University Climate Impact: Comparing Modeling Approaches for Predicting Greenhouse Gas Emissions from Higher Education Institutions

Anna C. Smith

ABSTRACT

Carbon Footprint (CF) calculation and reporting allows companies and organizations to quantify their emissions of greenhouse gases (GHGs). Higher education institutions (HEIs) play an important role in promoting sustainable development and mitigating climate change, and are increasingly reporting their CFs. A major challenge in HEI CF reporting is that there is no international standard for GHG accounting in the higher education sector, which has resulted in notable gaps and inconsistencies in in reported CFs among HEIs. In order to address sub-scope level gaps in HEI CFs, this study compares average emissions intensity (EIs) and linear regression modeling as approaches to predict missing emissions. Data on GHG emissions and institutional characteristics was downloaded from the Sustainability Tracking, Assessment & Rating System (STARS), and used to train five models to predict missing data for a total of thirteen emissions categories. The study found that multiple regression models produced the most accurate predictions for most emissions categories; however, the simple linear regression models using gross square footage of building space (GSF) or full-time equivalent student and employees (FTE) as the predictor variables offered similar accuracy for most emissions categories, while requiring only one input feature. The highest prediction accuracy was achieved for gross emissions (training $R^2 = 0.798$, testing $R^2 = 0.765$). It was concluded that average EI is not a robust modeling approach. HEI CF reporting needs to be expanded, and an international standard for GHG accounting in the higher education sector must be developed and formally adopted.

KEYWORDS

Carbon footprint, average emissions intensity, linear regression, FTE, GSF.

INTRODUCTION

Climate change is one of the most critical challenges facing the world in the 21st century, threatening to undermine sustainable development and exacerbate inequality on a global scale (Cevik and Jalles 2023). Key social entities, including companies, organizations, and institutions, are increasingly recognizing, measuring, and acting on their climate impact in an effort to combat the climate crisis (Abad-Segura et al. 2019, Yilmaz and Can 2020). An important step in this process involves quantifying emissions of greenhouse gases (GHGs) that are produced as a result of an organization's activities. Seeing as GHGs trap heat in the atmosphere and are the main driver of anthropogenic climate change, quantifying GHG emissions allows emitters to track emissions performance over time, and assess which activities have the greatest impact. To this end, GHG emissions are commonly calculated and represented in a carbon footprint (CF) that is specific to the organization (Global Footprint Network 2022). In the corporate world, CF reporting has become best practice in the sphere of corporate social responsibility (Walenta 2021, Carbon Disclosure Project 2022). Calculating and reporting CFs is an important step in taking responsibility for an entity's contribution to climate change.

Universities and higher education institutions (HEIs) are increasingly following suit by calculating and reporting their own CFs, in order to signal accountability, transparency, and leadership on sustainability matters (Li et al. 2021, Valls-Val and Bovea 2021). HEIs are considerable emitters of GHG emissions; in 2005, U.S. HEIs contributed an estimated 121 million $MtCO_2e$, or around 2% of U.S. GHG emissions – comparable to one-quarter of California's total emissions that year (Sinha et al. 2010). Trends in HEI sustainability action and reporting vary across the globe, and HEIs in North America and Europe are currently leading initiatives on these fronts (Alonso-Almeida et al. 2015, Amaral et al. 2020, Perchinunno and Cazzolle 2020, De Iorio et al. 2022). In the U.S., the American College & University Presidents' Climate Commitment (ACUPCC) has encouraged hundreds of universities to create climate action plans, compile GHG emissions inventories, and publish regular progress reports with the goal of reaching climate neutrality. The program has been successful in spurring action and reducing emissions amongst participating HEIs (Foust 2016, Dyer and Dyer 2017). As institutions on the frontier of research and innovation, HEIs are expected to take charge and help guide societies towards sustainable development (Peer and Stoeglehner 2013, Lozano et al. 2013, Sedlacek 2013, McCowan 2020);

HEIs play a central role in educating future leaders, and have significant influence on governance at the regional and national level, which places these institutions in a position of power when it comes to promoting sustainable decision-making (Sedlacek 2013, Cordero et al. 2020). Creating and diffusing knowledge on climate change and sustainability challenges, expanding efforts to measure the climate impact of the higher education sector, and developing accurate and comprehensive CF reporting are central aspects to HEIs fulfilling their social responsibility of researching, addressing, and mitigating climate change.

Key challenges that remain are that CF reporting is not practiced among all HEIs, and that there is no international GHG accounting framework that is specifically designed for the university context (Robinson et al. 2018, Valls-Val and Bovea 2022). As of 2016, only around 32% of U.S. HEIs had formally committed to GHG accounting and reduction (Foust 2016). Moreover, HEI sustainability reporting has been found to be in an early stage across the board (Ceulemans et al. 2015, Alonso-Almeida et al. 2015). While there have been some efforts in the literature to create universal GHG accounting frameworks for HEIs, no tool or method has successfully been adopted as the international standard (Robinson et al. 2018, Valls-Val and Bovea 2021, 2022). Consequently, HEIs have mostly relied on various GHG emissions calculation frameworks, such as the GHG Protocol, which were designed predominantly for profit-making enterprises; adapting and interpreting these corporate standards comes with a broad range of assumptions and caveats that limits methodological consistency and accuracy of the resulting CFs across different institutions (Robinson et al. 2018). These developments have led to notable inconsistencies and gaps in reported HEI CFs, which makes comparability of emissions and climate impact between HEIs a challenge (Robinson et al. 2018, Clabeaux et al. 2020, Helmers et al. 2021, Second Nature 2022a, Sprenk 2022). In order to, nonetheless, leverage available data and analyze the climate impact of HEIs, gaps in HEI-reported GHG emissions data need to be bridged.

In this study, I am addressing the central research question: How can HEI GHG emissions be most accurately predicted for specific scopes and sub-scope categories of emissions? The sub-questions guiding this research include (1) Can median emissions intensities for different scopes and sub-scope categories be used to accurately predict missing emissions data? (2) Can linear regression models for different scopes and sub-scope categories produce more accurate predictions? (3) Overall, how accurate and generalizable are these prediction models for emissions from different scopes and sub-scope categories? The research objectives for this study include accessing data from the Sustainability Tracking, Assessment & Rating System (STARS) and analyzing it in a Python Jupyter Notebook. HEI emissions data will be used to train prediction models for scope 1, 2, and 3 emissions, and their respective sub-scope categories of emissions, using median emissions intensities (SQ1), as well as simple and multiple regression models (SQ2). The resulting models will be used to generate emissions predictions from training data to test the internal validity of the models ; test data will be reserved to assess the external validity of the models. Model performance will be compared using RMSE and the R^2 coefficient of determination. My working hypotheses are that median emissions intensities can be used to create relatively reliable models for overarching emissions scopes, especially scopes 1 and 2; however, I predict that regression models can create more reliable predictions across the board, and are especially important for predicting emissions in sub-scope categories. I also expect that model performance will be inconsistent between emissions categories, and be largely affected by data availability.

BACKGROUND

Climate impact assessments and GHG accounting for HEIs

right° based on science

For this thesis project, I am collaborating with the research outreach team *right. open* at *right° based on science* (hereafter, right°). Right° is a company based in Frankfurt, Germany, that has created a climate impact assessment tool, the X-Degree Compatibility (XDC) Model, which quantifies and contextualizes the contribution to climate change of a certain company, building, or financial portfolio (right° based on science 2023). One challenge with traditional CFs is that climate impact is only gaged through emissions reported in metric tons of carbon dioxide equivalent (MtCO2e). Although CFs reported in MtCO2e make it possible to compare the relative magnitudes of emissions arising from different sources or entities, the MtCO2e unit does not effectively communicate how much these emissions are contributing to climate change (Helmke et al. 2020, right° based on science 2022). If we compare entity A and B, which produce 10

MtCO2e and 1,000 MtCO2e respectively, we can see that entity B is a larger emitter than entity A; however, it is not obvious how much these emissions contribute to climate change.

In response to this inherent weakness of CF reporting, right° developed the XDC Model, which translates CFs in MtCO2e to a climate impact metric in °C of global warming. The XDC model addresses the question: "How much global warming would occur if the entire world had the same climate performance as the entity in question?" (right° based on science 2023). Data on an entity's GHG emissions and economic contribution are used as inputs to the XDC model, which then scales up the entity's impact to the global scale to output predicted global temperature increase in °C that would arise from the resulting emissions. A direct comparison between the °C warming output of the model and the 1.5°C or 2.0°C warming goals defined by the Paris Agreement allows the entity to assess how well their operations are aligned with international climate change mitigation goals. This analysis presents the activities of an individual entity in the context of a global, concerted effort to combat the climate crisis, making it a much more powerful and comprehensible climate impact metric than MtCO2e alone.

Currently, right° has applied its model to industrial companies, financial portfolios, and the real estate sector. Now, they are in the process of adapting their methodology and software model to the higher education sector. My research contributes to these efforts, by analyzing how gaps in HEI GHG emissions data can be overcome.

GHG Protocol, emissions scopes, and sub-scope categories

The World Resources Institute (WRI) and the World Business Council for Sustainable Development (WBCSD) together created the GHG Protocol Corporate Standard, which is the most commonly used GHG emissions accounting standard (Greenhouse Gas Protocol 2022). The GHG Protocol outlines three scopes of emissions, and provides guidelines on how to calculate and report these emissions. Within each scope, the GHG Protocol defines several sub-scope categories that describe specific sources or activities from which emissions arise (Figure 1); taken in sum, the emissions from the sub-scope categories make up the total emissions in a scope (Table 1). The scopes and sub-scope categories defined in the GHG Protocol, or a variation of them, are used by most HEIs to calculate and report their emissions (Valls-Val and Bovea 2021); therefore, understanding this framework is central to this thesis, and more broadly to studying GHG emissions from HEIs.

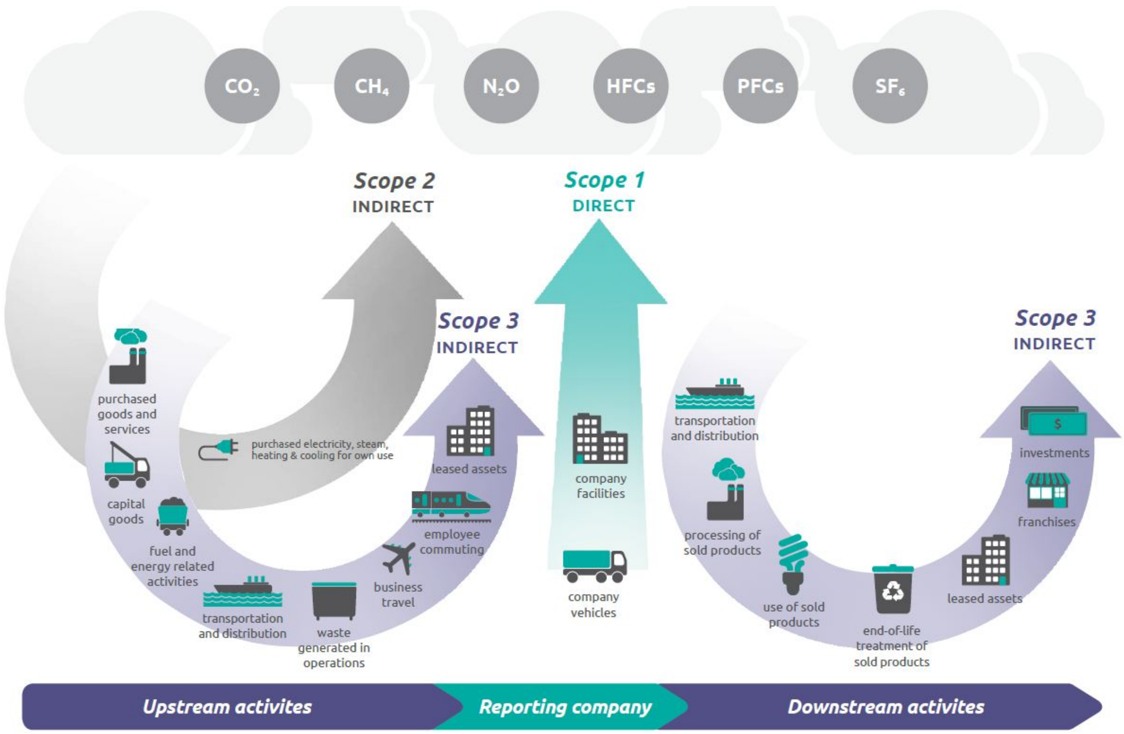

Figure 1. **Overview of GHG Protocol scopes and emissions sources across an entity's value chain**. Image from Greenhouse Gas Protocol's "Corporate Value Chain (Scope 3) Accounting and Reporting Standard" 2011 report.

Scope 1 ("direct emissions"). Scope 1 emissions are GHG emissions that arise from sources directly owned or controlled by a company or entity (Greenhouse Gas Protocol 2004). This typically includes emissions from on-site machinery such as generators, furnaces, boilers, or vehicles. The GHG Protocol defines the following sub-scope emissions categories for scope 1: (1) stationary combustion (e.g. boilers and furnaces), (2) mobile combustion (e.g. vehicle fleet), (3) fugitive emissions (e.g. GHG leaks from refrigeration or air-conditioning units), and (4) process emissions (e.g. emissions resulting from industrial processes like cement manufacturing).

Scope 2 ("indirect emissions"). Scope 2 emissions are emissions associated with the production purchased energy – primarily, purchased electricity (Greenhouse Gas Protocol 2004). The GHG Protocol scope 2 framework also applies to purchased steam, heating, and cooling; however, not all entities or companies purchase steam, heating, or cooling (Greenhouse Gas Protocol 2015).

While the generation of these resources produces GHG emissions, the emissions do not occur at sites or sources controlled by the reporting entity – therefore, these emissions are the indirect result of an entity's activities. The following sub-scope categories for scope 2 are defined by the GHG Protocol: (1) purchased electricity, (2) purchased steam, (3) purchased heating, and (4) purchased cooling.

There exist two common accounting methods for purchased electricity: market-based accounting, and location-based accounting (Greenhouse Gas Protocol 2015). Both methods calculate emissions by multiplying consumption data (kWh consumed) by an emissions factor that gives the amount of GHG produced by consuming a single unit of energy (MtCO2e/kWh). The market-based accounting approach calculates emissions based on the electricity that entities choose to purchase from their supplier. This approach considers the relative impact, or emissions factors, of the energy sources included in the mix that is supplied to the entity. Notably, in this approach, renewable energy sources have an emissions factor of 0, meaning that if an entity uses only electricity from renewable sources, their reported emissions for purchased electricity amount to zero. The location-based accounting approach, on the other hand, uses the emissions factors that are based on the average emissions intensity of the local grid area. This approach does not discount emissions based on the specific electricity mix purchased by an entity or the use of on-site renewables. The distinction between these approaches is important when discussing scope 2 emissions from purchased electricity.

Scope 3 ("other indirect and value chain emissions"). Scope 3 emissions are emissions that are produced as the consequence of an entity's activities, but are not emitted from sources controlled or owned by the company. These emissions are not included in scope 2, and are associated with the value chain of the entity, i.e. any upstream or downstream activities related to the entity's operations (Greenhouse Gas Protocol 2011). This may include the commuting of employees, or any emissions associated with the goods and services purchased by the entity, for example. Scope 3 therefore comprises many different sources of emissions, and this scope often accounts for the majority an entity's total emissions (Clabeaux et al. 2020, Valls-Val and Bovea 2021, García-Alaminos et al. 2022). At the same time, these emissions are often the most difficult to quantify, and therefore the calculation and reporting of scope 3 emissions is typically more limited (Larsen et al. 2013, Robinson et al. 2018, Valls-Val and Bovea 2021). Additionally, not all sub-scope categories apply to every entity, and some entities do not have the means or capacity to collect all the necessary data to calculate these emissions (Robinson et al. 2018).

The GHG Protocol defines the following 15 sub-scope categories of scope 3 emissions: (1) purchased goods and services, (2) capital goods, (3) fuel- and energy-related activities (not included in scope 1 or scope 2), (4) upstream transportation and distribution, (5) waste generated in operations, (6) business travel, (7) employee commuting, (8) upstream leased assets, (9) downstream transportation and distribution, (10) processing of sold products, (11) use of sold products, (12) end-of-life treatment of sold products, (13) downstream leased assets, (14) franchises and (15) investments.

The American College & University Presidents' Climate Commitment (ACUPCC) requires HEIs to report their scope 1 and scope 2 emissions, but only requires reporting for two sub-scope categories of scope 3 emissions: employee commuting and air travel paid for by or through the university (i.e. business travel). The ACUPCC "strongly encourages" signatory HEIs, "to investigate and report on additional Scope 3 emissions, especially those from sources that are large and can be meaningfully influenced by the institution" (Second Nature 2022b). The ACUPCC also leaves it up to the individual HEI's discretion to develop their own guide for completing GHG inventories, and allows a choice of GHG inventory calculators, as long as they are consistent with the standard of the GHG Protocol (Second Nature 2022b). This goes to show that guidance on GHG emissions accounting and reporting for HEIs is mostly aligned with the GHG Protocol, but is very limited, especially for scope 3. The diversity of reported emissions categories, GHG inventory collection guides, and GHG inventory calculators makes it so that the approaches and scopes of reported CFs are different between ACUPCC signatories.

