vehicle, Vehicle | Snapshot, Year | Transportation, Transportation |
| Fuel economies        | Survey                                     | Vehicle        | Year             | Transportation       |
| Emissions factor      | Varied means for estimating mass CO2 monitors | Vehicle        | Snapshot, Near continuous | Transportation, Electricity production |
| Ambient CO₂ concentration | In-situ infrared natural gas analyzers, Remote spectrometer | Spatial point(s), Spatially averaged (~3.5 km² for OCO-2) | Near continuous, TBD | Mixed |

Customer-level billing data, typically produced monthly, are summed to the sector scale to protect confidential information. These data often cover the entire population. Uncertainty in these methods is limited to IE from power flow meters, CO₂ meters, fuel mass scales, and small sampling biases (SS) in combustion emission factors and line loss estimates. While methods utilizing primary data (fuel sales and surveys) for transportation emissions are available [6], we identified only one article that used primary observations of fuel sales to estimate transportation emissions [32]. Utility billing records had previously cleanly distinguished emissions by end-use sector. However, the growth of electrified transportation will increasingly challenge source and sector attribution with utility records. Theoretical modeling and downscaling are the predominant methods used to estimate travel demands, individual building energy use, and unmetered industrial sources [6, 18, 33–35]. Common theoretical modeling methods use a mix of primary and secondary data, introducing sampling bias (SB). For example, the highly cited VULCAN dataset adjusts inputs to the building energy simulator eQUEST using high-level property characteristics typical of real estate tax assessments, such as floor space, building age, primary activity, class, and end-use sector [33]. Given these data are calibrated to national samples, eQUEST estimates would be analogous to estimating a local distribution (shown as a dashed red line in figure 2) from an estimated national distribution (shown as a dotted red line in figure 2). If models are not validated using primary data, resulting estimates include unknowable model uncertainty (MU). Examples of downscaled process-based data publications include ACES, ODIAC, HESTIA, and EDGAR [18, 35].

3.2.2. Metered estimates

Atmospheric CO₂ concentrations are typically measured in situ using an infrared gas analyzer. Scaling methods require complementary meteorological information (wind speed, wind direction, temperature, humidity, and precipitation) at high frequencies. As such, no temporal downscaling is needed for metered estimates. Temporal upscaling involves integrating or averaging these high-frequency readings over time horizons of interest, using error propagation to estimate relatively small IEs.

With respect to spatial scaling, researchers use micro-meteorological theory to estimate spatial characteristics of sources associated with point CO₂ concentrations. Within this broad framing, the literature demonstrates two predominant methods: the eddy covariance technique and atmospheric dispersion modeling.

The eddy covariance method assumes that turbulence from prevailing winds moves CO₂ from sources (or sinks) near the earth's surface vertically and measures the CO₂ flux based on correlations between directional wind speed and CO₂ concentration [31]. Fluxes are reported in units of mass of CO₂ per time per area and thus represent emissions averaged over the observed area called a ‘flux footprint.’

While the technique was developed for relatively homogeneous agricultural surfaces (e.g. a wheat grass prairie), applications of the eddy covariance technique to urban systems require estimation of the flux footprint. Here, the eddy covariation method can be integrated into one of many ‘footprint models,’ which make assumptions about the spatial distribution of source contributions to the observed CO₂ signal. The sheer number of different footprint models demonstrates the considerable unknowable model uncertainty (MU) in footprint models [31]. Confounding, transboundary sources are a particular
challenge for eddy covariance footprinting. Researchers can attempt to isolate confounding background emissions from urban emissions by placing towers near borders, using two towers, or using isotope analysis to isolate fossil-based emissions [36–38]. Despite the potential of multi-site sampling to improve urban estimates, we found this to be rarely practiced in our review.

