Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

TimoDS: A comprehensive and synthetically generated dataset of Timoshenko steel beams with general elastic supports and responses to elementary unit load cases

Youssef Derrazi

e.yderrazi@ms.ugr.es

ETSIIT, University of Granada

Diego Hernán Peluffo-Ordóñez

Universidad Yachay Tech

Juan Carlos Torres Cantero

ETSIIT, University of Granada

data-descriptor

Keywords:

Posted Date: August 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9663285/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

TimoDS: A comprehensive and synthetically generated dataset of TIMOSHENKO steel beams with general elastic supports and responses to elementary unit load cases

Youssef DERRAZI*[1,2,5], Diego Hernán PELUFFO-ORDÓÑEZ[3,4], and Juan Carlos TORRES CANTERO[1]

[1]Virtual Reality Laboratory, ETSIIT, Department of Computer Languages and Systems, University of Granada, c/Periodista Manuel Saucedo Aranda, s/n,Granada 18071, España
[2]SDAS Research Group (https://sdas-group.com/), Hacienda San José, Urcuquí 100115, Ecuador
[3]School of Mathematical and Computational Sciences, Universidad Yachay Tech, Hacienda San José, Urcuquí 100115, Ecuador
[4]Investigación, Universidad Ecotec, Universidad Ecotec, Km. 13.5 Samborondón, Samborondón 092302, Ecuador
[5]EMINES School of Industrial Management, University Mohammed 6 Polytechnic, Lot 660, Hay Moulay Rachid, Benguerir, 43150, Morocco

Y. DERRAZI
D. PELUFFO
J.C. TORRES

Contents

1 Introduction **3**

2 Methods **4**
2.1 Mechanical model 4
2.2 Data Generation Pipeline 11

3 Data Records **15**

4 Data Overview and technical validation **15**
4.1 Statistical characterization of the input space 16
4.2 Statistical characterization of the output space 22
4.3 Physical consistency 40
4.4 Strain energy consistency and correction 47

5 Baseline machine learning experiment **49**
5.1 Example protocol 49
5.2 Results 50
5.3 Discussion 51

6 Usage Notes **52**
6.1 Superposition and generalization 52
6.2 Potential applications 53
6.3 Limitations and future extensions 54

Data Availability **55**

Code Availability **55**

*Corresponding author: e.yderrazi@ms.ugr.es

Bibliography **56**

Author contributions **57**

Competing interests **57**

Acknowledgements **57**

Y. Derrazi
D. Peluffo
J.C. Torres

Abstract

This paper presents TimoDS, a synthetically generated dataset of 60000 Timoshenko steel beam instances with general elastic supports, together with a baseline machine learning experiment demonstrating its utility for data-driven structural mechanics. Each instance corresponds to a fully defined linear static beam problem, characterised by a beam length, a European steel cross-section from the IPE, HEA or HEB profile families, and twelve elastic support stiffness parameters spanning the full continuum from free to quasi-clamped boundary conditions. Six elementary unit load cases, comprising three unit point loads and three unit point moments, are provided independently for each instance, enabling the reconstruction of responses to arbitrary loading distributions via superposition. For each problem, the spatial distributions of kinematic quantities, internal forces and strain energy are sampled at 21 equally spaced positions along the beam axis, yielding an output vector of 253 quantities per instance. The dataset was generated through a fully reproducible pipeline combining Latin Hypercube Sampling with a fixed random seed and an automated parametric graph implemented in Grasshopper and Karamba3D. A feedforward neural network surrogate trained on a single load case demonstrates accurate reconstruction of response fields and confirms that the superposition principle is satisfied by the trained model on out-of-sample loading scenarios, establishing the dataset as a viable benchmark for surrogate modelling, physics-informed learning and operator learning in structural engineering.

1 Introduction

The forward simulation of structural mechanical behaviour is a central task in civil and structural engineering. Traditional numerical methods, such as the Finite Element Method (FEM), are well established and physically interpretable, but remain computationally intensive for iterative workflows such as design optimization, uncertainty quantification, or inverse problems. Data driven surrogate models have emerged as a complementary approach, offering the prospect of significant computational speedups once trained on sufficiently large and representative datasets.