Table 1. Summary of scopes and sub-scope categories of emissions. Scope 1 emissions are classified as direct emissions resulting from on-site sources and activities. Scope 2 emissions are energy-related indirect emissions, while scope 3 emissions are indirect emissions arising from upstream and downstream sources and activities along the entity's value chain.

Direct Emissions	**Indirect Emissions**		
Scope 1	Scope 2	Scope 3	
		Upstream emissions	***Downstream emissions***
1. Stationary combustion 2. Mobile combustion 3. Fugitive emissions 4. Process emissions	1. Purchased electricity 2. Purchased steam 3. Purchased heating 4. Purchased cooling	1. Purchased goods and services 2. Capital goods 3. Fuel- and energy-related activities (not included in scope 1 or scope 2) 4. Upstream transportation and distribution 5. Waste generated in operations 6. Business travel 7. Employee commuting 8. Upstream leased assets	1. Downstream transportation and distribution 2. Processing of sold products 3. Use of sold products 4. End-of-life treatment of sold products 5. Downstream leased assets 6. Franchises 7. Investments

Normalizing, modeling, and predicting HEI GHG emissions

HEI emissions intensities

Emissions intensity (EI) is a commonly used metric that is used to normalize GHG emissions between actors within a certain sector. Fundamentally, an EI is simply the amount of GHG emissions produced per one standard unit, given as a fraction with emissions as the numerator and one standard unit as the denominator:

$$Emissions\ Intensity\ (EI) = \frac{emissions}{standard\ unit}$$

EIs can be adapted and modified by changing the standard units in the denominator. For example, the denominator could be GDP (gross-domestic product) when discussing emissions on the national level, GVA (gross-value added) on the industrial or corporate level, or kilocalorie produced on the cropland level. Manufacturers in the agricultural sector may calculate their EI in

MtCO2e/kcal produced, which helps make GHG emissions from farms of different sizes comparable (Carlson et al. 2017). Similarly, the EI of nations can be calculated and compared in MtCO2e per unit GDP (Goldemberg 2020) or per capita (Desme and Smart 2018). Generally, EIs are useful as they help normalize emissions, typically by a metric related to size, which helps make emissions from entities of different sizes comparable. Furthermore, EIs from different entities in a sector can be averaged to calculate how much emissions are produced per standard unit, on average, by entities in that sector. Such a sector average EI could be used to approximate the amount of GHG emissions produced by an entity of a given size which has not collected or reported emissions data:

$$Predicted\ emissions = \left(\frac{emissions}{standard\ unit}\right)_{average} \cdot Observed\ standard\ units$$

EIs specific to the higher education sector have been developed and explored in the literature, where full-time equivalent students or employees (FTE) and gross-square footage of occupied building space (GSF) are the most commonly used standard units (Zhaurova 2008, Sinha et al. 2010, Klein-Banai and Theis 2013, Larsen et al. 2013, Helmers et al. 2021). Other, less-commonly used standard units include economic metrics, such as dollar expenditure (Larsen et al. 2013, Helmers et al. 2021). Average emissions intensities, with emissions normalized by FTE and GSF, were compared between different levels of Carnegie classification (Doctorate-granting Universities, Master's Colleges and Universities, Baccalaureate Colleges, Associate's Colleges, Special Focus Institutions, and Tribal Colleges and Universities) by Sinha et al (2010). This study used average EI for each Carnegie classification level, with emissions normalized by FTE, and enrollment data from the U.S. Department of Education's National Center for Education Statistics, to estimate total emissions from the higher education sector in 2005.

Logically, EIs help reveal the *emissions intensity* of individual HEIs. For example, an HEI with a hypothetical GSF of 400,000 is likely to produce more emissions than an HEI with only 100,000 GSF; however, by calculating their respective EI, we can see which HEI produces more emissions per square foot of building space. EIs can be a quick tool for predicting emissions for an HEI that does not report emissions – by multiplying the available standard unit data (e.g. FTE or GSF) and the average sector EI to give a predicted emissions value. For example, if HEIs produce on average 2 MtCO2e per square foot of building space, an HEI with GSF of 200,000 is expected to emit around 400,000 MtCO2e. Using average sector EI as an estimation method is advantageous since it is straightforward, easily interpretable, and only requires one input data point.

Prediction Modeling for HEI GHG Emissions

Previous studies have identified institutional characteristics that affect an HEI's GHG emissions, and can therefore be used as predictor variables in linear regression models to predict emissions. FTE and GSF are particularly important predictors for HEI GHG emissions, and have been studied repeatedly (Fetcher 2009, Sinha et al. 2010, Klein-Banai and Theis 2013). In the literature, FTE is typically used to describe full-time equivalent enrollment of students; however, FTE can be extended to include full-time equivalent employees (Nebe 2022, Sprenk 2022). Right° is using a combined student and employee FTE metric for their model development. Therefore, the FTE metric I will be using in my analysis, and referring to going forward, describes full-time equivalent students and employees.

Other predictors observed in the literature include variables related to local climate, such as mean July temperature and mean January temperature (Fetcher 2009), or heating degree days, and cooling degree days (Klein-Banai and Theis 2013). Predictors related to building area and space usage include GSF (Zhaurova 2008, Fetcher 2009, Sinha et al. 2010, Klein-Banai and Theis 2013), lab square footage, residential square footage, and health care square footage (Klein-Banai and Theis 2013, Gui et al. 2020). Select economic predictor variables that have been explored in the literature, including total income (Wadud et al. 2019) and total expenditure (Helmers et al. 2021). Other predictors related to institutional characteristics, such as Carnegie classification (Zhaurova 2008, Fetcher 2009, Sinha et al. 2010, Klein-Banai and Theis 2013), medical school status (Klein-Banai and Theis 2013, Gui et al. 2020), and public versus private institution type (Zhaurova 2008) have also been studied. These studies have shown that FTE, GSF, medical school status, and Carnegie classification are especially important determinants of HEI CFs (Fetcher 2009, Sinha et al. 2010, Klein-Banai and Theis 2013, Gui et al. 2020).

In terms of regression modeling observed in the literature, Fetcher (2009) and Klein-Banai and Theis (2013) are the most notable examples. Both studies used ACUPCC signatories as their

HEI samples. Fetcher predicted only combined scope 1 and 2 emissions data as the outcome variable, and used a stepwise multiple regression modeling approach. Emissions, FTE and GSF data was transformed using $\log_{10}$ to stabilize variance. Klein-Banai and Theis (2013) trained multiple regression models for combined scope 1 and 2 emissions, as well as for gross emissions from all scopes combined. Similar to Fetcher (2009), Klein-Banai and Theis (2013) log-transformed emissions and GSF data. This study used principal component analysis and stepwise multivariate regression to find important determinants and produce emissions predictions. What has not been observed in the literature is prediction modeling for sub-scope categories of emissions, which is necessary to fill specific gaps in CFs.

Data sources and data availability

Sustainability Indicator Management & Analysis Platform (SIMAP)

Seeing as the goal of my study is to predict GHG emissions data for different scopes and sub-scope categories of GHG emissions from HEIs, I require data for GHG emissions from HEIs with sub-scope level granularity, as well as data on institutional characteristics that could be used as predictor variables. The most commonly used data source for HEI GHG emissions in the literature is data from ACUPCC signatories (Fetcher 2009, Sinha et al. 2010, Klein-Banai and Theis 2013). The ACUPCC was developed in 2006 as a collaboration between twelve college and university presidents, and three organizations, namely Second Nature, ecoAmerica, and the Association for the Advancement of Sustainability in Higher Education (AASHE). As of 2011, Second Nature was the only remaining supporting organization of the ACUPCC, and rebranded the ACUPCC as the Presidents' Climate Leadership Commitments (PCLC) signature program in 2015. The PCLC is an expansion of the ACUPCC, and comprises three separate commitments: a Climate Commitment, a Resilience Commitment, and a Carbon Commitment (Second Nature 2023a).

In 2022, Second Nature partnered with the University of New Hampshire Sustainability Institute (UNHSI) to offer the Sustainability Indicator Management & Analysis Platform (SIMAP) as the public reporting platform for PCLC signatories (Second Nature 2023b). SIMAP is a "carbon and nitrogen-accounting platform that offers campuses a simple, comprehensive, and affordable online tool to track, analyze, and improve campus-wide sustainability" (SIMAP 2023). SIMAP was launched in 2017, and is based on the GHG Protocol standard. Prior to 2022, Second Nature had its own reporting platform for ACUPCC/PCLC signatory emissions data; however, this reporting platform is no longer accessible, and data can now be accessed only through SIMAP.

I requested data from the SIMAP Public Reporting Module, and was able to access GHG emission data with a level of granularity even smaller than the sub-scope categories defined by the GHG Protocol; emissions were reported by specific activities and sources, some of which needed to be aggregated to match sub-scope categories defined by the GHG Protocol. In terms of institutional characteristics, SIMAP collects data for Carnegie classification, U.S. state, total enrollment, and total building square footage. This presented a significant limitation for my research purposes, since I was interested in including further institutional characteristics to make predictions in my multiple regression modeling. Moreover, I needed full-time equivalent employee data to get my desired FTE count. Although additional data could be collected manually or downloaded from other data sources such as the Integrated Postsecondary Education Data System (IPEDS), it was difficult to get data for the same sample of HEIs and the same reporting years represented in the SIMAP dataset; also, combining data collected from different sources makes it more difficult to ensure consistency and reliability of the data. A final limitation presented by SIMAP data was sample size; although the ACUPCC/PCLC has over 700 signatories, only 398 HEIs were active signatories to the most recent version as of Noveember 2022, of which only 194 HEIs had publicly reported data to SIMAP. Reports from these 194 HEIs also stemmed from different years, ranging from 2013 and 2022. Using a smaller timeframe further limited my sample size. Ultimately, I did not use SIMAP data given limitations in institutional characteristics data and sample size.

Sustainability Tracking, Assessment & Rating System (STARS)

The Sustainability Tracking, Assessment & Rating System (STARS) is a "transparent, self-reporting framework for colleges and universities to measure their sustainability performance" (STARS 2023a). STARS was developed in 2006 in response to a call from the Higher Education Associations Sustainability Consortium (HEASC) for the AASHE to develop a campus sustainability rating system. The AASHE was established in 2005, and supports "the advancement of sustainability in higher education by empowering higher education faculty, administrators, staff and students to be effective change agents and drivers of sustainability innovation" (AASHE 2023). In total, 1,147 HEIs have registered to use AASHE's STARS reporting tool as of March 2023 (STARS 2023b).

STARS has a Content Display webpage that provides exact responses to specific reporting fields in STARS reports, including GHG emissions categories and a wide range of institutional characteristics. Several versions of STARS have been developed, with the sixth and most recent version being STARS version 2.2. The reporting framework changes slightly between versions, and data can be downloaded separately for each version. As of February 2023, 310 HEIs had reported under STARS version 2.2, with reporting years ranging between 2019 and 2023. Compared to SIMAP, STARS has a coarser granularity in terms of emissions categories; sub-scope categories were included for all three scopes of emissions, but some sub-scope categories represented aggregated sources or activities.

Ultimately, STARS presents a slight tradeoff in sub-scope emissions category granularity, but offered advantages over SIMAP in terms of sample size, availability of institutional characteristics data, and geographic coverage; STARS has international HEI participation, while SIMAP only included HEIs from the United States. In the literature, STARS and SIMAP have been contrasted as follows: "The findings showed that the STARS were by far the most comprehensive assessment framework for university sustainability. ACUPCC, on the other hand, fulfilled a different purpose of being a target-oriented assessment framework, primarily focusing on GHG reduction" (Shi and Lai 2013). This underscores the notion that SIMAP is closely aligned with the GHG Protocol and provides access to very granular GHG emissions data, while STARS collects more comprehensive data on sustainability-related variables, including a wide range of relevant institutional characteristics.

METHODS

Mapping emissions categories between STARS, SIMAP, and GHG Protocol

In order to produce generalizable analysis and results, I mapped emissions categories collected by STARS to the emissions categories included in SIMAP and the GHG Protocol framework. I created this mapping in Excel, listing all GHG Protocol categories in one column, and matching emissions categories from SIMAP to the GHG Protocol categories in an adjacent column. Allison Leach, Program Manager at SIMAP, signed off on the mapping from SIMAP to GHG Protocol. Finally, I matched STARS categories to SIMAP categories in a third column using STARS's 'Guidance for STARS participants using SIMAP' (see Appendix A). Monika Urbanski, Senior Manager for Data, Resources & Publications at AASHE, helped me develop and verify my mapping framework in regards to STARS. I used this mapping to aggregate emissions categories from STARS for my analysis, to better align with the GHG Protocol's sub-scope emissions categories.

Data collection, exploration, and preparation

In this study, I analyzed data on HEI GHG emissions and institutional characteristics from AASHE's STARS database. I downloaded STARS version 2.2 data from the Content Display webpage (see Appendix A). I used different queries to access the desired institutional characteristics and GHG emissions data, which were then directly downloaded as Excel files. I recorded all queries in an Excel sheet (see Appendix B). Each query produced a separate Excel sheet with data for one variable using the same sample of HEIs; I manually combined data columns for all variables into one combined Excel sheet, which I then uploaded to DataHub in order to perform analysis in a Jupyter Notebook.

From STARS, I retrieved GHG emissions data for all possible scope and sub-scope level emissions categories, including scope 1 (stationary combustion, other sources, total), scope 2 (imported electricity, imported thermal energy, total), and scope 3 (business travel, commuting, purchased goods and services, capital goods, fuel- and energy-related activities, waste, other sources, total).

The institutional characteristics data that I downloaded included institution type/Carnegie classification (qualitative categorical – Associate's, Baccalaureate, Master's, Doctorate/Research), institutional control (qualitative categorical – public, private non-profit), medical school (binary), hospital (binary), satellite campuses (binary), endowment size (quantitative continuous - USD), gross floor area of building space (quantitative continuous – square feet), floor area of laboratory space (quantitative continuous – square feet), floor area of healthcare space (quantitative continuous – square feet), full-time equivalent student enrollment (quantitative discrete – individuals), and full-time equivalent employees (quantitative discrete – individuals). The FTE metric used in my analysis was calculated manually and added as a feature, by adding full-time equivalent student enrollment and full-time equivalent employees together.

Monika Urbanski confirmed that "0" values in the data are an indication of performance rather than missing data, while missing values are signified by "—" or N/A (i.e. null). Ultimately, all null and zero values were removed from the emissions data, because for certain emissions categories, many institutions reported zero emissions, which prevented me from producing accurate, non-zero predictions. Although reported zeros are technically a measure of performance in STARS, it is not obvious whether zeros from all HEIs are indeed an indication of zero emissions, or whether emissions for that category were simply not measured. For most emission categories, it is unlikely that an HEI produced no emissions at all. For example, 31 HEIs in my sample reported zero emissions for scope 3 emissions from commuting, and 51 HEIs reported zero emissions for scope 3 emissions from waste; however, it is highly unlikely that an HEI has zero emissions associated with the commuting of its students and staff, or the treatment of its waste. Furthermore, the objective of my study was to produce non-zero emissions estimates. If an HEI knows a certain emissions category does not apply to them, it is easy to fill in zero emissions for that category. Although STARS uses the market-based approach for scope 2 emissions from electricity, I removed zeros for this category too. While these zeros have a different meaning, since they indicate that all electricity is derived from renewable sources, it is easy to fill in zero emissions knowing that an HEI uses only renewable electricity; it is more difficult, however, to get accurate non-zero emissions predictions, which I was only able to achieve by removing zeros.

For my analysis, I created a test-train split in the data, with 85% of my data being used for training, and 15% of my data being reserved for testing. Before beginning model training, I plotted histograms and boxplots of my emissions categories and predictor variables to investigate the distributions of the data. Normally distributed data is best suited for linear regression modeling, therefore I used the Shapiro-Wilk test to assess normality. I repeated this process of testing for normality for distributions of emissions intensities that I calculated using FTE and GSF as standard units. All modeling was performed in a Python Jupyter Notebook.

Average emissions intensities

In order to address SQ1, I used my training data to calculate average emissions intensities for all emissions categories using FTE and GSF as standard normalizing units. GSF here corresponds to gross floor area of building space from STARS, and FTE corresponds to the sum of full-time equivalent student enrollment and full-time equivalent employees data from STARS.

The exploratory data analysis stage showed that emissions, FTE, and GSF data was approximately normally distributed after transforming the data using $\log_{10}$ and removing outliers from the log-transformed emissions data below the 25th percentile and above the 75th percentile. For all log-transformed emissions categories, outliers fell below the 25th percentile, ranging between -2 and 2 on the $\log_{10}$ scale (i.e. between 0.01 and 100 MtCO2e); except for scope 3 emissions from purchased goods and services, where all outliers fell above the 75th percentile at 10,000 MtCO2e. Most outliers that were removed were therefore associated with particularly small reported emissions, with the exception of emissions in the purchased goods and services category. The HEI most commonly identified as an outlier was the University of Hawai'i Maui College; this HEI has a combined FTE of 1,492, and reports only 15 MtCO2e as its gross emissions.