Several eddy covariance studies leverage temporal variability in emission activities and spatial variability in land use to downscale emissions. In explaining emission variability, some studies apply ad hoc filters to temporally and spatially varying emissions concentrations to attribute emissions by land use, transportation, or building heating [36, 39–41]. These methods still require a footprint model for spatial attribution, demonstrating unknowable model uncertainty (MU). However, several studies regress emission variability against traffic counts or land use data [42–44] such that the uncertainty in downscaling can be quantified (MO). Regression can potentially associate upwind source proxies with measured fluxes by their covariation with wind speed and direction, controlling for unobserved background emissions. However, samples may still be spatially biased based on tower location and imbalanced prevailing wind speeds and directions (SB).

In contrast to eddy covariance footprint methods, dispersion methods use multiple $\text{CO}_2$ observations and make explicit assumptions about the location and timing of sources [45]. All dispersion methods simulate the transport of $\text{CO}_2$ from a source to a receptor, assuming the micro-meteorological theory. Dispersion methods generally use optimization to constrain solutions to observed concentrations, meteorological conditions, and source locations. As such, the emission uncertainty is estimated by error propagation. Methods vary in complexity, calibration criteria, and how they dredge the solution space. Most reviewed articles employ ‘inverse’ meteorological methods (primarily Lagrangian backward transport), meaning they model the flow of $\text{CO}_2$ from a receptor to a source as opposed to following the natural direction from source to receptor [4, 5].

Relative to eddy covariance methods, dispersion models capture more spatial heterogeneity in sources and may be more suited to overcome spatial sampling bias when scaling. While dispersion models can be validated or calibrated to primary observations of atmospheric $\text{CO}_2$, calibration data describing emission sources are missing. Recent dispersion models utilize process-based simulations to estimate strengths. Existing synthetic source strength databases have been shown to be highly inconsistent, and underreporting of uncertainty in source strength data is highly likely [8]. Uncertainties from dispersion models thus inherit these uncertainties (from process-based simulations discussed above) but also include those related to micrometeorological methods. The use of nationally-averaged synthetic data for validation and calibration of dispersion likely misrepresents actual spatial-temporal correlation of emissions at fine space scales. In particular, existing proxies used to distribute emissions in time and space have yet to recognize important explanatory sources of variability in emissions, such as demographics, end-use efficiency, or building envelope. For these reasons, emissions estimated from dispersion models demonstrate unknowable uncertainty (MU), which is expected to increase at finer time and space scales.

### 3.2.3. Remotely sensed estimates

Remotely sensed estimates are sampled frequently, such that temporal scaling is unnecessary. However, individual observations may be of insufficient quality owing to interference from moisture, aerosols, and ground objects. Extensive research has developed algorithms that filter suboptimal readings, which make up the majority of Orbiting Carbon Observatory-2 (OCO-2) observations [29, 46–48]. Filtering algorithms alone do not introduce uncertainty in observations. However, these algorithms introduce sampling biases, particularly over water, snow, ice, clouds, and urban environments. At the city scale, additional sampling biases relate to satellite orbit cycles, fixed spatial observation windows incongruous with municipal boundaries, and the confounding influence of transboundary sources. Researchers have partly overcome the latter bias by using statistical models to differentiate urban from surrounding rural emissions and to develop associations between emissions and population [30, 49, 50]. The use of proximate non-urban areas as a control has been criticized as simplistic in that it may not reflect the movement of confounding upwind sources, and researchers have employed dispersion models as an alternative [51]. Of course, dispersion models have their own strengths and weaknesses, as discussed above, and controls for upwind sources could be employed in statistical methods.

