SE 2050 Commitment Program 2023 Data Analysis and Findings Report
Authors: Demi Fang, Mel Chafart, Martín Torres, Jonathan Broyles
Editors: Lauren Wingo, Zachary Chabot, Eric Borchers

Executive Summary

This report details the methods, results, implications, and recommendations of a comprehensive data analysis of the SE 2050 Database, conducted by the Data Science Team (DST) internal working group within the SE 2050 Committee. The SE 2050 Commitment Program was established in 2020 with the vision to empower the structural engineering industry to understand, reduce, and ultimately eliminate embodied carbon in building structures. As structural engineering firms sign on to the commitment as SE 2050 Signatory Firms, they submit anonymized data on building structures that they have designed to the SE 2050 Database. On November 1, 2023, the database had 522 project submissions, which were primarily US projects. The projects captured in the database encompass a range of building types, each with different building features, structural engineering features, and embodied carbon performance of the structural system, measured as Global Warming Potential (GWP). This analysis is the first large-scale study in North America to investigate the statistical relationship between design inputs of the structural system and GWP from real-world data. In support of the vision of the SE 2050 Commitment Program, the goals of this report are to: (1) define the data filtering and analysis methods, (2) interpret the analysis results, (3) propose recommendations for design practitioners and suggestions for future benchmarks, and (4) provide recommendations to the SE 2050 Committee to improve future data collection and analyses.

Methodology

The data was first filtered to remove values deemed out of scope (infeasible or potentially incorrect). The filtered project data was categorized into two datasets according to the Life Cycle Assessment (LCA) stages reported: “Upfront Carbon,” which included projects with GWP intensities associated with the A stages (A1-A3 at minimum, with A4 and A5 if provided for the project); and “A to C,” which included projects with GWP intensities associated with stages A through C. The two data subsets had 241 and 100 data entries, respectively (Figure 1). Projects assessed for life-cycle stages including module D (potential loads and benefits beyond the system boundary) are separated from this analoysis because the fate of construction products is often unknown, resulting in an inability to reliably assume that calculated benefits will be realized.

Fang, Demi, Mel Chafart, Martín Torres, and Jonathan Broyles. “Executive Summary of the SE 2050 Commitment Program 2023 Data Analysis and Findings Report.” Edited by Lauren Wingo, Zachary Chabot, and Eric Borchers. SE 2050, in press with the American Society of Civil Engineers ASCE.

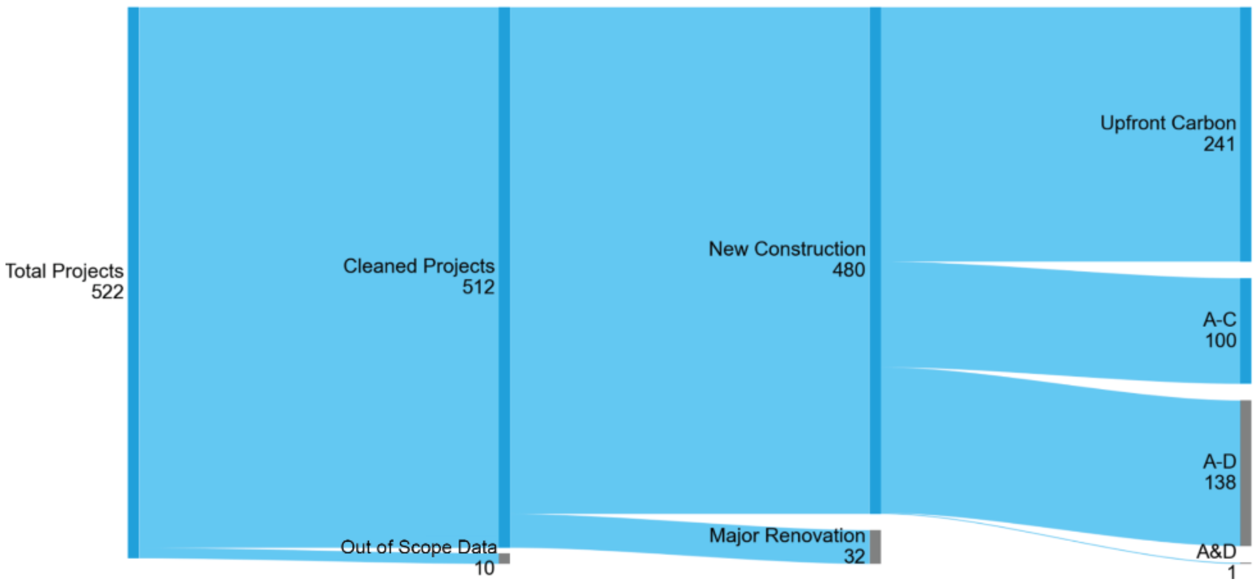

Figure 1. Sankey diagram demonstrating number of projects available throughout data filtering steps. The subsets from the last filtering step refer to life-cycle stages, for example “A-C” refers to projects with GWP intensities associated with life-cycle stages A through C.