To get a more normal distribution of emissions intensities, I log-transformed emissions data and removed outliers, and divided each resulting emissions observation by respective log-transformed FTE or GSF data. EIs for each category of emissions for all HEIs were therefore calculated according to the following equation:

$$log - transformed\ Emissions\ Intensity\ (EI) = \frac{log_{10}(emissions)}{log_{10}(standard\ unit)}$$

Next, I took EIs for each emissions category from all HEIs and selected the average value. Seeing as emissions were not initially normally distributed, I chose median as the average metric, since the median is less affected by outliers. Ultimately, I used the this median EI value to make emissions predictions within each emissions category, according to the following equation:

$$log_{10}(predicted\ emissions) = \left(\frac{log_{10}(emissions)}{log_{10}(standard\ unit)}\right)_{median} \cdot log_{10}(observed\ standard\ units)$$

Treating median EI as a slope parameter (θ_1), the above equation can be rewritten as follows, where $\hat{y}$ is predicted emissions, *x* is observed normalizing units, and θ_1 is median EI value:

$$log_{10}(\hat{y}) = \theta_1 \cdot log_{10}(x)$$

Finally, I visualized emissions predictions by plotting log-transformed emissions versus log-transformed standard units in a scatter plot, alongside median EI model predictions as a linear function, for each emissions category. I also calculated residuals, and visualized these in a scatter residual plot to further asses model fit.

I repeated this entire process twice, first using FTE, and then GSF as the standard unit.

Linear regression modeling

Simple Linear Regression (SLR)

As an alternative prediction method to median emissions intensity, I used linear regression models to predict emissions for each emissions category using my training data. The main difference between the SLR and EI approach is that an intercept parameter (θ_0) is fitted to the data, along with a slope parameter (θ_1). The EI analysis also contains a type of slope term, since median EI can be thought of as the slope as shown above; however, this value was calculated manually, and was not fit to the data to reduce sum of squared residuals, as it is for SLR. Seeing as log-transformed emissions versus log-transformed FTE or GSF seemed to best support a linear model in the exploratory data analysis, I used log-transformed emissions as the outcome variable, and log-transformed FTE and GSF data as my predictor variables. The equation used to make emissions predictions is the following, where $\hat{y}$ is predicted emissions, *x* is either observed FTE or GSF, and θ_0 and θ_1 are the fitted intercept and slope terms:

$$log_{10}(\hat{y}) = \theta_0 + \theta_1 \cdot log_{10}(x)$$

I plotted scatterplots and residuals for all trained SLR models that show the model predictions alongside log-transformed emissions and standard units data.

I repeated this entire process twice, first using FTE, and then GSF as the input data.

Multiple Linear Regression

For MLR, I included the following institutional characteristics as possible predictor variables: institution type/Carnegie classification, institutional control, medical school, hospital, satellite campuses, endowment size, FTE, GSF, floor area of laboratory space, and floor area of healthcare space. I manually checked null-values, and was able to correct one null value for the satellite campuses feature, where SUNY Oswego reported null despite having a satellite campus. The remaining null values came from three binary variables (medical school, hospital, satellite campuses) and one numerical feature (health care area), and were filled with zeros.

In order to prepare my data for MRL modeling, I converted categorical variables into binary features using One-Hot encoding. I also added log-transformed features of all my quantitative, continuous features (GSF, FTE, endowment size, floor area of laboratory space, and floor area health care space). Next, I standardized quantitative, continuous variables, by subtracting the mean and dividing by the standard deviation of each continuous feature variable. In order to mitigate collinearity between my variables, I explored correlations between features visually, and used variance inflation factor (VIF) to iteratively filter out highly correlated variables.

I used two subsequent feature selection approaches to find significant features and train model coefficients for each outcome emissions category. First, I selected significant features for each outcome emissions variable using p-values, and fed these into a forward selection algorithm to select features, and train ordinary least squares models. For this step, I split my training data into a training and validation set, and used the R^2 calculated for the validation set to select features. Next, I compared the results with models trained using LASSO regularization, which included 10-fold cross-validation for tuning the penalty-term hyperparameter. I compared the trained LASSO regression models with the forward selected models based on visual fit and R^2 calculated on training data. Lastly, I plotted actual versus predicted emissions data in scatterplots, where a hypothetical perfect model follows the line $y = x$, i.e. predicted emissions = actual emissions. For each emissions category, I was left with two MLR models with between 1 and n features included as predictor variables and an intercept term:

$$log_{10}(\hat{y}) = \theta_0 + \theta_1 \cdot x_1 + \cdots + \theta_n \cdot x_n$$

Model evaluation

To assess the internal validity of my models, I calculated emissions predictions and R^2 for my training data set. Next, I calculated emissions predictions and R^2 for the test data to assess external validity. I also calculated RMSE for all models first using my training data, and next using my test data. I used summary tables showing RMSE and R^2 as model performance metrics both for my training and test data to make final recommendations for the best prediction method for each emissions category.

RESULTS

Mapping emissions categories between STARS, SIMAP, and GHG Protocol

By mapping emissions categories between GHG Protocol, SIMAP, and STARS, I found that most emissions categories from the GHG Protocol framework are covered by the STARS data (Table 2). STARS documentation shows that multiple categories from SIMAP and the GHG Protocol were aggregated in the STARS framework, and cannot be represented individually. For example, scope 1 emissions from three separate categories, namely mobile combustion, process emissions, and fugitive emissions, according to the GHG Protocol, were combined into "Gross scope 1 GHG emissions from other sources (i.e. mobile combustion, process emissions, fugitive emissions)" in STARS. SIMAP has a higher granularity of emissions sources compared to both GHG Protocol and STARS, but highlights only very specific emissions sources within each GHG Protocol category. Meanwhile, the granularity of STARS is generally consistent with GHG Protocol categories, asides from mobile combustion, fugitive emissions, process emissions, purchased heating, purchased cooling, purchased steam, and employee commuting. In order to better align STARS with GHG Protocol, I combined "Scope 3 emissions from commuting" and "Scope 3 emissions from other sources not included in scope 1 or scope 2" into one category to align with GHG Protocol's employee commuting category. This aggregation was informed by the by 'Guidance for STARS participants using SIMAP' document provided by STARS (see Appendix A). I added a gross emissions category by summing total gross scope 1 emissions, total gross scope 2 emissions, and total gross scope 3 emissions from STARS. Overall, this left me with fourteen individual emissions categories.

GHG Protocol categories for emissions from upstream leased assets, downstream transportation and distribution, processing of goods sold, use of sold products, end of life treatment of sold products, downstream leased assets, franchises, upstream waste and distribution, and investments were not accounted for in STARS or SIMAP. Process emissions and capital goods were covered by STARS but not SIMAP, although process emissions are not entirely relevant to HEIs. Overall, of the 28 emissions categories from the GHG Protocol, seventeen are represented in SIMAP via 24 categories, while eighteen are represented in STARS via fourteen total categories.

Table 2. Mapping GHG emissions categories between GHG Protocol, SIMAP and STARS. SIMAP to STARS mapping informed by 'Guidance for STARS participants using SIMAP' document provided by STARS, SIMAP to GHG Protocol mapped manually. (a) Scope totals and gross emissions were mapped across frameworks; STARS has no explicit total/gross emissions category. (b) Four sub-scope categories from GHG Protocol for scope 1 covered in two categories by STARS. SIMAP does not cover process emissions. (c) Scope 2 purchased electricity covered by SIMAP and STARS, while purchased heating, cooling, and steam is aggregated. (d) Fifteen scope 3 sub-scope categories mapped, only five of which are covered by SIMAP and six by STARS.

(a)

GHG Protocol	SIMAP	STARS
Totals		
Total emissions	Gross Emissions	-
Scope 1 Total	Scope 1 Emissions	Total gross Scope 1 GHG emissions
Scope 2 total	Scope 2 Emissions	Total gross Scope 2 GHG emissions
Scope 3 Total	Scope 3 Emissions	Total gross Scope 3 emissions

(b)

<table>
<tr><th>GHG Protocol</th><th>SIMAP</th><th>STARS</th></tr>
<tr><th colspan="3">Scope 1</th></tr>
<tr><td rowspan="3">Stationary Combustion</td><td>Co-gen Electricity</td><td rowspan="3">Gross Scope 1 GHG emissions from stationary combustion</td></tr>
<tr><td>Co-gen Stem</td></tr>
<tr><td>Other on-campus Stationary</td></tr>
<tr><td>Mobile Combustion</td><td>Direct Transportation</td><td rowspan="4">Gross Scope 1 GHG emissions from other sources (i.e. mobile combustion, process emissions, fugitive emissions)</td></tr>
<tr><td rowspan="2">Fugitive Emissions</td><td>Refrigerants & Chemicals</td></tr>
<tr><td>Fertilizer & Animals</td></tr>
<tr><td>Process Emissions</td><td>-</td></tr>
</table>

(c)

GHG Protocol	SIMAP	STARS
Scope 2		
Purchased Electricity	Purchased Electricity	Gross Scope 2 GHG emissions from purchased electricity
Purchased Heating	Purchased Steam/Chilled Water	Gross Scope 2 GHG emissions from purchased heating and cooling
Purchased Cooling		
Purchased Steam		

(d)

GHG Protocol	SIMAP	STARS
Scope 3		
Purchased Goods and Services	Food	Scope 3 GHG emissions from purchased goods and services
	Paper Purchasing	
Waste	Solid Waste	Scope 3 GHG emissions from waste generated in operations
	Wastewater	
Business Travel	Study Abroad Air Travel	Scope 3 GHG emissions from business travel
	Directly Financed Air Travel	
	Other Directly Financed Travel	
Employee Commuting	Faculty Commuting	Scope 3 GHG emissions from commuting
	Staff Commuting	
	Student Commuting	
	Student Travel to/from Home	Scope 3 GHG emissions from other sources not included in Scope 1 or Scope 2
Fuel and Energy Related Activities	Trasnmission & Distribution Losses	Scope 3 GHG emissions from fuel- and energy-related activities not included in Scope 1 or Scope 2
Capital Goods	-	Scope 3 GHG Emissions from capital goods
Upstream Leased Assets	-	-
Downstream Transportation and Distribution	-	-
Processing of Goods Sold	-	-
Use of Sold Products	-	-
End of Life Treatment of Sold Products	-	-
Downstream Leased Assets	-	-
Franchises	-	-
Upstream Waste and Distribution	-	-
Investments	-	-

Data collection, exploration, and preparation

The complete STARS version 2.2 data had a total of 356 rows, each row corresponding to an individual STARS report submitted by an HEI. Since the launch of STARS version 2.2 in 2019, certain HEIs have submitted up to three reports, resulting in multiple rows corresponding to the same HEI. I filtered the data to get only the most recent report per institution, leaving 310 rows, one for each unique HEI included in the sample. Reporting years range from 2019 to 2023. The data represented HEIs from fourteen countries spread across six continents (Figure 2). There were 70 medical schools in the sample.

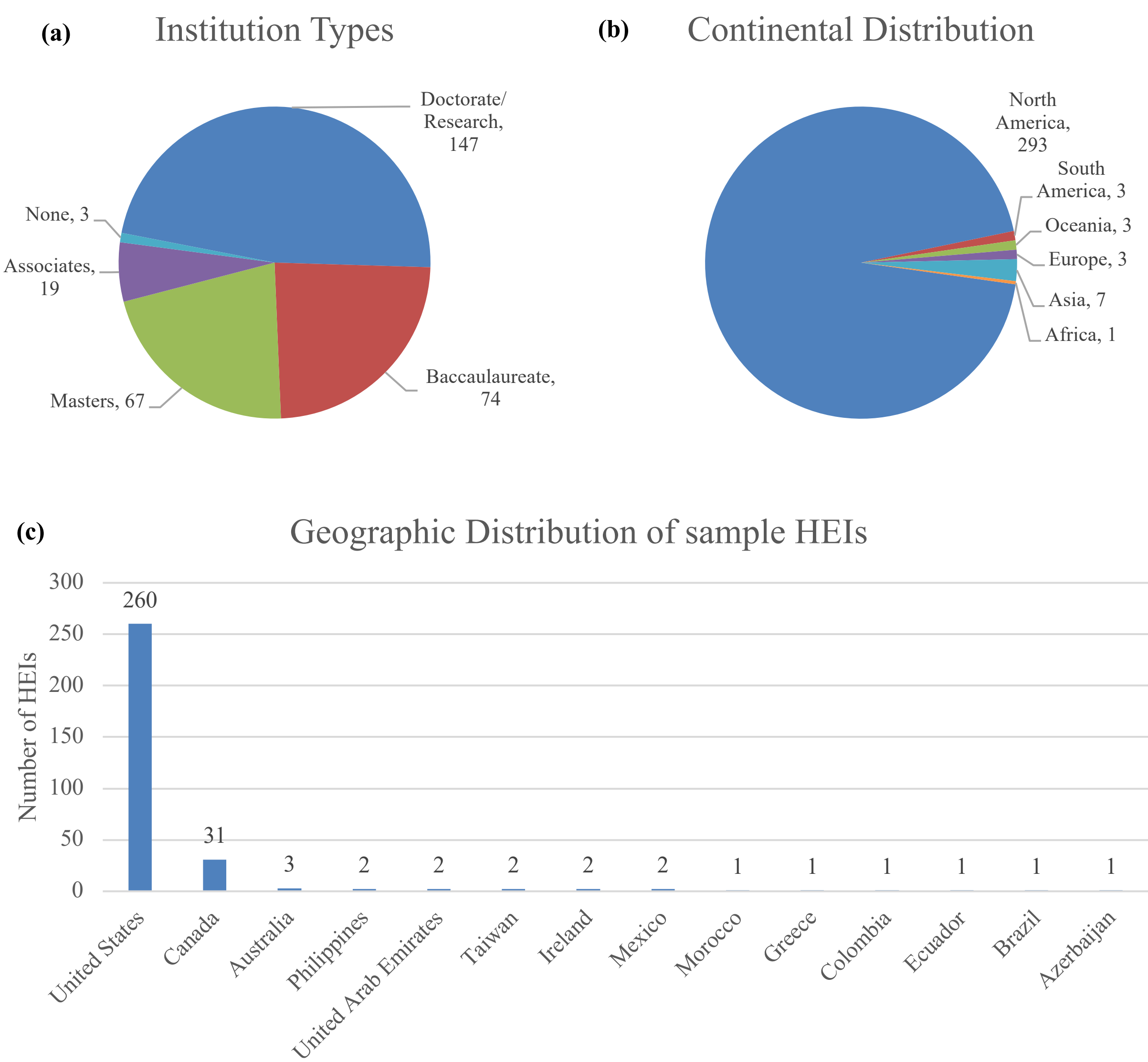

Figure 2. Analyzing HEIs included in STARS version 2.2 data sample. (a) Almost half (~ 47%) of all HEIs were Doctorate-granting/Research institutions, and close to a quarter were both Baccalaureate institutions (~ 24%) and Master's colleges (~ 22%). (b) Almost all HEIs (~ 95%) of HEIs in the sample were from North America, with the second largest group representing Asia, accounting for only around 2% of sample HEIs. (c) Out of 310 HEIs total, 260 are located in the United States and 31 are located in Canada, together comprising around 94% of the total sample.

The histograms and boxplots for the original emissions data, with zeros and nulls removed, showed that emissions distributions were mostly unimodal, but skewed significantly to the right.

The Shapiro-Wilk test for normality showed that emissions from no emissions category were normally distributed. After applying a $\log_{10}$ transformation, emissions distributions looked significantly more unimodal and symmetric, although for most categories, distributions were still somewhat skewed, this time mostly to the left. Only scope 2 emissions from imported thermal energy were normally distributed at this stage, according to the Shapiro-Wilk test (p-value > 0.05), despite the visual distribution being bimodal; with a sample size of only 39, normality according to the Shapiro-Wilk test does not seem reliable for this category. After removing outliers, all emissions categories were normally distributed according to the Shapiro-Wilk test, except scope 2 electricity, scope 2 total, scope 3 business travel, and scope 3 purchased goods and services.

Roughly the same trends were observed for emissions intensity distributions, with emissions normalized by FTE and GSF, where no emissions category had normally distributed data originally or after log-transforming, except scope 2 thermal. The skew and modality of the distributions again went from significantly skewed right, to being more symmetric with a slight left-skew. After removing outliers, skew was reduced, however not many emissions categories achieved normally distributed EI data. For EIs calculated using FTE, only scope 2 thermal and scope 3 fuel- and energy-related emissions were normally distributed according to the Shapiro-Wilk test; for EIs calculated with GSF, scope 1 emissions from other sources, scope 2 thermal, and scope 3 fuel- and energy-related emissions were normally distributed according to the Shapiro-Wilk test.

Seeing as log-transformed emissions data with outliers removed below the 25% percentile and above the 75% percentile were mostly normally distributed, emissions data for each category was transformed in this way and used in the remaining analysis. Although EI distributions were mostly not normal, the above set of transformations made the distributions appear the closest to normal visually. Therefore EIs were ultimately calculated by dividing log-transformed emissions with outliers removed, by log-transformed FTE or GSF data.

After removing zeros, nulls, and outliers, sample sizes varied across emissions categories (Table 3). The scope 3 category for capital goods only had eleven valid, non-zero observations and was therefore eliminated. Overall, thirteen emissions categories had training sample sizes sufficiently large ($n > 30$), and were included in modeling and subsequent analysis. Categories in scope 1 had the most observations, followed by scope 2, and finally scope 3. Aggregated categories (scope 1 total, scope 2 total, scope 3 total, and gross emissions) had among the most observations, and consistently had the greatest sample size within their respective scope.