Table 2 summarizes and compares methods used to adjust and attribute emission observations to different temporal, spatial, or sector source/scales. Table 2 indicates that only one method does not introduce model uncertainty, which is the sum of customer billing data to the sector level. Metered and sensed instruments capture high-frequency observations, allowing them to use temporal variability to attribute emissions to sources or end-use sectors. The regression of these data against source proxies allows for explicit observation of uncertainty (MO). The introduction of theoretical models (building energy simulation, transportation demand simulation, dispersion models) without distributions of primary data for validation or calibration data introduces unknowable model uncertainty. Only a
| Category                                                                 | Estimation method(s)                                | Purpose                                                                 | Sampling uncertainty                                                                 | Model uncertainty                        | Spatial coverage | Temporal resolution | Scopes covered | Sectors and end-uses covered |
|------------------------------------------------------------------------|---------------------------------------------------|-----------------------------------------------------------------------|------------------------------------------------------------------------------------|------------------------------------------|-----------------|---------------------|----------------|----------------------|
| Process-based                                                          | Summing billed energy use                        | Sum emissions to protected customer information                        | SS in emissions factors and line losses                                           | MU                                        | Entire city     | Year                | Scopes 1 and 2 | Buildings + electrified transportation, Metered industrial end uses |
| Building energy simulation                                             | Sum emissions at varied resolutions               | Estimate high-resolution emissions                                     | SS, SB depending on use of primary data                                           | MU                                        | Source          | High frequency      | Scopes 1 and 2 | Buildings             |
| Travel demand simulation                                              | Estimate emissions at varied resolutions          | Estimate high-resolution emissions                                     | SS, SB depending on use of primary data                                           | MU                                        | Block to city   | Day to year         | Scopes 1 and 2 | Transportation        |
| Proxy-based (Downscaling)                                             | Spatially and temporally distributed aggregated emissions | Estimate emissions at varied resolutions                              | SS, SB depending on use of primary data                                           | MU                                        | Entire city     | Day to year         | Scopes 1 and 2 | All                   |
| Metered                                                                | Eddy covariance footprint                         | Estimate spatial extents of sources contributing to CO₂ concentrations | SB in confounding sources                                                         | MU                                        | ~1 km²          | 30–60 min           | Potentially mixed scopes | Mixed sources          |
| Eddy covariance with temporal filtering                                | Attribute CO₂ to end use sectors by filtering observations in space and time |                              | SB from filtered sample                                                         | ME                                        | ~1 km²          | 30–60 min           | Scope 1        | All                   |
| Eddy covariance w/ regression                                         | Attribute CO₂ to end use sectors by regressing against source proxies |                              | Depends on sample                                                              | MO                                        | ~1 km²          | 30–60 min           | Scope 1        | All                   |
| Dispersion modeling plus process-based simulations                     | Attribute CO₂ to sources using meteorological theory and process-based source estimates |                              | SB pending sampling locations; SS in meteorological model                       | MU                                        | Entire city     | 30–60 min           | Sources         | All                   |
| Remote sensing                                                         | Regression                                        | Attribute CO₂ to end use sectors by regressing against source proxies | SB depending on sample                                                              | MO                                        | Entire city     | Day to year         | Mixed sources   | All                   |
| Dispersion modeling                                                   | Attribute CO₂ to sources using meteorological theory and process-based source estimates |                              | SB depending on sample; SS in meteorological model                           | MU                                        | Entire city     | ~hour               | Mixed sources   | All                   |
few methods (summing billing records, dispersion modeling, sensed emissions) cover the entire city, and sensed emissions can only do so if cities fall within the satellite observation pathways.

Table 3 summarizes the trade-offs inherent in local greenhouse gas emission estimation methods. For example, dispersion modeling provides high-resolution emission estimates but at the expense of relatively high uncertainty. In contrast, utility billing records provide high-fidelity estimates for buildings, but not transportation. Regressing eddy covariance observations against traffic counts can provide high-fidelity estimates, but many installations are needed to completely cover the city. Only two methods are available for estimating scope 2 emissions: building energy simulation and billing records. Unfortunately, billing records are increasingly unable to attribute scope 2 emissions by sector, given the growth of electrified transportation.