After obtaining filtered subsets of data, an exploratory data analysis was conducted to evaluate relationships between the performance objective and data input variables. The performance objective for this analysis was defined as GWP Intensity Excluding Biogenic Carbon: kilograms of CO_2 equivalent per square meter of floor area ($kgCO_2e/m^2$), where biogenic carbon was excluded from the analysis to avoid unknown and potentially non-uniform approaches to quantifying biogenic carbon across projects. In some cases, Total GWP Excluding Biogenic Carbon (kgCO2e) was also examined as a performance objective. For conciseness, the exclusion of biogenic carbon is assumed when using the terms “Total GWP” and “GWP Intensity” throughout the rest of the Executive Summary.

Data input variables available in the database include building features, such as Gross Floor Area and Stories Above Grade, as well as structural engineering features, such as the Typical Floor Live Load and Primary Horizontal Gravity System. Variables were analyzed according to committee priority and data type, with high-priority and secondary-priority variables summarized in Table 1.

Fang, Demi, Mel Chafart, Martín Torres, and Jonathan Broyles. "Executive Summary of the SE 2050 Commitment Program 2023 Data Analysis and Findings Report." Edited by Lauren Wingo, Zachary Chabot, and Eric Borchers. SE 2050, in press with the American Society of Civil Engineers ASCE.

Table 1. Summary of data input variables identified as high-priority and secondary-priority ahead of data analysis, and their respective data types. This is not an exhaustive list of all variables analyzed. A full table of all variables is provided in Appendix A of the full report.

Data input	**Priority**	**Variable type**
Building Use Type	High	Discrete, un-ordered
Primary Horizontal Gravity System	High	Discrete, un-ordered
Gross Floor Area	High	Continuous
Mean Roof Height	High	Continuous
Stories Above Grade	High	Continuous
LCA Tool	Secondary	Discrete, un-ordered
Project Phase at LCA	Secondary	Discrete, ordered
Project Region	Secondary	Discrete, un-ordered
Stories Below Grade	Secondary	Continuous

The exploratory data analysis was followed by more detailed statistical analyses to clarify the strength of relationships in the data. These analyses include linear regressions, machine learning regressions with feature importance, and other statistical models.

Results

The report and this Executive Summary focus on results from the Upfront Carbon dataset, though results from the A to C dataset are also presented in the appendix of the full report and noted when differing significantly from the findings from the Upfront Carbon dataset. A snapshot of the spread of GWP Intensity across the Upfront Carbon dataset is presented in Figure 2.

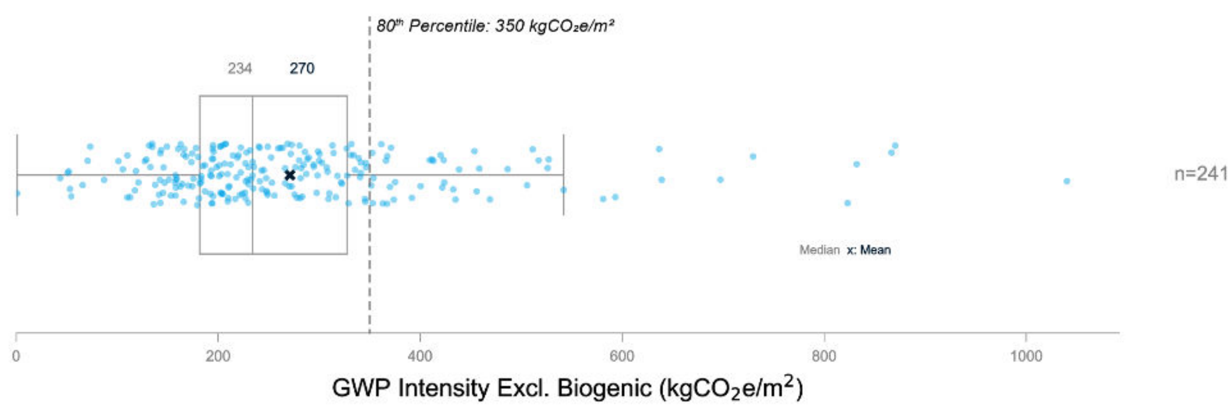

Figure 2. Distribution of GWP Intensity for projects in the Upfront Carbon data set.