Table 3. Final sample sizes. Out of an original sample size of 310, the following number of observations remained after removing zeros, nulls, and outliers. Valid data resulted from removing nulls and zeros; after additionally removing outliers, the remaining data was split into training and test sets for each emissions category. S3 Capital Goods was eliminated from analysis (n < 30). Note small sample sizes for S2 Thermal and S3 Goods and Services.

Emissions Categories	Valid Data	Outliers removed	Training Data	Test Data
S1 Stationary	278	8	229	41
S1 Other Sources	253	7	209	37
S1 Total	282	9	232	41
S2 Electricity	255	10	208	37
S2 Thermal	39	0	33	6
S2 Total	256	6	212	38
S3 Business Travel	183	1	154	28
S3 Commuting	135	2	113	20
S3 Goods and Services	85	6	67	12
S3 Capital Goods	11	-	-	-
S3 Fuel and Energy	107	2	89	16
S3 Waste	147	5	120	22
S3 Total	232	4	193	35
Gross Emissions	275	1	232	42

Average emissions intensities

Scatterplots of log-transformed emissions versus log-transformed FTE and GSF showed a positive association between the variables; while the relationship appeared linear, the spread of data points was relatively wide for certain categories. Median EI calculated using log-FTE seemed to predict $\log_{10}$(emissions) relatively well, with the predicted emissions line following general trend present in the data (Figure 3). R^2 was lowest for scope 2 categories, and highest among certain scope 3 categories; R^2 values ranged from 0.068 for scope 2 thermal emissions to 0.684 for scope 3 commuting emissions. The residual plots showed a relatively even scatter of points for all categories, although there seemed to be a slightly wider spread among negative residuals compared to positive residuals (Figure 4).

Median EI calculated by log-GSF predictions saw higher R^2 values for most emissions categories compared to median EI predictions using log-FTE. R^2 ranged between 0.105 for scope

3 goods and services, and 0.637 for gross emissions. Visually, the predicted values did not follow the general trend in the data (Figure 5). The residual plots clearly show that median EI by log-GSF tended to overpredict emissions for lower values of GSF, and underpredict emissions for higher values of GSF. This trend can be observed for almost every emissions category (Figure 6).

Generally, EIs were higher for FTE than for GSF, i.e. there were more emissions per full-time equivalent student or employee than per square foot of building space. This can logically be explained as FTE values ranged between 330 and 114,870 individuals, while GSF ranged between 106,223 and 49,663,410 square feet. Given the larger denominator for GSF, EIs were smaller compared to EIs for FTE. Median log-emissions per log-FTE or log-GSF units are summarized for all thirteen emissions categories in Table 4.

Table 4. Median emissions intensities. A log-transformed emissions intensity, with $\log_{10}$(emissions) divided by $\log_{10}$(FTE) or $\log_{10}$(GSF), was calculated for HEIs in the sample. The median values are captured below. Median intensities are denoted by θ_1, since they can be thought of as a slope term in a model of the form $y = \theta_1 \cdot x$.

Emissions Category	FTE θ_1	GSF θ_1
S1 Stationary	1.004	0.606
S1 Other Sources	0.685	0.413
S1 Total	0.996	0.606
S2 Electricity	0.984	0.606
S2 Thermal	0.81	0.529
S2 Total	0.987	0.607
S3 Business Travel	0.827	0.512
S3 All Commuting	0.931	0.578
S3 Goods and Services	0.489	0.309
S3 Fuel and Energy	0.749	0.462
S3 Waste	0.635	0.395
S3 Total	0.986	0.604
Gross Emissions	1.116	0.684

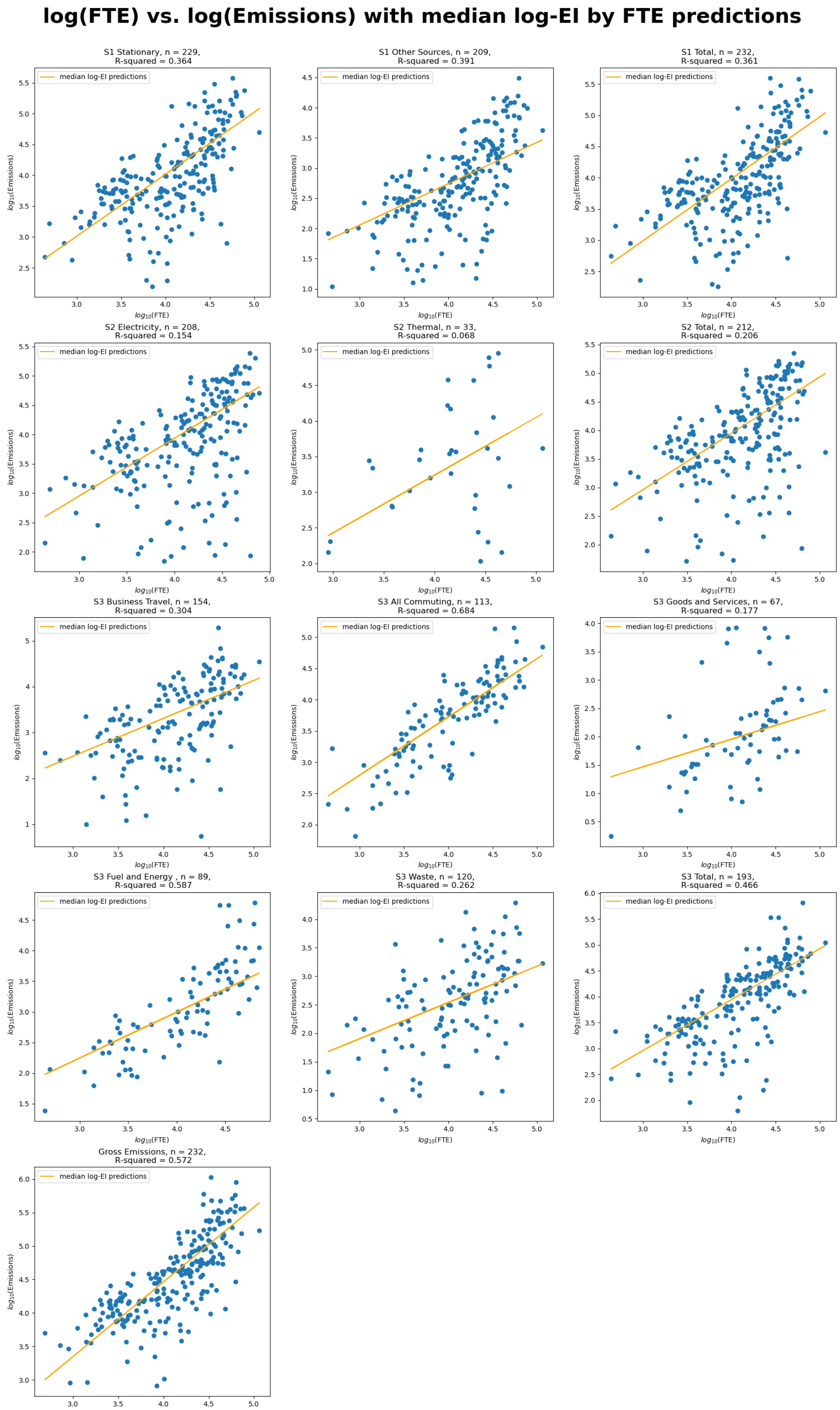

Figure 3. Scatterplots of $\log_{10}$(FTE) vs. $\log_{10}$(emissions) data alongside predictions made using median log-emissions per log-FTE unit. Good fit between predicted and actual emissions data. Positive linear trend seems to intensify with greater sample size.

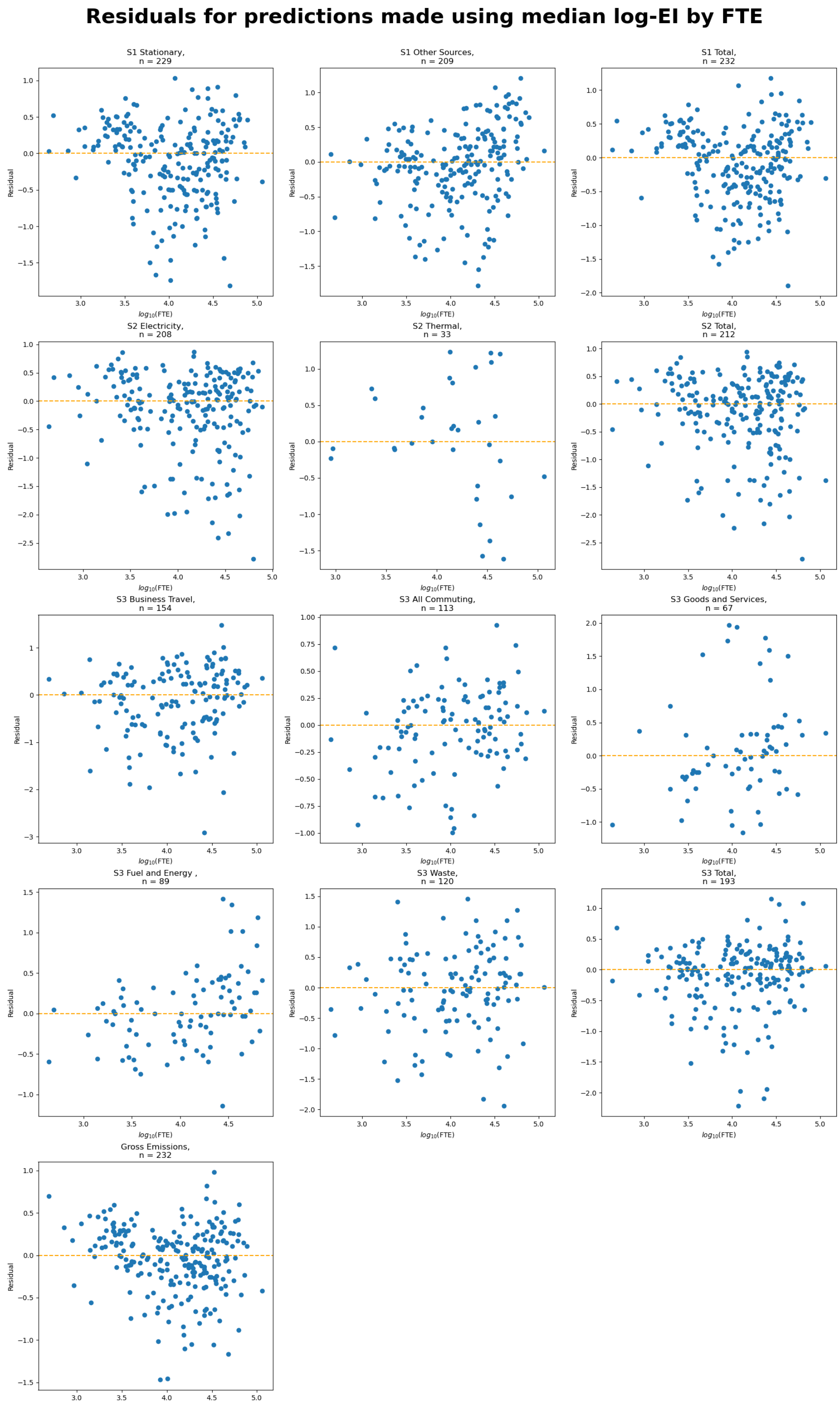

Figure 4. Residuals for predictions made using median log-emissions per log-FTE unit. Relatively even scatter of residuals for all categories, with no distinct, non-linear trends. Wider spread of negative residuals compared to positive residuals for most categories.

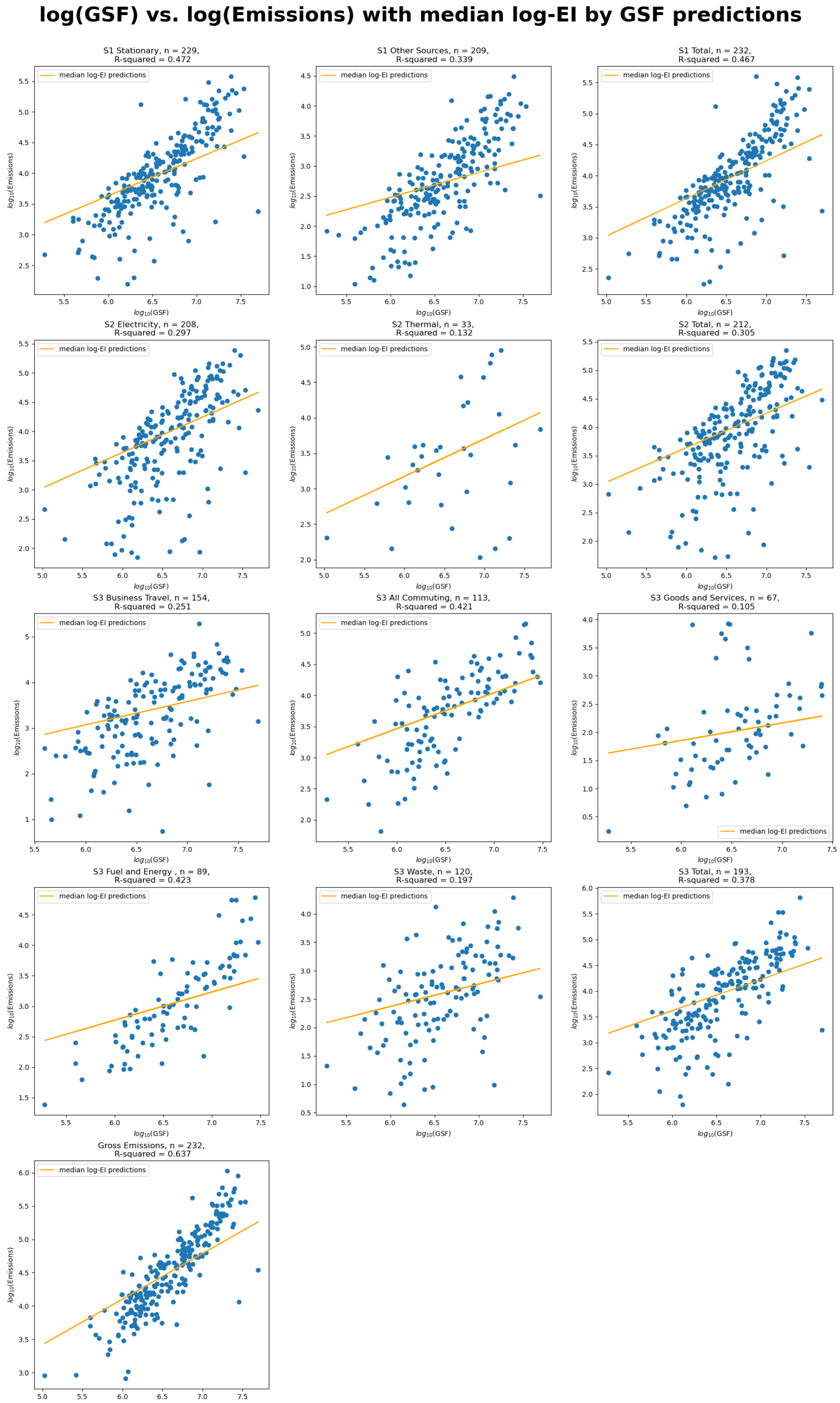

Figure 5. Scatterplots of $\log_{10}$(GSF) vs. $\log_{10}$(emissions) data alongside predictions made using median log-emissions per log-GSF. Poorer fit compared to median EI by log-FTE. Prediction line seems to cut through data cloud unevenly. Visually, we start to see tighter linear correlations between $\log_{10}$(GSF) and $\log_{10}$(emissions), than between $\log_{10}$(FTE) and $\log_{10}$(emissions).