3.2.4. Proxy-based methods
Proxy-based methods distribute aggregate emissions in time and space (e.g. ‘downscale’ emissions). Minimum uncertainty in these methods is related to the uncertainty of the proxies used for downscaling. A complete review of the many spatial proxies used to distribute emissions in time and space is beyond the scope of this study. We note that some demographic proxies involve relatively high uncertainty and point readers to articles that cover this topic in depth [8, 20]. Beyond uncertainty in proxies used for scaling and assumptions of linearity, proxy-based methods introduce unknowable uncertainty owing to a lack of validation data.

4. Discussion
Developing independent methods to observe and estimate emissions can improve understanding of emission sources and mitigation strategies. However, this review demonstrates that different methods may be too incongruous to draw fair comparisons. ‘Ground truth’ process-based data are available by source in utility billing records. However, they do not cover all fossil-fuel emissions and are aggregated to protect confidential consumer information. As a result, the underlying distribution by customer is unavailable to validate process-based theoretical models. Researchers can simulate process-based estimates by source using theoretical models, but at the expense of introducing unknowable uncertainty. Researchers can also ‘downscale’ aggregate estimates using source proxies, but these methods also introduce unknowable uncertainty. ‘Ground truth’ data are also available post-combustion as point atmospheric CO2 concentrations, but these data reflect mixed sources and do not cover scope 2 emissions. Regression, dispersion modeling, and ad hoc filters use spatial and temporal variability in atmospheric CO2 to attribute upwind sources, each with distinct uncertainty sources.

Theoretical models, such as building energy simulation, transportation demand models, and dispersion models, are useful for estimating the spatial and temporal distribution of emissions. To date, however, these methods have only been compared against city totals derived using independent methods. Alignment by city total does not imply that the underlying spatial and temporal source distributions are correct. Comparisons by source are needed for method validation. City totals often reflect estimates from mixed methods with mixed empirical fidelity. For example, Gurney et al. [7] compared theoretical estimates with ‘self-reported’ totals. Most self-reported inventories reflect a combination of primary utility billing records (covering most building end uses) and theoretical estimates for the remainder. A deterministic sum of emissions from mixed methods is a dubious benchmark, clouding underlying methodological and empirical differences.
As a result, we emphasize that no emission estimation method is universally superior. Methods vary by measurement frequency, spatial, sector, and source coverage, spatial, sector, and source resolution, as well as uncertainty sources and quantification. More importantly, true ground-truth data are rare and nearly missing altogether for transportation, indicating that the efficacy of existing theoretical methods is ultimately unknown.

Tools that integrate multiple estimates would improve decision-making. Researchers would benefit from comparing estimates across methods, where divergence would signal worthwhile inquiry. Decision-makers could use convergence to make more robust decisions. Tools that include new ontological and metadata labels, such as those presented herein, would promote a fairer application of emission estimates and facilitate new social science inquiry related to how decision-makers respond to varying fidelity. For example, how do decision-makers tradeoff increased spatiotemporal resolution with increased uncertainty? These questions are important as decision-makers attempt to integrate co-benefits (e.g. public health, equity) into emissions management.

Primary source data for buildings are increasingly available for methodological improvements [52–54]. These data, as well as emissions from municipal buildings, are not subject to the confidentiality constraints typical of utility data. As such, these primary samples present opportunity for experimentation to compare estimates at different spatiotemporal scales derived using different estimation methods. Regression of these data, as well as national samples [12, 55], against source proxies could produce source distributions with improved empirical fidelity. An alternative, but less robust approach, would be to encourage utilities to publish source distributions as opposed to the sector totals (the integral of the distribution), which would preserve confidentiality but offer more empirical insight. Despite these opportunities, electrified transportation will challenge sector attribution with utility data, and surveys may be needed to overcome this challenge.