It was determined that there is a clear correlation between Total GWP and Gross Floor Area (Figure 3) (log-log axes offer a way to measure linear relationships between continuous distributions with a strong right skew, as the data showed). Most continuous variables did not individually exhibit a correlation with GWP or GWP Intensity, but the feature importance results from machine learning regressions suggested Gross Floor Area, Typical Floor Live Load, and

Fang, Demi, Mel Chafart, Martín Torres, and Jonathan Broyles. “Executive Summary of the SE 2050 Commitment Program 2023 Data Analysis and Findings Report.” Edited by Lauren Wingo, Zachary Chabot, and Eric Borchers. SE 2050, in press with the American Society of Civil Engineers ASCE.

Mean Roof Height, to be important predictors of GWP Intensity. The discrete variables Building Use Type, Primary Horizontal Gravity System, Primary Vertical Gravity System were found to be influential predictors of GWP Intensity through the analyses (Figures 4-6).

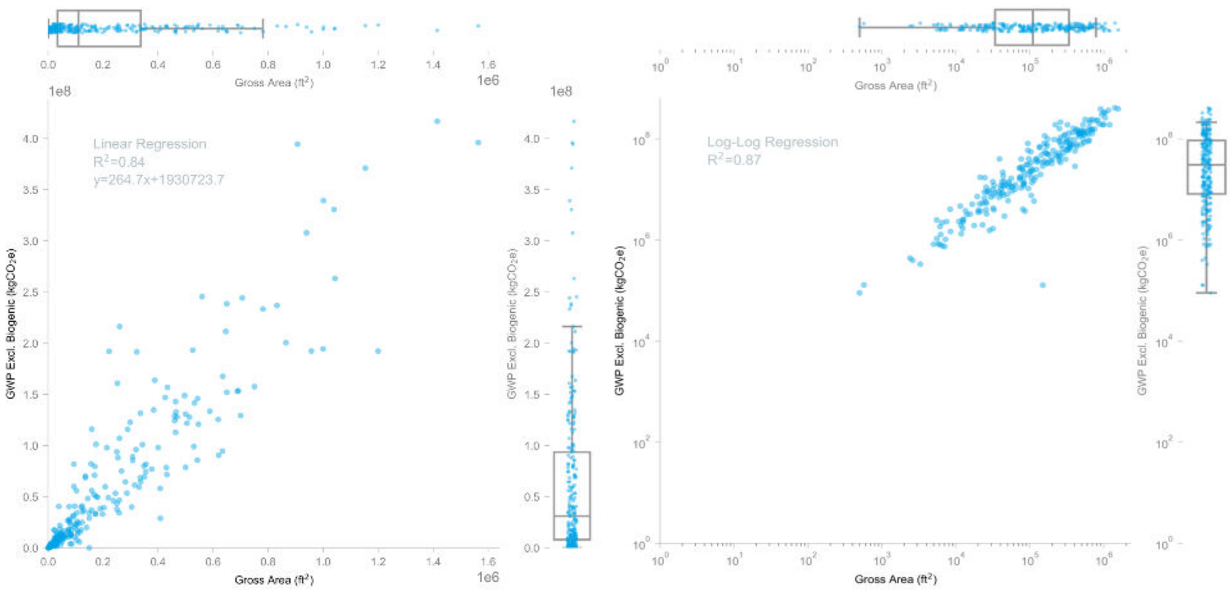

Figure 3. Scatter plot of GWP against Gross Floor Area for Upfront Carbon projects, on (left) linear and (right) log axes.