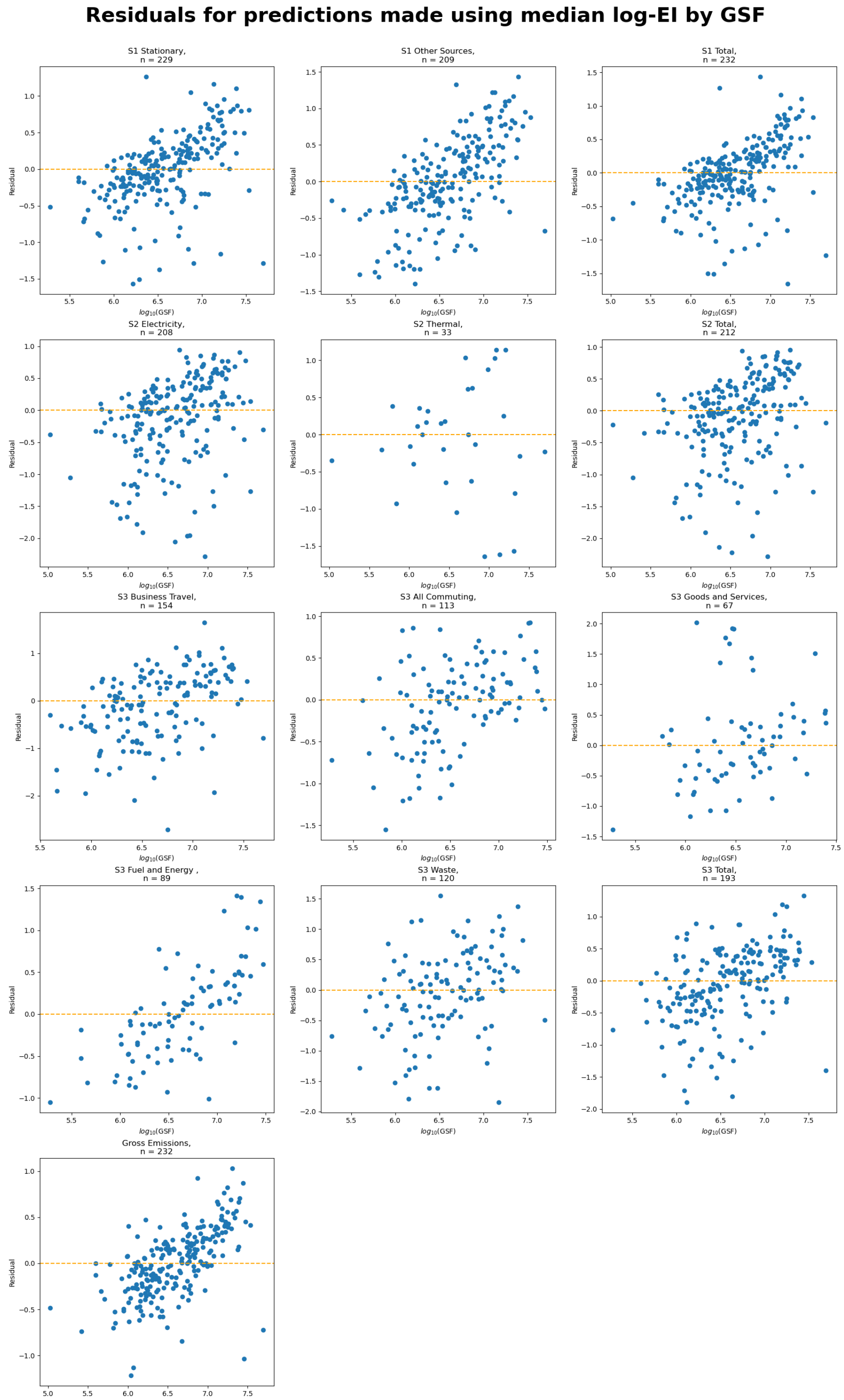

Figure 6. Residuals for predictions made using median log-emissions per log-GSF unit. Poor residual plots, with systematic over- and underprediction depending on size of GSF.

Linear regression modeling

Simple Linear Regression (SLR)

Simple linear regression (SLR) models for log-emissions predicted by log-FTE and log-GSF showed generally good fit to the data. For models using log-FTE as the predictor variable, scope 2 categories saw the poorest fit, as was observed in the median EI analysis using log-FTE (Figure 7). R^2 values ranged from 0.103 for scope 2 thermal, to 0.700 for scope 3 commuting. These were the same low and high categories as observed in the median EI analysis using log-FTE; however, the R^2 value for both showed improvement from the median EI approach. Other relatively high R^2 values were observed for scope 3 fuel- and energy related emissions and gross emissions, at 0.652 and 0.590 respectively. The residual plots showed a good, even scatter of residuals (Figure 8).

Using log-GSF as the predictor in the SLR models resulted in a particularly good fit, much improved from median EI using log-GSF, and for most categories, also improved fit compared to log-FTE predictions (Figure 9). The only categories where the SRL model using log-FTE predicted log-emissions better than log-GSF was emissions for scope 3 commuting and scope 3 goods and services. R^2 values ranged from 0.141 for scope 2 thermal, to 0.774 for gross emissions. Several other emissions categories saw relatively high R^2 values, including scope 3 fuel- and energy-related emissions at 0.716, scope 1 stationary at 0.610, scope 1 other sources at 0.600, scope 1 total at 0.590, scope 3 commuting at 0.551, and scope 3 total at 0.503. Residual plots for SLR models using log-GSF also saw a significant improvement from the median EI analysis using log-GSF. For the SLR models, residuals are scattered evenly and show no distinct pattern. (Figure 10). Intercept and slope terms for models fitted to the training data are summarized in Table 5.

Table 5. Intercept and slope terms for SLR models. The SLR models predict $\log_{10}$(emissions) data using $\log_{10}$(FTE) and $\log_{10}$(GSF) as the input data. All slopes were positive, showing that emissions increase with FTE or GSF. θ_0 is the intercept term, and θ_1 is the slope term in the equation $\hat{y} = \theta_0 + \theta_1 \cdot x$.

Emissions Category	FTE θ_0	FTE θ_1	GSF θ_0	GSF θ_1
S1 Stationary	0.457	0.867	-3.605	1.153
S1 Other Sources	-1.145	0.954	-5.244	1.21
S1 Total	0.376	0.882	-3.353	1.115
S2 Electricity	0.713	0.766	-3.264	1.081
S2 Thermal	1.251	0.52	-0.021	0.521
S2 Total	0.565	0.814	-3.031	1.052
S3 Business Travel	-0.907	1.018	-4.798	1.217
S3 All Commuting	-0.632	1.079	-3.394	1.086
S3 Goods and Services	-1.338	0.839	-3.307	0.825
S3 Fuel and Energy	-1.324	1.085	-5.419	1.281
S3 Waste	-0.664	0.792	-3.481	0.917
S3 Total	-0.429	1.064	-3.72	1.159
Gross Emissions	0.536	0.974	-3.254	1.179

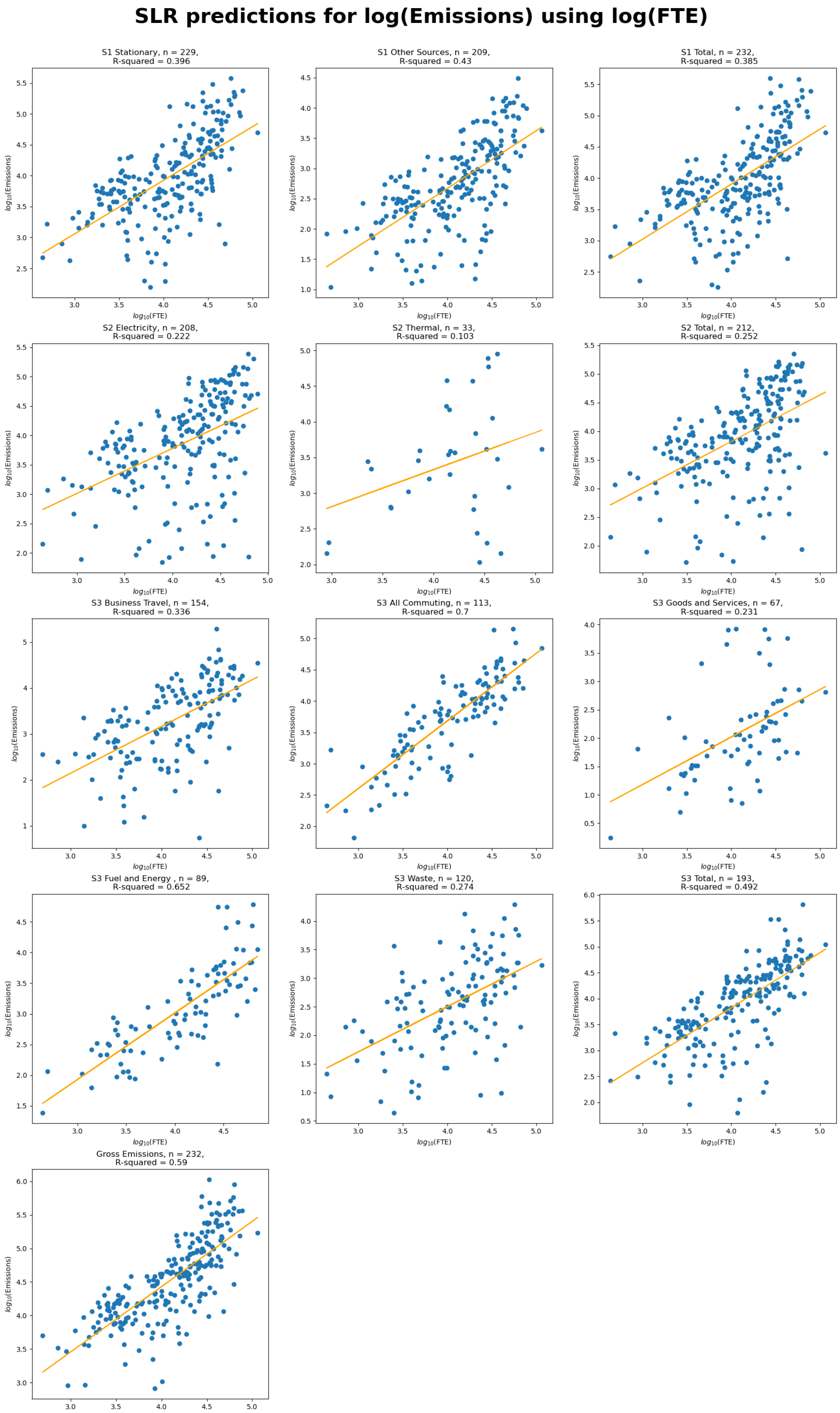

Figure 7. Scatterplots of $\log_{10}$(FTE) vs. $\log_{10}$(emissions) data alongside SLR model predictions with log-FTE input. Good fit between predicted and actual emissions data. Wider spread of data points compared to $\log_{10}$(GSF) vs. $\log_{10}$(emissions).

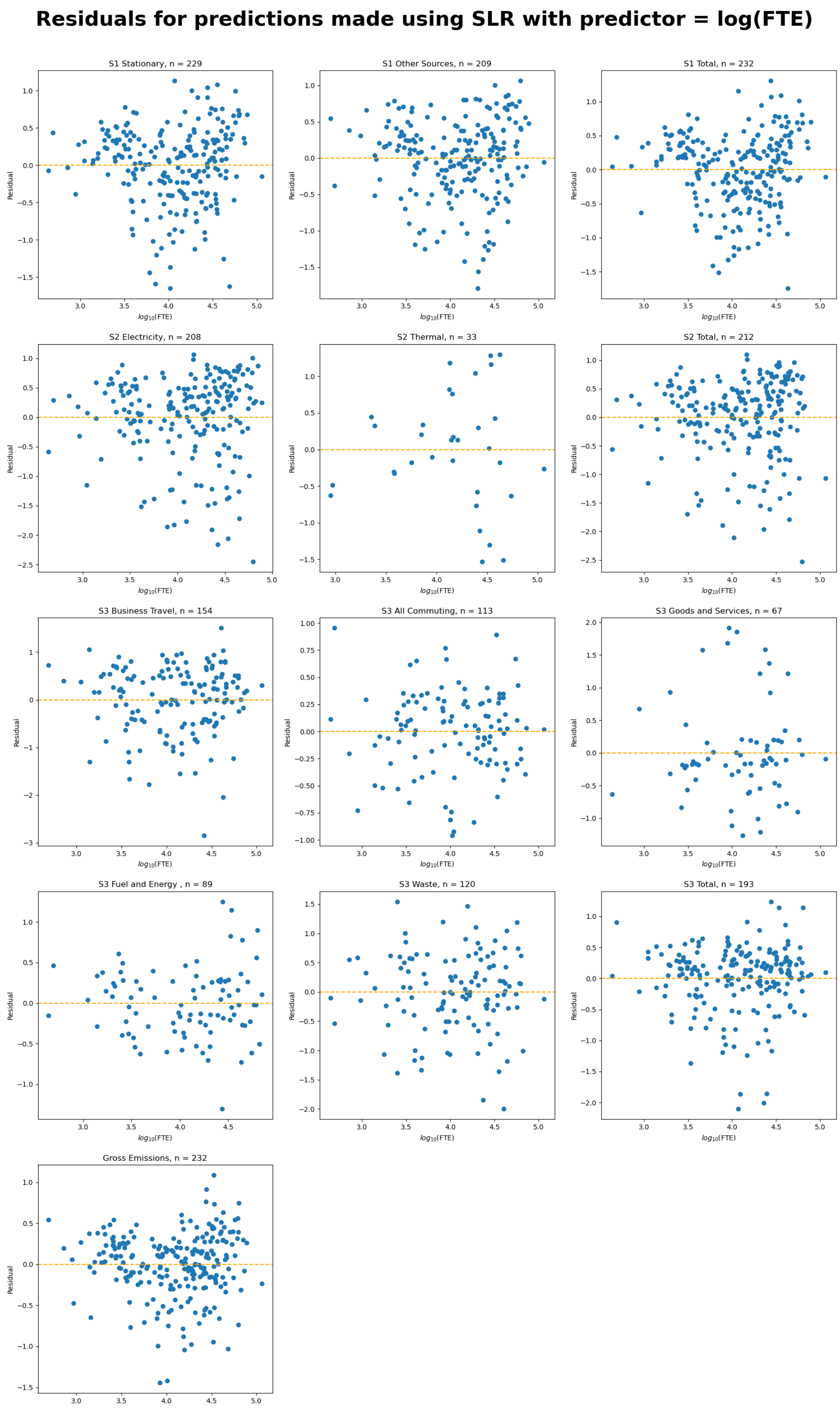

Figure 8. Residuals for predictions made using SLR model with log-FTE input. Relatively even scatter of residuals for all categories, with no distinct, non-linear trends.

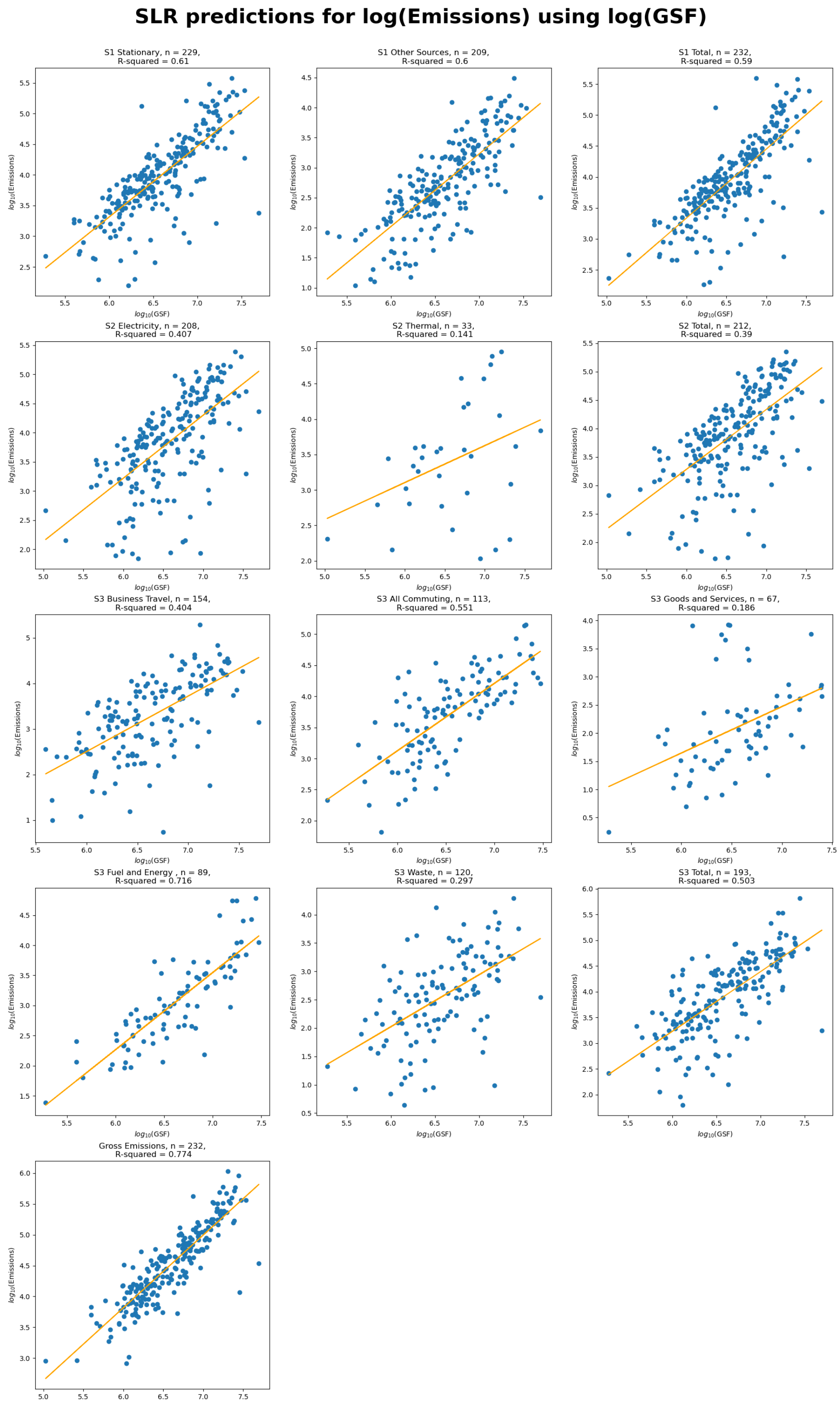

Figure 9. Scatterplots of $\log_{10}$(GSF) vs. $\log_{10}$(emissions) data alongside SLR model predictions with log-GSF input. Tighter fit of actual data around prediction line. Improved visual fit compared to median EI using log-GSF, and tighter fit compared to log-FTE predicted data.