Regression also has potential to attribute atmospheric CO₂ to sources. Unlike theoretical models that require untested assumptions, regression models can fit atmospheric CO₂ to flexible source proxies (e.g. land use, population, and roadway classification) with explicit uncertainty quantification. While missing primary data have previously constrained applications, new satellite data streams are designed to capture city-level emissions [56, p 3]. Regression has added methodological advantage of parsing confounding emissions by way of a global intercept. Regression methods are also better suited for measurement and verification of mitigation measures. Panel analyses, such as difference-in-difference methods, require few explanatory parameters to measure and verify emissions changes from planned interventions, where fixed time, subject, and group effects can serve as robust proxies for more detailed and elusive measures of mitigation (e.g. energy efficiency swaps). It is unclear how simulation methods, either process-based or dispersion models, can measure and verify mitigation measures owing to over-parameterization and a lack of primary data available for proper error propagation.

Despite this potential, it is highly unlikely that existing methods can support benchmarking against historically planned reductions. Historical reductions are nearly all presented as deterministic and on the order of 1% per year [57], which is smaller than expected uncertainty bounds. Uncertainty and variability in historical aggregated utility data may be estimated by error propagation conditional on variation presented in secondary data, but the efficacy of this approach is currently unclear. Instead of benchmarking against city-wide target reductions, narrowing measurement and verification to place-based or specific mitigation actions (e.g. a solar panel incentive program) may be more effective, which would require sub-sector emissions resolution.

Consistent source proxies for emission sources are needed to make comparisons across both estimation methods and cities. For example, observed differences between top-down estimates and those estimated by in situ meters may simply be due to difference in categorization of sources by end use sector as opposed to true analytical differences. However, it is unclear how well land use categories map both across cities and to end-use sector assignments in process-based methods. Developing and applying crosswalks between land use and end-use sectors could render more fair comparisons of estimates across methods and cities.

Data availability statement

All data that support the findings of this study are included within the article (available online at stacks.iop.org/ERL/17/053002/mmedia).

Acknowledgments

The authors would like to thank Dr Mario Bergès at Carnegie Mellon University and Dave Johnson at LI-COR for their helpful suggestions on this manuscript. The authors acknowledge no external funding sources supporting this work.

ORCID iDs

Michael Blackhurst, https://orcid.org/0000-0001-8193-1365
H Scott Matthews, https://orcid.org/0000-0002-4958-5981
References