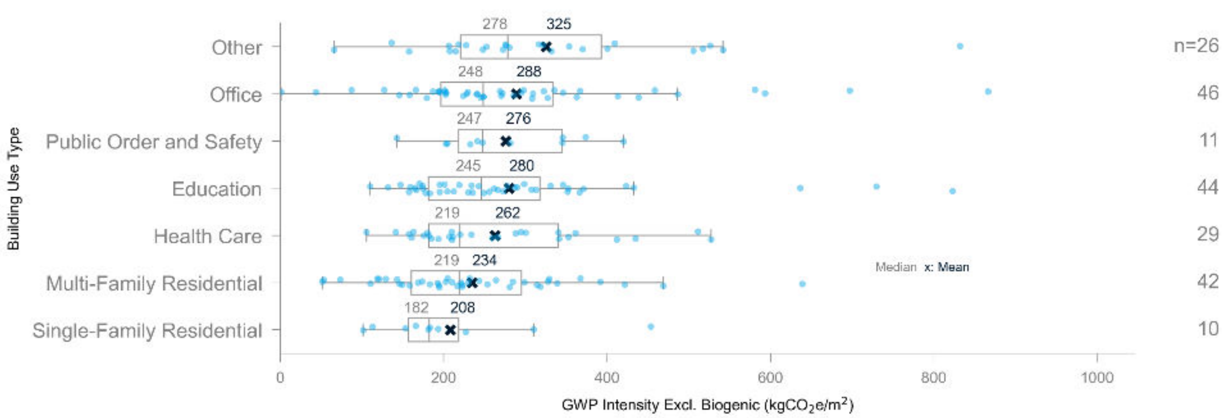

Figure 4. Distribution of GWP Intensity by Building Use Type in the Upfront Carbon dataset. Categories with fewer than 10 data entries are not shown.

Fang, Demi, Mel Chafart, Martín Torres, and Jonathan Broyles. “Executive Summary of the SE 2050 Commitment Program 2023 Data Analysis and Findings Report.” Edited by Lauren Wingo, Zachary Chabot, and Eric Borchers. SE 2050, in press with the American Society of Civil Engineers ASCE.

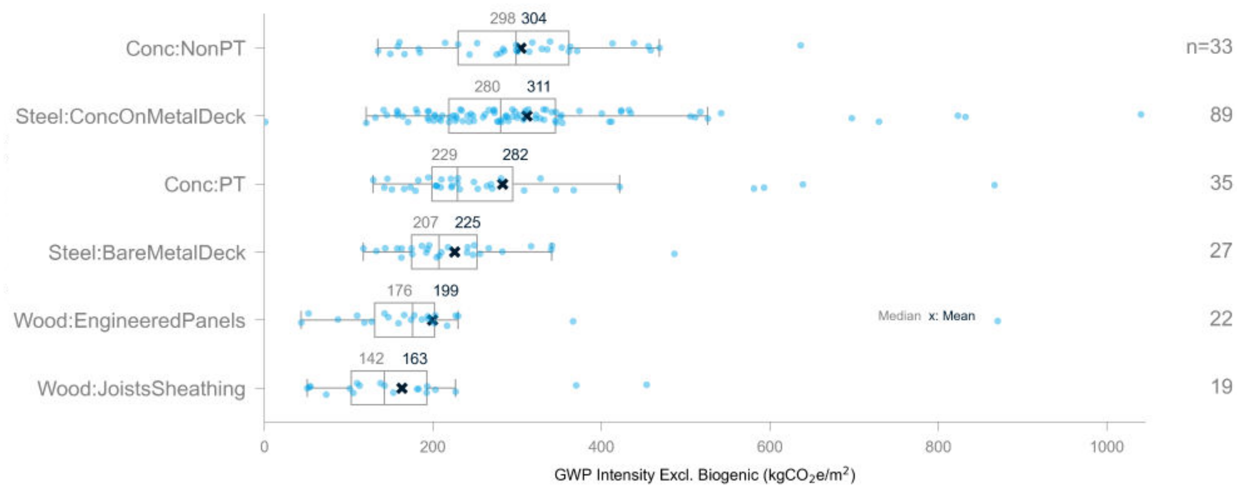

Figure 5. Distribution of GWP Intensity by Primary Horizontal Gravity System in the Upfront Carbon dataset. Categories with fewer than 10 data entries are not shown.