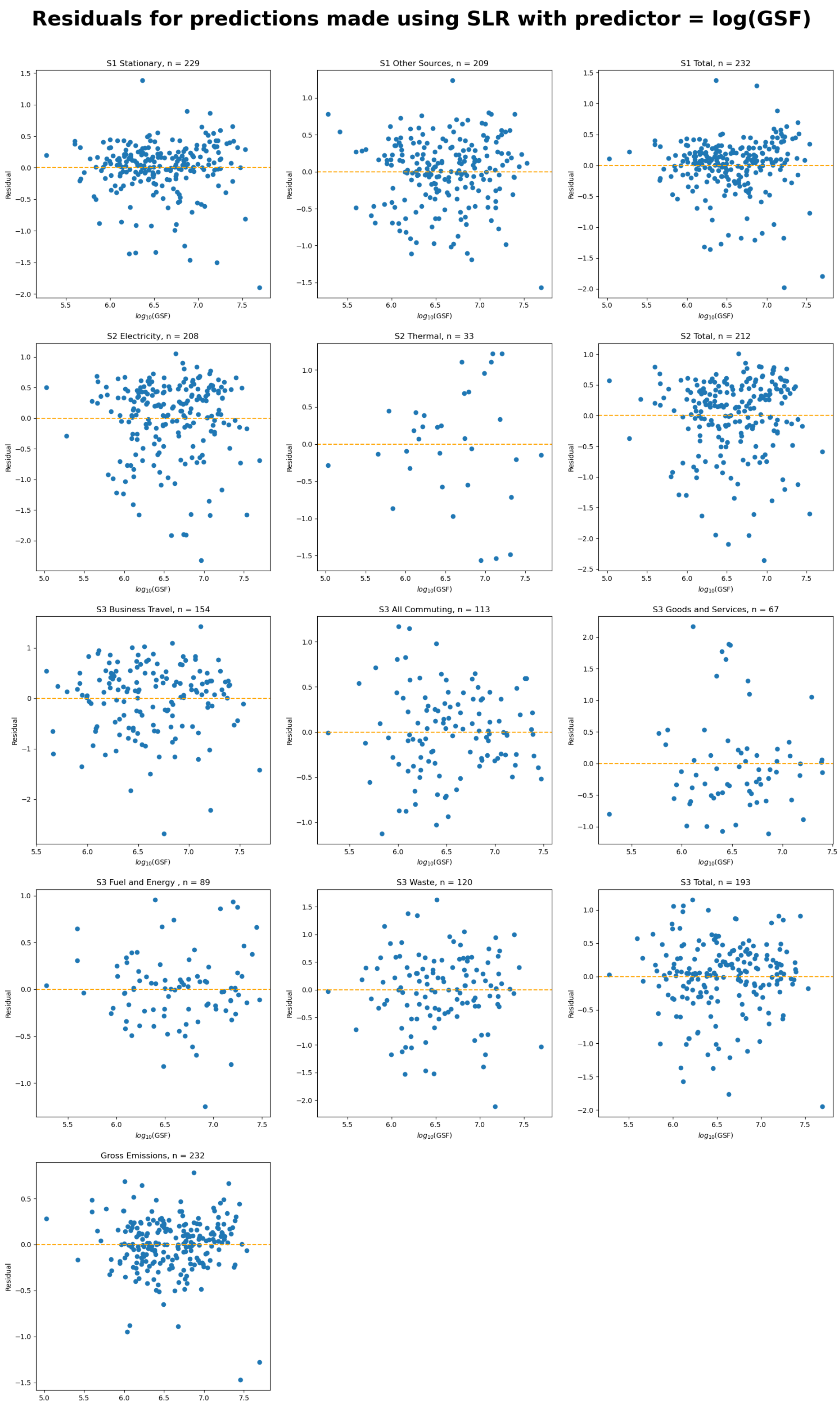

Figure 10. Residuals for predictions made using SLR model with log-GSF input. Even scatter of residuals, with smaller spread compared to median EI and SLR using log-FTE residual plots.

Multiple Linear Regression

A total of eighteen features were considered for multiple linear regression (MLR) modeling, including eight one-hot encoded features: medical school, hospital, satellite campuses, Associate's College, Baccalaureate college, Master's university, Doctorate-granting/Research university, and public institution. Ten quantitative, continuous features were considered and standardized, including FTE, $\log_{10}$(FTE), GSF, $\log_{10}$(GSF), endowment size, $\log_{10}$(endowment size), laboratory area, $\log_{10}$(lab area), health care area, and $\log_{10}$(health care area).

The VIF analysis showed that FTE, GSF, log-FTE, and log-GSF are very highly correlated amongst each other, with correlation coefficients around 0.8 and 0.9. Therefore, it was not possible to keep both FTE and GSF features, or standard unit and log-transformed versions of those features, in the model at the same time without subjecting the model to high collinearity. Seeing as log-GSF had the best predictive power for most emissions categories in the SLR analysis, I prioritized keeping log-GSF as a feature. Iteratively filtering out highly collinear features resulted in the following features being eliminated: log-FTE, FTE, GSF, public institution, and lab area. Thirteen features remained and were used in feature selection and model training.

Overall, LASSO regression showed improved model fit compared to forward selection models across emissions categories. For simplicity's sake, only the models resulting from LASSO regression will be presented and discussed (Table 10). The number of features selected varied across emissions categories, ranging between one and eleven features; scope 1 stationary had eleven features plus an intercept term, while scope 3 goods and services had one feature (log-GSF) plus an intercept term. The binary feature for hospital was never assigned a non-zero coefficient for any model, while log-GSF was included in every model as a non-zero coefficient feature.

The LASSO regression models showed a very good fit between predicted and actual data (Figure 11). Scope 2 thermal (R^2 = 0.230) and scope 3 goods and services (R^2 = 0.169) saw the poorest model fit, likely due to the small sample sizes; these two emissions categories had the fewest observed data points, with training sample sizes of 33 for scope 2 thermal and 67 for scope 3 goods and services. Gross emissions saw the highest R^2 value at 0.798 – this was the best model performance observed in the entire analysis. Other categories also saw very high R^2 values, including scope 3 commuting at 0.720, scope 3 fuel- and energy-related emissions at 0.738.

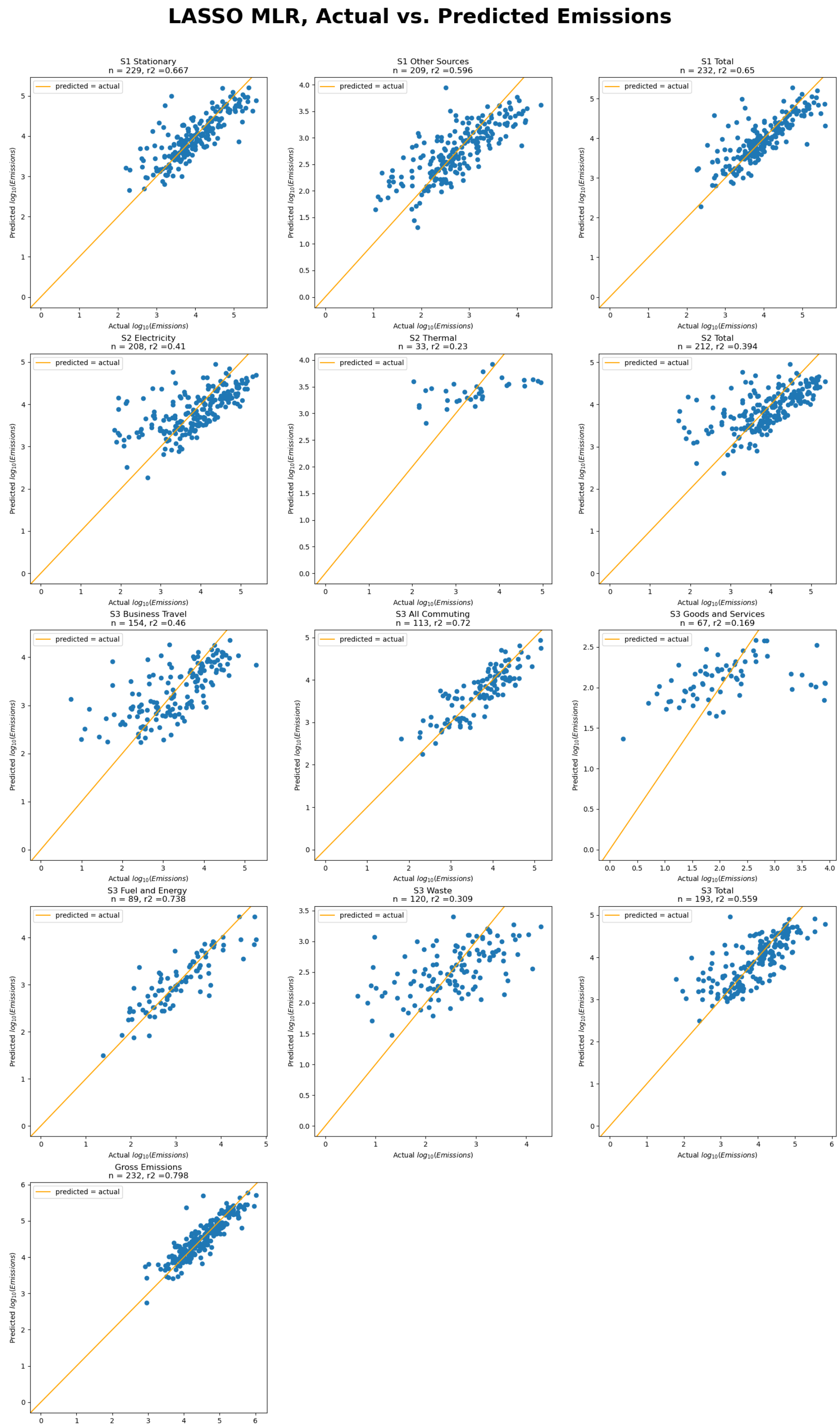

Figure 11. Actual vs. predicted $\log_{10}$(emissions) from LASSO regression. All emissions categories, except scope 2 thermal and scope 3 goods and services see a good fit between the actual and predicted data.

Model evaluation

Model performance for the training data showed the same trends across both R^2 and RMSE performance metrics. For all emissions categories except scope 3 goods and services and scope 1 other sources, MLR LASSO model predictions had the highest R^2 and lowest RMSE scores (Table 6, 7). The SLR model using GSF typically achieved the second highest performance score and demonstrated very similar model performance to the MLR LASSO model, especially for scope 2 electricity (MLR: $R^2 = 0.410$, SLR GSF: $R^2 = 0.407$), scope 2 total (MLR: $R^2 = 0.394$, SLR GSF: $R^2 = 0.390$), and scope 3 waste (MLR: $R^2 = 0.309$, SLR GSF: $R^2 = 0.297$). The exception is scope 3 goods and services, for which the SLR model using FTE achieved the best performance ($R^2 = 0.231$, RMSE = 0.721); however, with an R^2 of only 0.231, this category seems difficult to accurately predict using linear models in general. For scope 1 other sources, SLR using GSF performed best (SLR GSF: $R^2 = 0.600$), closely followed by MLR (MLR: $R^2 = 0.596$). The categories with the highest training accuracy were gross emissions (MLR: $R^2 = 0.798$), scope 3 fuel and energy (MLR: $R^2 = 0.738$), and scope 3 commuting (MLR: $R^2 = 0.720$). The categories with the lowest training accuracy were scope 2 thermal (MLR: $R^2 = 0.230$), scope 3 goods and services (MLR: $R^2 = 0.231$), and scope 3 waste (MLR: $R^2 = 0.309$).

Overall, model performance using training data shows that the highest accuracy predictions can be achieved using the MLR LASSO model for almost all categories; however, using the SLR GSF model gives similar accuracy, while requiring only one data input. As was noted in the scatterplots previously, $\log_{10}$(GSF) versus $\log_{10}$(emissions) showed a tighter linear trend in the data compared to $\log_{10}$(FTE). This is reaffirmed by the R^2 scores, where predictions made using GSF typically outperformed their FTE counterpart, except for scope 3 commuting and scope 3 goods and services.

Table 6. Model performance measured by R^2 for training data. R^2 ranges between 0 and 1, and represents the proportion of variance in emissions explained by model. Desirable scores are as close to 1 as possible. Here, highest scores within each category are bolded and marked with an asterisk. Training sample size for each category is summarized in the n column. Note similarities between MLR LASSO and SLR GSF performance scores.

Emissions Category	n	Median EI FTE	Median EI GSF	SLR FTE	SLR GSF	MLR LASSO
S1 Stationary	229	0.364	0.472	0.396	0.610	**0.667***
S1 Other Sources	209	0.391	0.339	0.430	**0.600***	0.596
S1 Total	232	0.361	0.467	0.385	0.590	**0.650***
S2 Electricity	208	0.154	0.297	0.222	0.407	**0.410***
S2 Thermal	33	0.068	0.132	0.103	0.141	**0.230***
S2 Total	212	0.206	0.305	0.252	0.390	**0.394***
S3 Business Travel	154	0.304	0.251	0.336	0.404	**0.460***
S3 All Commuting	113	0.684	0.421	0.700	0.551	**0.720***
S3 Goods and Services	67	0.177	0.105	**0.231***	0.186	0.169
S3 Fuel and Energy	89	0.587	0.423	0.652	0.716	**0.738***
S3 Waste	120	0.262	0.197	0.274	0.297	**0.309***
S3 Total	193	0.466	0.378	0.492	0.503	**0.559***
Gross Emissions	232	0.572	0.637	0.590	0.774	**0.798***

Table 7. Model performance measured by RMSE for training data. RMSE scale is relative and is given in units of $\log_{10}$(emissions). Generally, lower RMSE means better model performance. Lowest scores per category are bolded and marked with and asterisk. Trends in model performance are consistent between RMSE and R^2 for the training data. Note similar scores between SLR GSF and MLR LASSO.

Emissions Category	n	Median EI FTE	Median EI GSF	SLR FTE	SLR GSF	MLR LASSO
S1 Stationary	229	0.535	0.487	0.521	0.419	**0.387***
S1 Other Sources	209	0.556	0.579	0.538	**0.451***	0.453
S1 Total	232	0.535	0.489	0.525	0.429	**0.396***
S2 Electricity	208	0.734	0.669	0.704	0.615	**0.613***
S2 Thermal	33	0.776	0.748	0.761	0.745	**0.705***
S2 Total	212	0.699	0.654	0.679	0.613	**0.611***
S3 Business Travel	154	0.721	0.748	0.704	0.667	**0.635***
S3 All Commuting	113	0.381	0.515	0.371	0.454	**0.359***
S3 Goods and Services	67	0.745	0.777	**0.721***	0.741	0.749
S3 Fuel and Energy	89	0.463	0.548	0.425	0.385	**0.369***
S3 Waste	120	0.669	0.698	0.664	0.653	**0.648***
S3 Total	193	0.548	0.592	0.535	0.529	**0.498***
Gross Emissions	232	0.399	0.368	0.391	0.290	**0.275***

The generalizability of the models is estimated by measuring model performance using test data. This way, we see how well the model performs on data it has not "seen" during training. I observed different trends in model performance using the test data compared to the training data. Overall, MLR LASSO performed best for four categories, SLF GSF for five categories, and SLR FTE for two categories, again including scope 3 goods and services (Table 8, 9). Therefore, in the test data, both SLR models competed more with MLR model performance than in the training data results. For several categories, MLR LASSO scores were very similar to SLR GSF scores, especially for scope 1 total (MLR: R^2 = 0.670, SLR GSF: R^2 = 0.667), scope 2 electricity (MLR: R^2 = 0.564, SLR GSF: R^2 = 0.562), scope 2 total (MLR: R^2 = 0.510, SLR GSF: R^2 = 0.515), scope 3 commuting (MLR: R^2 = 0.772, SLR GSF: R^2 = 0.779), and gross emissions (MLR: R^2 = 0.765, SLR GSF: R^2 = 0.770).

Scope 2 thermal and scope 3 business travel observed negative R^2 values for some or all models. Although negative R^2 is uncommon, it is in fact possible. A negative R^2 score can occur when the model predictions fit the data worse than the null model, or the mean of the target variable. Therefore, I cannot be certain that my models are reliable for scope 2 thermal emissions or scope 3 business travel. While two non-negative R^2 scores were observed for scope 3 business travel, the highest positive R^2 was 0.118 from median EI using GSF; not only is this prediction accuracy very low, but predictions made using median EI using GSF fit the actual data very poorly, as was observed previously in the analysis of scatterplots and residuals. Seeing as median EI using FTE never showed the highest model performance scores, regression modeling is the more favorable modeling approach compared to median EIs.

The categories with the highest testing accuracy were scope 3 commuting (SLR GSF: R^2 = 0.779), gross emissions (SLR GSF: R^2 = 0.770), scope 1 stationary (MLR: R^2 = 0.736). The categories with the lowest training accuracy were scope 3 business travel (median EI by GSF: R^2 = 0.118), scope 1 stationary (MLR: R^2 = 0.222), and scope 3 waste (MLR: R^2 = 0.398).

Table 8. Model performance measured by R^2 for test data. Good model performance is evidenced by R^2 close to 1. Model performance using test data shows a greater variety of best performing models. R^2 from best model was higher for test data than training data for scope 1 stationary, scope 1 total, scope 2 electricity, scope 2 total, scope 3 commuting, scope 3 goods and services, scope 3 waste, and scope 3 total. Lower R^2 was observed for scope 1 other sources, scope 2 thermal, scope 3 business travel, scope 3 fuel- and energy-related emissions, and gross emissions.