[1] Wei T, Wu J and Chen S 2021 Keeping track of greenhouse gas emission reduction progress and targets in 167 cities worldwide Front. Sustain. Cities 3 64
[2] ICLEI Pressing forward: 2020 annual report March 2021 (available at: https://icleiusa.org/resources/pressing-forward-2020-anual-report/)
[3] Lauvaux T et al 2020 Policy-relevant assessment of Urban CO2 emissions Environ. Sci. Technol. 54 10237–45
[4] Hutchinson M, Oh H and Chen W-H 2017 A review of source term estimation methods for atmospheric dispersion events using static or mobile sensors Inf. Fusion 36 130–48
[5] Shankar Rao K 2007 Source estimation methods for atmospheric dispersion Atmos. Environ. 41 6964–73
[6] Fong W K, Sotos M, Douct M, Schultz S and Deng-Beck C 2014 Global protocol for community-scale greenhouse gas emission inventories ICLEI (available at: https://jghgrprotocol.org/sites/default/files/standards/GHGP_GPC_0.pdf)
[7] Gurney K R, Liang J, Roest G, Song Y, Mueller K and Lauvaux T 2021 Under-reporting of greenhouse gas emissions in U.S. cities Nat. Commun. 12 1
[8] Hutchins M G, Colby J D, Marland G and Marland E 2017 A comparison of five high-resolution spatially-explicit, fossil-fuel, carbon dioxide emission inventories for the United States Mitig. Adapt. Strateg. Glob. Change 22 497–72
[9] Zheng B, Chevallier F, Ciais P, Broquet G, Wang Y, Lian J and Zhao Y 2020 Observing carbon dioxide emissions over China’s cities and industrial areas with the Orbiting Carbon Observatory-2 Atmos. Chem. Phys. 20 8501–10
[10] Watham T, Padalia H, Rani S, Verma P A and Chauhan P 2021 Seasonal dynamics and impact factors of atmospheric CO2 concentration over sub tropical forest canopies: observations from eddy covariance tower and OCO-2 satellite in Northwest Himalaya, India Environ. Monit. Assess. 193 106
[11] Fu P, Xie Y, Moore C E, Myint S W and Bernacchi C J 2019 A comparative analysis of anthropogenic CO2 emissions at city level using OCO-2 observations: a global perspective Earth’s Future 7 1058–70
[12] DOE (U.S. Department of Energy) 2018 Residential energy consumption survey, 2015 U.S. Department of Energy, Energy Information Administration (EIA), Microdata (available at: www.eia.gov/consumption/residential/)
[13] Incropera F P, DeWitt D P, Bergman T L and Lavine A S 2006 Introduction to Heat Transfer 5th edn (Hoboken, NJ: Wiley)
[14] James J Hirsch & Associates (JJH) 1990-1999 Eddy Covariance: A Practical Guide to Measurement and Data Analysis vol 12 Springer Atmospheric Sciences (Berlin: Springer)
[15] Guterek K R, Hutrya L R and Wing I S 2015 Cities, traffic, and CO2: a multidecadal assessment of trends, drivers, and scaling relationships Proc. Natl Acad. Sci. USA 112 4999–5004
[16] Zhou Y and Gurney K 2010 A new methodology for quantifying on-site residential and commercial fossil fuel CO2 emissions at the building scale spatial and hourly scale Carbon Manage. 1 45–56
[17] Oda T, Maksyutov S and Andres R J 2017 The open-source data inventory for anthropogenic CO2 (ODIAC) fossil fuel emission model version 3.0 (ODIAC v3.0) Earth Syst. Sci. Data 10 87–107
[18] Gurney K R, Patarauskas R, Liang J, Song Y, O’Keefe D, Rao P, Whetstone J R, Duren R M, Eldering A and Miller C 2019 The Hestia fossil fuel CO2 emissions data product for the Los Angeles megacity (Hestia-LA) Earth Syst. Sci. Data 11 1309–35
[19] Kordowski K and Kuttler W 2010 Carbon dioxide fluxes over an urban area Atmos. Environ. 44 2722–30
[20] Jarvi L, Rannik Ü, Kokkonen T, Kurppa M, Karppinen A, Rasouli S and Timmermans H 2012 Uncertainty in travel demand forecasting models: literature review and research agenda Transp. Lett. 4 55–73
[21] Roy C J and Oberkampf W L 2011 A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing Comput. Methods Appl. Mech. Eng. 200 2131–44
[22] Eisenhowe B, ONeil Z, Fonoherov V and Mezi 2012 Uncertainty and sensitivity decomposition of building energy models J. Build. Perform. Simul. 5 171–84
[23] Gately C K and Hutrya L R 2017 Large uncertainties in Urban-scale carbon emissions J. Geophys. Res.: Atmos. 122 11242–60