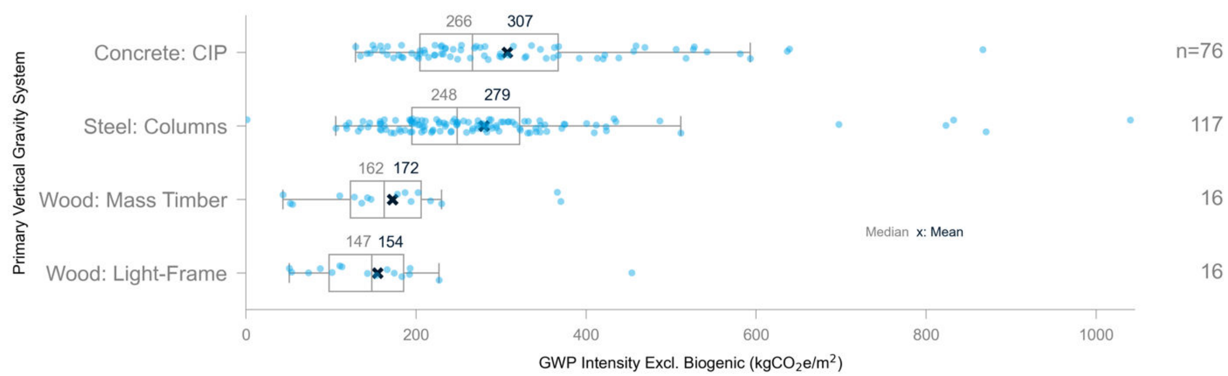

Figure 6. Distribution of GWP Intensity by Primary Vertical Gravity System in the Upfront Carbon dataset. Categories with fewer than 10 data entries are not shown.

Readers interested in the relationship between pairs of data variables in the datasets, such as the relative frequency of data projects of a particular Building Use Type and Typical Floor Live Load available, are referred to Appendix C of the full report. These analyses do not account for performance objectives such as Total GWP or GWP Intensity but give other information about the dataset. The most salient findings are intuitive, such as the high positive correlation between Stories Above Grade and Mean Roof Height.

Recommendations to practitioners

While no universal solution can address embodied carbon reduction across all projects due to the inherent complexities of structural engineering, analyzing large datasets of buildings can provide valuable insights that advance state-of-practice. Building designers and structural engineers can take note of the salient relationships identified in this data analysis. Gross Floor Area and Total

Fang, Demi, Mel Chafart, Martín Torres, and Jonathan Broyles. "Executive Summary of the SE 2050 Commitment Program 2023 Data Analysis and Findings Report." Edited by Lauren Wingo, Zachary Chabot, and Eric Borchers. SE 2050, in press with the American Society of Civil Engineers ASCE.

GWP had the strongest correlation that was statistically significant (Figure 2). Among a variety of regression techniques and feature importance models, the following design variables emerged as being meaningfully correlated with GWP Intensity, in no particular order:

- Gross Floor Area
- Mean Roof Height
- Typical Floor Live Load
- Primary Horizontal Gravity System
- Primary Vertical Gravity System
- LCA Tool
- Building Use Type

More data and analyses are needed to confidently determine the hierarchy of these variables' importance relative to GWP Intensity. The combinations of architecturally- and structurally-driven variables emerging as important predictors of GWP Intensity suggest the importance of architect-engineer collaboration during the design process for embodied carbon reductions. Possible explanations for the appearance of "LCA Tool" as a predictor of GWP Intensity include differences in underlying data, assumptions, and input prompts; these explanations are further discussed in the full report.

Another clear result from the analysis is that the distribution of GWP Intensity in the current database exhibits high variance, which is influenced by a variety of factors, and therefore cannot yet be used to establish precise benchmarks. Although embodied carbon benchmarks for classifications such as Building Use Type and Primary Horizontal Gravity System are not provided, a recommended limit for upfront carbon is proposed. The 80th percentile of GWP Intensity from the Upfront Carbon (A1-A5) dataset was selected as a conservative upper limit for structural embodied carbon: 350 $kgCO_2e/m^2$ (Figure 2). As the database grows, it is important to regularly update these results to ensure that the industry is supplied with the most up-to-date, science-based recommendations to support impactful embodied carbon reduction efforts. Furthermore, analyzing larger amounts of data in future iterations can lead to more definitive conclusions to inform embodied carbon benchmarks for structural engineers.

Recommendations to the SE 2050 Committee

The report successfully addresses most of the goals (data analysis, interpretation, and recommendations) established for the inaugural report. However, the provision of substantial feature-specific embodied carbon benchmarks requires additional round(s) of collecting high-fidelity project data based on several recommendations. These recommendations are to (1) require the submission of environmental impacts per life cycle stage, (2) require the submission of quantities granular to assembly (super- vs. substructure) and assumed GWP factors, and (3) provide more guidance on submission of biogenic carbon uptake and emissions. These recommendations are provided to the SE 2050 Committee to inform their continued efforts to support the pursuit of net-zero structural systems.