Emissions Category	n	Median EI FTE	Median EI GSF	SLR FTE	SLR GSF	MLR LASSO
S1 Stationary	41	0.218	0.467	0.306	0.612	**0.736***
S1 Other Sources	37	0.110	0.215	0.046	0.181	**0.222***
S1 Total	41	0.474	0.490	0.411	0.667	**0.670***
S2 Electricity	37	0.504	0.436	0.415	0.562	**0.564***
S2 Thermal	6	-0.173	-0.062	-0.118	-0.075	-0.056
S2 Total	38	0.260	0.344	0.296	**0.515***	0.510
S3 Business Travel	28	0.042	**0.118***	-0.049	-0.006	-0.075
S3 All Commuting	20	0.760	0.540	0.778	**0.779***	0.772
S3 Goods and Services	12	0.270	0.171	**0.441***	0.427	0.322
S3 Fuel and Energy	16	0.432	0.146	**0.470***	0.320	0.340
S3 Waste	22	0.337	0.246	0.343	**0.398***	0.379
S3 Total	35	0.557	0.460	0.558	**0.651***	0.632
Gross Emissions	42	0.457	0.580	0.482	**0.770***	0.765

Table 9. Model performance measured by RMSE for test data. Trends in best performing model are generally consistent with R^2 trends. Median EI using GSF should be disregarded due to poor visual fit.

Emissions Category	n	Median EI FTE	Median EI GSF	SLR FTE	SLR GSF	MLR LASSO
S1 Stationary	41	0.578	0.478	0.545	0.407	**0.336***
S1 Other Sources	37	0.662	0.621	0.685	0.634	**0.618***
S1 Total	41	0.435	0.428	0.460	0.346	**0.344***
S2 Electricity	37	0.582	0.621	0.632	0.547	**0.546***
S2 Thermal	6	0.814	0.774	0.794	0.779	**0.772***
S2 Total	38	0.811	0.764	0.791	**0.657***	0.660
S3 Business Travel	28	0.746	**0.716***	0.781	0.764	0.790
S3 All Commuting	20	0.309	0.428	**0.297***	**0.297***	0.301
S3 Goods and Services	12	0.869	0.927	**0.761***	0.770	0.838
S3 Fuel and Energy	16	0.519	0.636	**0.501***	0.567	0.559
S3 Waste	22	0.511	0.545	0.509	**0.487***	0.494
S3 Total	35	0.485	0.536	0.485	**0.431***	0.442
Gross Emissions	42	0.524	0.461	0.511	**0.341***	0.344

Table 10. MLR model parameters. Model parameters trained using LASSO regularization, which drives coefficient sizes of less significant features to zero in order to minimize model complexity. Data was standardized, therefore coefficient sizes represent relative importance.

Emissions Category	Intercept	Medical School	Satellite Campuses	Hospital	Endowment Size	Health Care Area	log GSF	log Endowment Size	log Lab Area	log Health Care Area	Institution Type Associate	Institution Type Baccalaureate	Institution Type Doctorate	Institution Type Master
S1 Stationary	3.982	-0.146	-0.067	0	0.019	0.043	0.429	0.136	0.074	0.049	0.197	-0.018	0	-0.107
S1 Other Sources	2.671	0	0	0	0	0	0.51	0	0.015	0	0	0	0	0
S1 Total	3.97	-0.087	-0.053	0	0.025	0.055	0.399	0.12	0.092	0.022	0.139	0	0	-0.11
S2 Electricity	3.798	0	0	0	0	0	0.463	0	0	0.044	0	0	0	0
S2 Thermal	3.41	0	0	0	-0.037	-0.101	0.198	0	0	0	0	0	0	0
S2 Total	3.849	0	0	0	0	0.001	0.443	0	0	0.045	0	0	0	0
S3 Business Travel	2.99	0	0	0	0.032	0.013	0.349	0.096	0.002	0	0	0	0.378	0
S3 All Commuting	3.768	0	0.076	0	0.042	0.079	0.43	-0.093	0.041	-0.072	0.211	-0.424	0	-0.019
S3 Goods and Services	2.088	0	0	0	0	0	0.271	0	0	0	0	0	0	0
S3 Fuel and Energy	2.952	0	0	0	0.069	0	0.536	0	0	0	0	-0.014	0	0
S3 Waste	2.517	0	0	0	0	0	0.343	0	0	0.091	0	0	0	0
S3 Total	3.876	0	0.051	0	0.025	0.014	0.377	0	0	0.038	0	-0.322	0.101	0
Gross Emissions	4.43	0	0	0	0.046	0.008	0.488	0.045	0.022	0.025	0.169	-0.029	0.016	0

DISCUSSION

This study presents a novel contribution to research on HEI GHG emissions by training and evaluating prediction models for sub-scope categories of emissions defined by the GHG Protocol; previous studies have focused only on scope totals or gross emissions. Furthermore, the mapping between the GHG Protocol, SIMAP, and STARS has not been seen or done before in the literature or by the respective organizations, and may offer novel insight on how these HEI-specific frameworks align with the GHG Protocol. The key findings of this study are that linear regression models produce more reliable emissions estimates compared to median emissions intensity. While multiple linear regression models can typically achieve the highest accuracy, simple linear regression models using GSF or FTE as data inputs can achieve similar accuracy while requiring only one input data point. The tradeoff between data input requirements, prediction accuracy, and model interpretability is central to my research.

Average emissions intensities

Right° suggested average emissions intensity as a prediction method, seeing as it is a highly interpretable approach that is straight-forward and requires little computing power. The model performance of median emissions intensities using FTE as the normalizing factor could not compete with other models analyzed in this study, although visually the predictions followed the general trend in the data. For some categories, using GSF as the normalizing factor produced better model performance compared to FTE in terms of R^2 and RMSE; however, visually, the predictions made using the GSF median EI fit the general trend in the data relatively poorly. Therefore, neither median EI model can be recommended as a

Thus far, average emissions intensity remains largely unexplored as a prediction method in the literature. Sinha et al. (2010) is a rare example where average EI was used for prediction purposes. Further metrics could be explored as normalizing factors for making predictions using average EIs, such as unit expenditure (Larsen et al. 2013, Helmers et al. 2021); other economic factors, like endowment size, have not yet been studied as normalizing factors for EIs at all. The approach used in this study, of log-transforming the numerator and denominator for EIs has also not been observed previously, and could be further investigated.

Linear regression modeling

One important finding from this research is that SLR using GSF as the predictor variable produced very similar training accuracy to the MLR model. Previous studies have also used regression modeling approaches to predict GHG emissions from HEIs. Klein-Banai and Theis (2013) trained an SLR model to predict $\log_{10}$(gross emissions) using log-GSF data as the input, and achieved an R^2 of 0.795; my SLR model for log-gross emissions using log-GSF as the predictor achieved a training and testing R^2 of 0.774 and 0.770 respectively, which is comparable to the prediction accuracy achieved by Klein-Banai and Theis (2013). For all emissions categories, there was a positive slope for FTE and GSF SLR models, which underlines the positive association between emissions and institution size measured by GSF or FTE. Additionally, the log-transformation of the variables that is necessary to achieve a linear relationship aligns with the findings from previous studies and underscores the power-law nature of these relationships (Fetcher 2009, Klein-Banai and Theis 2013, Wadud et al. 2019). Additional quantitative features could be analyzed as input features to SLR models, such as expenditure or endowment size.

Like Klein-Banai and Theis (2013) and Fetcher (2009), I found GSF to be the better predictor of GHG emissions from HEIs compared to FTE. SLR GSF models predicted better than SLR FTE models for most emissions categories. Log-GSF was selected as a prediction feature for all MLR models, and log-GSF had the first or second greatest coefficient value in all MLR models – seeing as continuous features were standardized for MLR, the size of the coefficients indicate relative importance. Limiting the expansion of building spaces of HEIs will be important to mitigate future GHG emissions from HEIs. Alternatively, reducing the size of building space associated with an HEI, or condensing research and energy intensive spaces, might be beneficial to decrease the size of an HEI's CF (Fetcher 2009, Gui et al. 2020). The VIF analysis showed a high collinearity between GSF and FTE, which must be taken into account when predicting HEI GHG emissions, since both features cannot be used simultaneously in regression models.

Model evaluation

Overall, regression modeling produced more accurate predictions than median emissions intensity predictions, with MLR achieving the highest training accuracy for almost all emissions categories. The only better performing regression models in the literature were MLR models trained by Fetcher (2009) and Klein-Banai and Theis (2013). These models used certain features that were not included in my analysis. Fetcher predicted $\log_{10}$(scope 1 + 2) emissions with R^2 values up to 0.915 using July mean temperature, January mean low temperature, and the percentage of coal in the energy mix as features, in addition to gross building area or FTE data. Klein-Banai and Theis (2013) included features such as emissions from commuting, cooling degree days, heating degree days, square feet of residential facilities, and other net square feet to predict gross emissions, in addition to features included in my approach, and achieved R^2 values up to 0.954. STARS collects data for heating degree days and cooling degree days, therefore my analysis could be recreated and built upon using these features in MLR models. I chose to focus only on institutional characteristics that could be easily measured and provided by the HEI itself, and therefore did not use variables related to local climate in my analysis.

The models trained in this study did not achieve extremely high accuracy, as would be evidenced by training and test R^2 values of 0.9 or above; however, they can still be used to get general approximations for missing emissions data. The models can most closely predict emissions for scope 1 stationary (MLR: training R^2 = 0.667, test R^2 = 0.736), scope 1 total (MLR: training R^2 = 0.650, test R^2 = 0.670), scope 3 commuting (MLR: training R^2 = 0.720, test R^2 = 0.772), scope 3 total (MLR: training R^2 = 0.559, test R^2 = 0.632), and gross emissions (MLR: training R^2 = 0.798, test R^2 = 0.765), since good model performance was observed for both the training and test data for these emissions categories. For these categories, the models are considered relatively accurate and generalizable. Scope 1 other sources (SLR GSF: training R^2 = 0.600, test R^2 = 0.181) and scope 3 fuel and energy (MLR: training R^2 = 0.738, test R^2 = 0.340) had relatively high training accuracy and low test accuracy, which raises concerns about the external validity of these models.

Scope 3 emissions from business travel (MLR: training R^2 = 0.460, test R^2 = -0.075), goods and services (SLR FTE: training R^2 = 0.231, test R^2 = 0.441), and waste (MLR: training R^2 = 0.309, test R^2 = 0.379), could not be predicted very accurately, with training and test R^2 scores never exceeding 0.5. Models for all scope 2 categories generally performed poorly on the training data, although when using the test data, emissions from scope 2 electricity and scope 2 total were able to slightly surpass an R^2 of 0.5. The low predictive power of the models for scope 2 emissions is surprising, seeing as a strong correlation between HEI facility size, electricity consumption, and GHG emission was previously identified, which would suggest that scope 2 emissions could be predicted particularly well using variables such as GSF (Keegan 2006). Models for scope 2 thermal and scope 3 goods and services might have shown limited performance due to particularly small sample sizes. The other poor performing models applied to all the emissions categories that were not normally distributed after log-transforming the data and removing outliers, except scope 3 emissions from waste, which was normally distributed according to the Shapiro-Wilk test. My methodology could be refined and modified to better predict emissions for these categories. Low R^2 values for certain categories may also be an indication that that the linear relationship between log-emissions and the predictor data is not very strong.

In response to my central research question, "How can HEI GHG emissions be most accurately predicted for specific scopes and sub-scope categories of emissions?," I conclude that multiple linear regression can generally produce the highest accuracy predictions for most emissions categories. This is no surprise, since increasing model complexity typically guarantees better model performance; however, getting MLR estimates for different emissions categories involves collecting up to eleven data points on an HEI's institutional characteristics. Simple linear regression using GSF or FTE as the input feature can produce similarly accurate predictions while requiring only one input data point. SLR using FTE is favored over SLR using GSF for emissions from scope 3 commuting and scope 3 fuel and energy; for scope 3 commuting, this logically makes sense, since the extent commuting activities is much more closely related to the number of students and employees than to the gross square footage of building space.

Limitations and Future Directions

There are several limitations associated with my research. Data on GHG emissions and institutional characteristics is self-reported by HEIs to STARS. There is no assurance or verification of the validity of the reported data. Therefore, the accuracy of the data cannot be guaranteed with certainty. Another key limitation to my study is related to sample size. Participation in STARS is voluntary, which limits the number of HEIs for which data is available. Although STARS was chosen over SIMAP as a data source for this analysis, partly to achieve larger sample sizes, I still only had data for slightly over 300 HEIs. Reporting for different institutional characteristics and emissions categories was irregular between HEIs, which further limited my sample size and resulted in an inconsistent number of observations for each emissions category; in some cases, such as scope 2 thermal and scope 3 goods and services, the small sample size may have contributed to poorer model performance. Splitting my data into a training and test set was also a challenge, seeing as not much data was set aside for testing ($n < 50$ for all categories). This makes the interpretation of model performance on the testing data less conclusive, since the calculated R^2 and RMSE values may demonstrate greater variability or uncertainty; this means that the calculated metrics may not be as accurate as they would be with a larger sample size.

A final limitation regarding my data includes its geographic coverage; most HEIs in my sample were North American, which means that my analysis is mostly representative of trends in GHG emissions and institutional characteristics in North America. Sub-scope level data for HEI GHG emissions is not as readily accessible for institutions in other countries. The Higher Education Statistics Agency (HESA), for example, reports GHG emissions data for HEIs based in the United Kingdom; however, data reporting is not entirely aligned with the GHG Protocol, and most data is reported as scope totals (HESA 2023). Efforts like those undertaken by STARS or SIMAP to aggregate and publicly report HEI emissions and sustainability data should be expanded in other countries and regions.

One priority for my project was creating readily interpretable prediction models that could be implemented in a spreadsheet, instead of requiring more advanced tools such as a Jupyter Notebook and Python for model implementation. While linear regression models are easily interpretable, they are not necessarily the most powerful prediction models. More advanced machine learning models have been used to predict GHG emissions in other contexts (Ma et al. 2021, Heurtebize et al. 2022, Serafeim and Velez Caicedo 2022, Sun and Chenchen Huang 2022, Wang et al. 2023). Future research should focus on applying other such prediction models to the HEI context for predicting sub-scope level emissions.

Finally, the lack of formal guidance on HEI CF reporting, and the lack of official recommendations for the application of the GHG Protocol in the higher education sector leads to improper alignment between HEI frameworks such as SIMAP and STARS, and the GHG Protocol. Although the mapping stage of my analysis was included to address this shortcoming, it still relied on certain assumptions and interpretations that might not be completely accurate or appropriate for the HEI context. More research should be done to clarify the alignment between SIMAP, STARS and the GHG Protocol, and further guidance should be developed to advise which GHG Protocol sub-scope categories are truly relevant to the HEI context.

Broader Implications

Sustainability reporting among HEIs is in an early stage (Ceulemans et al. 2015, Sepasi et al. 2019). Universities and governments need to take steps to expand current emissions reporting efforts in the higher education sector, such as issuing declarations to promote sustainability in higher education, increasing governmental target-setting for emissions reductions, and expanding environmental auditing efforts (Grindsted 2011, Saha et al. 2021). GHG emissions disclosure and sustainability reporting are important as spur emissions reductions among HEIs, and offer a broad range of benefits across the entire university organization, such as increased participation of stakeholders in decision-making, stronger understanding and internalization of the institution's mission and values among stakeholders, and the development of new communication channels (Tehmina 2015, Yáñez et al. 2019). Beyond providing benefits to the university, GHG emissions disclosure is a key aspect of HEIs fulfilling their role and responsibility to promote sustainable development and contribute to climate change mitigation (Knuth et al. 2007, Sedlacek 2013). Actors across society, including governments, corporations, and institutions like HEIs must all do their part to reduce GHG emissions and combat the climate crisis.

My research is a response to a lack of comprehensive CF reporting across universities. The adaptation of corporate standards for sustainability reporting can increase comparability and consistency across HEIs, as was demonstrated by efforts to adapt the German Council for Sustainable Development's sustainability code to the higher education sector (Huber and Bassen 2018); however, in other cases, HEIs continue to adapt corporate standards in a less standardized and concerted manner (Robinson et al. 2018). CF reporting efforts among HEIs need to be expanded across the globe, and an international standard or guideline for GHG accounting in the higher education sector needs to be formally implemented and agreed upon. Only this way can CFs between HEIs truly be comparable, and reach a common level of depth and accuracy.

ACKNOWLEDGEMENTS

I would like to thank Tina Mendez for her wholehearted commitment to the ES Thesis program, and for guiding me through this exciting and challenging process. I thank Danielle Perryman and Robin Lopez for their feedback, and for all their time and effort. I am very grateful for the support I received from friends and family, including my parents, Max, John, Ina, Laura, Zack, and Will – thank you for cheering me up and cheering me on throughout this process! Thank you, Christina Schulte, for all your helpful peer-review feedback, and for being my thesis buddy.

I am very grateful to Hannah Helmke for the kind introduction for right° based on science, and I thank Franziska Neumann and Elnaz Roshan from right° open for enabling and supporting this research collaboration. Thank you, Elnaz, for your extensive advice and feedback on my work.

Thank you, Allison Leach and Monika Urbanski, for helping me access and wrangle HEI GHG emissions data. I would also like to thank Michael Xiao and Mehdi Badri for their feedback and modeling expertise – your contributions were integral to my work!

REFERENCES

AASHE. 2023. Association for the Advancement of Sustainability in Higher Education. Discover Who We Are. https://www.aashe.org/about-us/who-we-are/.

Abad-Segura, Cortés-García, and Belmonte-Ureña. 2019. The Sustainable Approach to Corporate Social Responsibility: A Global Analysis and Future Trends. Sustainability 11:5382.