[24] Gurney K R, Liang J, O’Keefe D, Patarauskas R, Hutchins M, Huang J, Rao P and Song Y 2019 Comparison of global downscaled versus bottom-up fossil fuel CO2 emissions at the urban scale in four U.S. urban areas J. Geophys. Res.: Atmos. 124 2832–40
[25] Geng G, Zhang Q, Martin R V, Lin J, Hao H, Zheng B, Wang S and He K 2017 Impact of spatial proxies on the representation of bottom-up emission inventories: a satellite-based analysis Atmos. Chem. Phys. 17 4131–45
[26] Parker S et al 2015 Metering best practices: a guide to achieving utility resource efficiency (Richland, WA: Pacific Northwest National Laboratory (PNNL)) (available at: https://datacenters.1bl.gov/sites/default/files/Metering%20Best%20Practices_2015.pdf)
[27] Quilcaile Y, Gasser T, Ciais P, Lecoq F, Janssens-Maenhout G and Mohr S 2018 Uncertainty in projected climate change arising from uncertain fossil-fuel emission factors Environ. Res. Lett. 13 044017
[28] Wong L 2011 A Review of Transmission Lenses in Planning Studies CEC: 200-2011-089 California Energy Commission
[29] Razanousky M, Short T and Swayne T 2012 Assessment of transmission and distribution losses in New York state (Electric Power Research Institute) (available at: www. nyserda.ny.gov/-/media/Files/Publications/Research/Electric-Power-Delivery/epri-assessment-losses.pdf) (Accessed 14 July 2021)
[30] Burba G and Anderson D 2010 A Brief Practical Guide to Eddy Covariance Flux Measurements: Principles and WorkFlow Examples for Scientific and Industrial Applications (LI-COR Biosciences)
[31] Frankenberg C et al 2015 The Orbiting Carbon Observatory (OCO-2): spectrometer performance evaluation using pre-launch direct sun measurements Atmos. Meas. Tech. 8 301–13
[32] Bell E et al 2020 Evaluation of OCO-2 X variability at local and synoptic scales using Lidar and in situ observations from the ACT-America campaigns J. Geophys. Res.: Atmos. 125 c02019(J03140)
[33] Wunch D et al 2010 Calibration of the total carbon column observing network using aircraft profile data Atmos. Meas. Tech. 3 1351–62
[34] Wunch D et al 2017 Comparisons of the Orbiting Carbon Observatory-2 (OCO-2) XCO2 measurements with TCCON Atmos. Meas. Tech. 10 2209–38
[35] Schwander F M et al 2017 Spaceborne detection of localized carbon dioxide sources Science 358
[36] Vesala T, Eugster W and Ojala A 2012 (Eddy Covariance: A Practical Guide to Measurement and Data Analysis vol 12 Springer Atmospheric Sciences (Berlin: Springer)) pp 365–76
[37] Gately C K, Hutrya L R and Wing I S 2015 Cities, traffic, and CO2: a multidecadal assessment of trends, drivers, and scaling relationships Proc. Natl Acad. Sci. USA 112 4999–5004
[38] Zhou Y and Gurney K 2010 A new methodology for quantifying on-site residential and commercial fossil fuel CO2 emissions at the building scale spatial and hourly scale Carbon Manage. 1 45–56
[39] Oda T, Maksyutov S and Andres R J 2017 The open-source data inventory for anthropogenic CO2 (ODIAC) fossil fuel emission model version 3.0 (ODIAC v3.0) Earth Syst. Sci. Data 10 87–107
[40] Gurney K R, Patarauskas R, Liang J, Song Y, O’Keefe D, Rao P, Whetstone J R, Duren R M, Eldering A and Miller C 2019 The Hestia fossil fuel CO2 emissions data product for the Los Angeles megacity (Hestia-LA) Earth Syst. Sci. Data 11 1309–35
[41] Kordowski K and Kuttler W 2010 Carbon dioxide fluxes over an urban area Atmos. Environ. 44 2722–30
[42] Jarvi L, Rannik Ü, Kokkonen T, Kurppa M, Karppinen A, Kozunternov R, Dantala P, Vesala T and Wood C R 2018 Uncertainty of eddy covariance flux measurements over an urban area based on two towers Atmos. Meas. Tech. 11 5421–38
[43] Weisett L J, Salmon J A, Turnbull J C and Schwendenmann L 2016 Temporal variability in the sources and fluxes of CO2 in a residential area in an evergreen subtropical city Atmos. Environ. 143 164–76
[44] Coutts A M, Beringer J and Tapper N J 2007 Characteristics influencing the variability of Urban CO2 fluxes in Melbourne, Australia Atmos. Environ. 41 51–62
[45] Dahlk F, Griessbaum F, Schmidt A and Klemm O 2010 Direct measurement of CO2 and particle emissions from an urban area Meteorol. Z. 19 1–11
[41] Kleingeld E, van Hove B, Elbers J and Jacobs C 2018 Carbon dioxide fluxes in the city centre of Arnhem, A middle-sized Dutch city *Urban Clim.* 24 994–1010

[42] Crawford B and Christen A 2015 Spatial source attribution of measured urban eddy covariance CO$_2$ fluxes *Theor. Appl. Climatol.* 119 753–55

[43] Nemitz E, Hargreaves K J, McDonald A G, Dorsey J R and Fowler D 2002 Meteorological measurements of the Urban heat budget and CO$_2$ emissions on a city scale *Environ. Sci. Technol.* 36 3139–46

[44] Hong J-W, Lee S-D, Lee K and Hong J 2020 Seasonal variations in the surface energy and CO$_2$ flux over a high-rise, high-population, residential urban area in the East Asian monsoon region *Int. J. Climatol.* 40 4384–407

[45] Ciais P, Rayner P, Chevallier F, Bousquet P, Logan M, Peylin P and Ramonet M 2010 Atmospheric inversions for estimating CO$_2$ fluxes: methods and perspectives *Clim. Change* 103 69–92

[46] O’Dell C W et al 2012 The ACOS CO$_2$ retrieval algorithm—part 1: description and validation against synthetic observations *Atmos. Meas. Tech.* 5 99–121

[47] O’Dell C et al 2018 Improved retrievals of carbon dioxide from Orbiting Carbon Observatory-2 with the version 8 ACOS algorithm *Atmos. Meas. Tech.* 11 6539–76

[48] Taylor T E et al 2016 Orbiting Carbon Observatory-2 (OCO-2) cloud screening algorithms: validation against collocated MODIS and CALIOP data *Atmos. Meas. Tech.* 9 3

[49] Labzovskii L D, Jeong S-J and Parazoo N C 2019 Working towards confident spaceborne monitoring of carbon emissions from cities using Orbiting Carbon Observatory-2 *Remote Sens. Environ.* 233 111359

[50] Wu D, Lin J C, Oda T and Kort E A 2020 Space-based quantification of per capita CO$_2$ emissions from cities *Environ. Res. Lett.* 15 035004

[51] Wu D, Lin J C, Fasoli B, Oda T, Ye X, Lauvaux T, Yang E G and Kort E A 2018 A Lagrangian approach towards extracting signals of urban CO$_2$ emissions from satellite observations of atmospheric column CO$_2$ (XCO$_2$): x-stochastic time-inverted Lagrangian transport model (‘X-STILT v1’) *Geosci. Model Dev.* 11 4843–71

[52] Mohammadizaii R, Copeland S and Bilec M M 2021 Urban building energy model: database development, validation, and application for commercial building stock *Energy Build.* 248 111175

[53] DOE (U.S. Department of Energy) Green Button Energy.gov (available at: www.energy.gov/data/green-button) (Accessed 18 August 2021)

[54] Seyrfar A, Ataei H and Derrible S 2020 A review of building energy benchmarking policies across the U.S. cities (6 August) (available at: www.energy-proceedings.org/wp-content/uploads/2020/12/aeab2020_paper_325.pdf)

[55] DOE (U.S. Department of Energy) 2022 Commercial buildings energy consumption survey, 2018 (Washington, DC: Energy Information Administration (EIA)) (available at: www.eia.gov/consumption/commercial/)

[56] Eldering A, Taylor T E, O’Dell C W and Parlick R 2019 The OCO-3 mission: measurement objectives and expected performance based on 1 year of simulated data *Atmos. Meas. Tech.* 12 2341–70

[57] Blackhurst M, Matthews H S, Sharrard A L, Hendrickson C T and Azevedo I L 2011 Preparing US community greenhouse gas inventories for climate action plans *Environ. Res. Lett.* 6 034003