Alonso-Almeida, M. del M., F. Marimon, F. Casani, and J. Rodriguez-Pomeda. 2015. Diffusion of sustainability reporting in universities: current situation and future perspectives. Journal of Cleaner Production 106:144–154.

Amaral, A. R., E. Rodrigues, A. R. Gaspar, and Á. Gomes. 2020. A review of empirical data of sustainability initiatives in university campus operations. Journal of Cleaner Production 250:119558.

Carbon Disclosure Project. 2022. CDP Homepage. https://www.cdp.net/en.

Carlson, K. M., J. S. Gerber, N. D. Mueller, M. Herrero, G. K. MacDonald, K. A. Brauman, P. Havlik, C. S. O'Connell, J. A. Johnson, S. Saatchi, and P. C. West. 2017. Greenhouse gas emissions intensity of global croplands. Nature Climate Change 7:63–68.

Ceulemans, K., R. Lozano, and M. Alonso-Almeida. 2015. Sustainability Reporting in Higher Education: Interconnecting the Reporting Process and Organisational Change Management for Sustainability. Sustainability 7:8881–8903.

Cevik, S., and J. T. Jalles. 2023. For whom the bell tolls: Climate change and income inequality. Energy Policy 174:113475.

Clabeaux, R., M. Carbajales-Dale, D. Ladner, and T. Walker. 2020. Assessing the carbon footprint of a university campus using a life cycle assessment approach. Journal of Cleaner Production 273:122600.

Cordero, E. C., D. Centeno, and A. M. Todd. 2020. The role of climate change education on individual lifetime carbon emissions. PLOS ONE 15:e0206266.

De Iorio, S., G. Zampone, and A. Piccolo. 2022. Determinant Factors of SDG Disclosure in the University Context. Administrative Sciences 12:21.

Desme, G., and L. Smart. 2018. Accounting For Carbon: Sovereign Bonds. Truscot S&P Global.

Dyer, G., and M. Dyer. 2017. Strategic leadership for sustainability by higher education: the American College & University Presidents' Climate Commitment. Journal of Cleaner Production 140:111–116.

Fetcher, N. 2009. Research and Solutions: Effects of Climate and Institution Size on Greenhouse Gas Emissions from Colleges and Universities in the United States. Sustainability: The Journal of Record 2:362–367.

Foust, V. S. 2016. An Analysis of Greenhouse-Gas Reduction Commitment, Diffusion and Success in the Higher Education Sector. Ph.D., North Carolina Agricultural and Technical State University, United States -- North Carolina.

García-Alaminos, Á., E. Gilles, F. Monsalve, and J. Zafrilla. 2022. Measuring a university's environmental performance: A standardized proposal for carbon footprint assessment. Journal of Cleaner Production 357:131783.

Global Footprint Network. 2022. Climate Change & the Carbon Footprint. https://www.footprintnetwork.org/our-work/climate-change/.

Goldemberg, J. 2020. The evolution of the energy and carbon intensities of developing countries. Energy Policy 137:111060.

Greenhouse Gas Protocol. 2004. The GHG Protocol Corporate Accounting and Reporting Standard.

Greenhouse Gas Protocol. 2011. The Corporate Value Chain (Scope 3) Accounting and Reporting Standard.

Greenhouse Gas Protocol. 2015. GHG Protocol Scope 2 Guidance.

Greenhouse Gas Protocol. 2022. Greenhouse Gas Protocol. https://ghgprotocol.org/.

Grindsted, T. S. 2011. Sustainable Universities From Declarations on Sustainability in Higher Education to National Law. SSRN Electronic Journal.

Gui, X., Z. Gou, and F. Zhang. 2020. The relationship between energy use and space use of higher educational buildings in subtropical Australia. Energy and Buildings 211:109799.

Helmers, E., C. C. Chang, and J. Dauwels. 2021. Carbon footprinting of universities worldwide: Part I—objective comparison by standardized metrics. Environmental Sciences Europe 33:30.

Helmke, H., H.-P. Hafner, F. Gebert, and A. Pankiewicz. 2020. Provision of Climate Services—The XDC Model. Pages 223–249 in W. Leal Filho and D. Jacob, editors. Handbook of Climate Services. Springer International Publishing, Cham.

HESA. 2023. Table 3 - Emissions and waste | HESA. https://www.hesa.ac.uk/data-and-analysis/estates/table-3.

Heurtebize, T., F. Chen, F. Soupé, and R. L. D. Carvalho. 2022. Corporate Carbon Footprint: A Machine Learning Predictive Model for Unreported Data. SSRN Electronic Journal.

Huber, S., and A. Bassen. 2018. Towards a sustainability reporting guideline in higher education. International Journal of Sustainability in Higher Education 19:218–232.

Keegan, B. 2006. Cooling Towers among Ivory Towers: A Comparative Analysis of Research Universities, Electricity Consumption, and Greenhouse Gas Emissions. Massachusetts Institute of Technology.

Klein-Banai, C., and T. L. Theis. 2013. Quantitative analysis of factors affecting greenhouse gas emissions at institutions of higher education. Journal of Cleaner Production 48:29–38.

Knuth, S., B. Nagle, C. Steuer, and B. Yarnal. 2007. Universities and Climate Change Mitigation: Advancing Grassroots Climate Policy in the US. Local Environment 12:485–504.

Larsen, H. N., J. Pettersen, C. Solli, and E. G. Hertwich. 2013. Investigating the Carbon Footprint of a University - The case of NTNU. Journal of Cleaner Production 48:39–47.

Li, Z., Z. Chen, N. Yang, K. Wei, Z. Ling, Q. Liu, G. Chen, and B. H. Ye. 2021. Trends in research on the carbon footprint of higher education: A bibliometric analysis (2010–2019). Journal of Cleaner Production 289:125642.

Lozano, R., R. Lukman, F. J. Lozano, D. Huisingh, and W. Lambrechts. 2013. Declarations for sustainability in higher education: becoming better leaders, through addressing the university system. Journal of Cleaner Production 48:10–19.

Ma, N., W. Y. Shum, T. Han, and F. Lai. 2021. Can Machine Learning be Applied to Carbon Emissions Analysis: An Application to the CO2 Emissions Analysis Using Gaussian Process Regression. Frontiers in Energy Research 9:756311.

McCowan, T. 2020. The impact of universities on climate change: a theoretical framework. Centre for Global Higher Education, University of Oxford.

Nebe, A.-K. 2022. Climate impact analysis in the higher education sector: An extension of the X-Degree Compatibility Model. Master Thesis, University of Mannheim.

Peer, V., and G. Stoeglehner. 2013. Universities as change agents for sustainability – framing the role of knowledge transfer and generation in regional development processes. Journal of Cleaner Production 44:85–95.

Perchinunno, P., and M. Cazzolle. 2020. A clustering approach for classifying universities in a world sustainability ranking. Environmental Impact Assessment Review 85:106471.

right° based on science. 2023. XDC Model - right°. https://www.right-basedonscience.de/en/xdc-model/.

Robinson, O. J., A. Tewkesbury, S. Kemp, and I. D. Williams. 2018. Towards a universal carbon footprint standard: A case study of carbon management at universities. Journal of Cleaner Production 172:4435–4455.

Saha, A. K., H. Al-Shaer, R. Dixon, and I. Demirag. 2021. Determinants of Carbon Emission Disclosures and UN Sustainable Development Goals: The Case of UK Higher Education Institutions. Australian Accounting Review 31:79–107.

Second Nature. 2022a. Second Nature Reporting Database. https://reporting.secondnature.org/.

Second Nature. 2022b. Measuring Progress. https://secondnature.org/signatory-handbook/measuring-progress/.

Second Nature. 2023a. Our History. https://secondnature.org/our-history/.

Second Nature. 2023b. SIMAP®: Sustainability Indicator Management and Analysis Platform. https://secondnature.org/signatory-handbook/simap/.

Sedlacek, S. 2013. The role of universities in fostering sustainable development at the regional level. Journal of Cleaner Production 48:74–84.

Sepasi, S., U. Braendle, and A. H. Rahdari. 2019. Comprehensive sustainability reporting in higher education institutions. Social Responsibility Journal 15:155–170.

Serafeim, G., and G. Velez Caicedo. 2022. Machine Learning Models for Prediction of Scope 3 Carbon Emissions. SSRN Electronic Journal.

Shi, H., and E. Lai. 2013. An alternative university sustainability rating framework with a structured criteria tree. Journal of Cleaner Production 61:59–69.

SIMAP. 2023. About SIMAP | SIMAP. https://unhsimap.org/cmap/about.

Sinha, P., W. A. Schew, A. Sawant, K. J. Kolwaite, and S. A. Strode. 2010. Greenhouse Gas Emissions from U.S. Institutions of Higher Education. Journal of the Air & Waste Management Association 60:568–573.

Sprenk, U. 2022. Assessing Emissions Data Quality for Higher Education Institutions: Development of an Evaluation Scheme for Different Emissions Scopes. Master Thesis, University of Mannheim.

STARS. 2023a. STARS, Sustainability Tracking Assessment & Rating System. https://stars.aashe.org/.

STARS. 2023b. Participants & Reports | Institutions | STARS Reports. https://reports.aashe.org/institutions/participants-and-reports/.

Sun, W. and Chenchen Huang. 2022. Predictions of carbon emission intensity based on factor analysis and an improved extreme learning machine from the perspective of carbon emission efficiency. Journal of Cleaner Production 338:130414.

Tehmina, S. 2015. Role of greenhouse gas inventories in climate change mitigation at Institutions of Higher Education. M.S., Duquesne University, United States -- Pennsylvania.

Valls-Val, K., and M. D. Bovea. 2021. Carbon footprint in Higher Education Institutions: a literature review and prospects for future research. Clean Technologies and Environmental Policy 23:2523–2542.

Valls-Val, K., and M. D. Bovea. 2022. Carbon footprint assessment tool for universities: CO2UNV. Sustainable Production and Consumption 29:791–804.

Wadud, Z., S. Royston, and J. Selby. 2019. Modelling energy demand from higher education institutions: A case study of the UK. Applied Energy 233–234:816–826.

Walenta, J. 2021. The making of the corporate carbon footprint: the politics behind emission scoping. Journal of Cultural Economy 14:533–548.

Wang, C., M. Li, and J. Yan. 2023. Forecasting carbon dioxide emissions: application of a novel two-stage procedure based on machine learning models. Journal of Water and Climate Change 14:477–493.

Yáñez, S., Á. Uruburu, A. Moreno, and J. Lumbreras. 2019. The sustainability report as an essential tool for the holistic and strategic vision of higher education institutions. Journal of Cleaner Production 207:57–66.

Yilmaz, V., and Y. Can. 2020. Impact of knowledge, concern and awareness about global warming and global climatic change on environmental behavior. Environment, Development and Sustainability 22:6245–6260.

Zhaurova, L. 2008, August. U.S. higher education and global climate change: An exploration of institutional factors that affect greenhouse gas emissions. Masters, Tufts University.

APPENDIX A: External Links

Name	Source	Link
Guidance for STARS participants using SIMAP	AASHE	a) AASHE website: Help Center > Operations > GHG Emissions: 'My institution uses SIMAP for its greenhouse gas emissions reporting. Are there any specific guidelines for SIMAP users?': https://stars.aashe.org/resources-support/help-center/operations/greenhouse-gas-emissions/#my-institution-uses-simap-for-its-greenhouse-gas-emissions-reporting--are-there-any-specific-guidelines-for-simap-users b) 'Guidance for STARS participants using SIMAP.pdf' in Google Drive: https://drive.google.com/file/d/1ph8wZdVXsf4znIwpk_cJeof67SXnO-RF/view?usp=sharing
STARS Report Content Data Display	STARS	https://reports.aashe.org/institutions/data-displays/2.0/content/?institution__institution_type=DO_NOT_FILTER

APPENDIX B: Data Queries from STARS Report Content Data Display

Variable	STARS Version	Type of Characteristic	Specific Characteristic	Category	Subcategory	Credit	Reporting Field	Date	Time
S1 Stationary Combustion	2.2	Institution Type	All Institutions	Operations	Air & Climate	OP-1: Emissions Inventory and Disclosure	Gross Scope 1 GHG emissions from stationary combustion, performance year	2/5/23	12:00
S1 Other Sources	2.2	Institution Type	All Institutions	Operations	Air & Climate	OP-1: Emissions Inventory and Disclosure	Gross Scope 1 GHG emissions from other sources, performance year	2/5/23	12:04
S1 Total	2.2	Institution Type	All Institutions	Operations	Air & Climate	OP-1: Emissions Inventory and Disclosure	Total gross Scope 1 GHG emissions, performance year	2/5/23	12:07
S2 Electricity	2.2	Institution Type	All Institutions	Operations	Air & Climate	OP-1: Emissions Inventory and Disclosure	Gross Scope 2 GHG emissions from imported electricity, performance year	2/5/23	12:13
S2 Thermal Energy	2.2	Institution Type	All Institutions	Operations	Air & Climate	OP-1: Emissions Inventory and Disclosure	Gross Scope 2 GHG emissions from imported thermal energy, performance year	2/5/23	12:14
S2 Total	2.2	Institution Type	All Institutions	Operations	Air & Climate	OP-1: Emissions Inventory and Disclosure	Total gross Scope 2 GHG emissions, performance year	2/5/23	12:16
S3 Business Travel	2.2	Institution Type	All Institutions	Operations	Air & Climate	OP-1: Emissions Inventory and Disclosure	Scope 3 GHG emissions from business travel, performance year	2/5/23	12:18
S3 Commuting	2.2	Institution Type	All Institutions	Operations	Air & Climate	OP-1: Emissions Inventory and Disclosure	Scope 3 GHG emissions from commuting, performance year	2/5/23	12:19
S3 Goods and Services	2.2	Institution Type	All Institutions	Operations	Air & Climate	OP-1: Emissions Inventory and Disclosure	Scope 3 GHG emissions from purchased goods and services, performance year	2/5/23	12:20
S3 Capital Goods	2.2	Institution Type	All Institutions	Operations	Air & Climate	OP-1: Emissions Inventory and Disclosure	Scope 3 GHG emissions from capital goods, performance year	2/5/23	12:21

S3 Fuel and Energy Related	2.2	Institution Type	All Institutions	Operations	Air & Climate	OP-1: Emissions Inventory and Disclosure	Scope 3 GHG emissions from fuel- and energy-related activities not included in Scope 1 or Scope 2, performance year	2/5/23	12:22
S3 Waste	2.2	Institution Type	All Institutions	Operations	Air & Climate	OP-1: Emissions Inventory and Disclosure	Scope 3 GHG emissions from waste generated in operations, performance year	2/5/23	12:23
S3 Other	2.2	Institution Type	All Institutions	Operations	Air & Climate	OP-1: Emissions Inventory and Disclosure	Scope 3 GHG emissions from other sources not included in Scope 1 or 2, performance year	2/5/23	12:25
S3 Total	2.2	Institution Type	All Institutions	Operations	Air & Climate	OP-1: Emissions Inventory and Disclosure	Total Scope 3 GHG emissions, performance year	2/5/23	12:27
Institutional Control	2.2	Institution Type	All Institutions	Report Preface	Institutional Characteristics	PRE-3: Institutional Boundary	Institutional control	2/5/23	12:42
Medical School	2.2	Institution Type	All Institutions	Report Preface	Institutional Characteristics	PRE-3: Institutional Boundary	Is a medical school included in the institutional boundary?	2/5/23	12:45
Professional Schools, Labs, Clinincs	2.2	Institution Type	All Institutions	Report Preface	Institutional Characteristics	PRE-3: Institutional Boundary	Are all other professional schools with labs or clinics included in the institutional boundary?	2/5/23	12:46
Satellite Campuses	2.2	Institution Type	All Institutions	Report Preface	Institutional Characteristics	PRE-3: Institutional Boundary	Are all the satellite campuses included in the institutional boundary?	2/5/23	12:49
Hospital	2.2	Institution Type	All Institutions	Report Preface	Institutional Characteristics	PRE-3: Institutional Boundary	Is the hospital included in the institutional boundary?	2/5/23	12:50
Endowment	2.2	Institution Type	All Institutions	Report Preface	Institutional Characteristics	PRE-4: Operational Characteristics	Endowment size	2/5/23	12:52
GSF	2.2	Institution Type	All Institutions	Report Preface	Institutional Characteristics	PRE-4: Operational Characteristics	Gross floor area of building space	2/5/23	12:55
Lab GSF	2.2	Institution Type	All Institutions	Report Preface	Institutional Characteristics	PRE-4: Operational Characteristics	Floor area of laboratory space	2/5/23	12:56
Health Care GSF	2.2	Institution Type	All Institutions	Report Preface	Institutional Characteristics	PRE-4: Operational Characteristics	Floor area of healthcare space	2/5/23	12:57

Student FTE	2.2	Institution Type	All Institutions	Report Preface	Institutional Characteristics	PRE-5: Academics and Demographics	Full-time equivalent student enrollment	2/5/23	12:59
Employee FTE	2.2	Institution Type	All Institutions	Report Preface	Institutional Characteristics	PRE-5: Academics and Demographics	Full-time equivalent of employees	2/5/23	13:00
Employee Researchers	2.2	Institution Type	All Institutions	Academics	Research	AC-9 Research and Scholarships	Total number of employees that conduct research	2/5/23	13:04