The development of such models in structural engineering is however constrained by the limited availability of high quality, publicly accessible training data. Experimental datasets from real structures are costly to produce, narrow in scope, and rarely cover the full parameter space required for generalization. Synthetic dataset generation, leveraging parametric simulation environments, offers a practical path forward, provided the generation pipeline is reproducible, physically consistent, and sufficiently diverse in its coverage of the input space.

Y. DERRAZI
D. PELUFFO
J.C. TORRES

The application of machine learning to structural engineering can be traced back to the late 1980s, when Adeli and Yeh first employed perceptron networks for the selection and design of steel members [1]. Over the following decade, neural network architectures were extended to structural analysis, design optimization and damage detection, establishing the foundations of what has since become a broad and active research field [2]. Comprehensive reviews now document applications spanning structural health monitoring through computer vision [8, 9], seismic response prediction, and performance assessment [3, 4]. Recent developments have focused on improving the physical consistency and generalization of learned models. Physics informed neural networks (PINNs) incorporate governing equations directly into the training loss, reducing data requirements and improving extrapolation to unseen loading scenarios [11]; finite element informed neural networks embed FEM residuals as soft constraints, achieving similar benefits within a variational framework [10]. Graph neural networks (GNNs) have been applied to frame structures by representing members as edges and joints as nodes, naturally encoding the structural topology and enabling design predictions at the component level [12]. Surrogate modelling approaches, including ensemble methods trained on parametric finite element databases, have been used to approximate expensive simulations for tasks such as connection capacity prediction and embodied carbon optimization [15, 16], while generative models have accelerated the production of structural layouts for shear wall buildings [13, 14]. At the dataset level, several community initiatives have begun to standardize and share structural data, and reproducible synthetic datasets generated from parametric simulation pipelines have started to appear in fields such as computational biomechanics [18]; within structural engineering, however, publicly available benchmarks remain largely focused on specific domains such as earthquake engineering or crack detection from images [4, 3, 5].

Several gaps nevertheless persist. Diverse, high quality datasets for structural analysis are rarely made publicly available, limiting reproducibility and comparison across studies [4, 5]. Most existing databases focus on isolated members under standard boundary conditions (simply supported or fully clamped) without covering the continuum of intermediate elastic restraint that characterizes real structural connections [3, 16]. Furthermore,

many synthetic datasets are generated through ad hoc procedures that lack a fully automated and reproducible pipeline, making independent verification or extension difficult [18]. Finally, there is a shortage of datasets providing full spatial distributions of structural responses as opposed to scalar outputs, which limits their utility for architectures such as PINNs and GNNs that operate on spatially distributed fields [11, 10, 12].

This paper presents a dataset of 60000 synthetically generated TIMOSHENKO steel beam instances with general elastic supports, intended to address these gaps and serve as a training and benchmarking resource for data driven structural engineering applications. Each instance corresponds to a fully defined linear static beam problem, solved under one of six elementary unit load cases, with outputs comprising the full spatial distributions of kinematic quantities, internal forces and strain energy sampled at 21 equally spaced positions along the beam axis. The dataset was generated through a fully automated and reproducible pipeline, described in detail in Methods, combining LATIN HYPERCUBE Sampling with a self driving parametric graph implemented in GRASSHOPPER and KARAMBA3D [19].

2 Methods

The dataset was generated through a two-stage procedure. In the first stage, a mechanical model of a Timoshenko beam with general elastic supports is established, providing the theoretical foundation for the structural problems to be solved. In the second stage, a fully automated and reproducible data generation pipeline is described, encompassing the sampling strategy for input parameters, the numerical resolution of each beam problem via the KARAMBA3D finite element plug-in for GRASSHOPPER, and the retrieval and structuring of output quantities. Both stages are described in detail in the following subsections. Particular attention is given to reproducibility, ensuring that the dataset can be regenerated exactly from the provided inputs and scripts; this is discussed in detail afterwards.

2.1 Mechanical model

A solid body $\mathcal{S}$ is modelled as a straight beam of length L with a constant cross-section Σ of area S. The volume domain is the Cartesian product $[0, L] \times \Sigma$, and each material point is referenced with respect to the principal centroidal frame $(G, \vec{e}_x, \vec{e}_y, \vec{e}_z)$, where G is the centroid of Σ, $\vec{e}_x$ is the axis of the beam, and $\vec{e}_y$, $\vec{e}_z$ are the principal axes of inertia of Σ. Figure 1 summarizes the one-dimensional beam idealization and the associated coordinate frame.

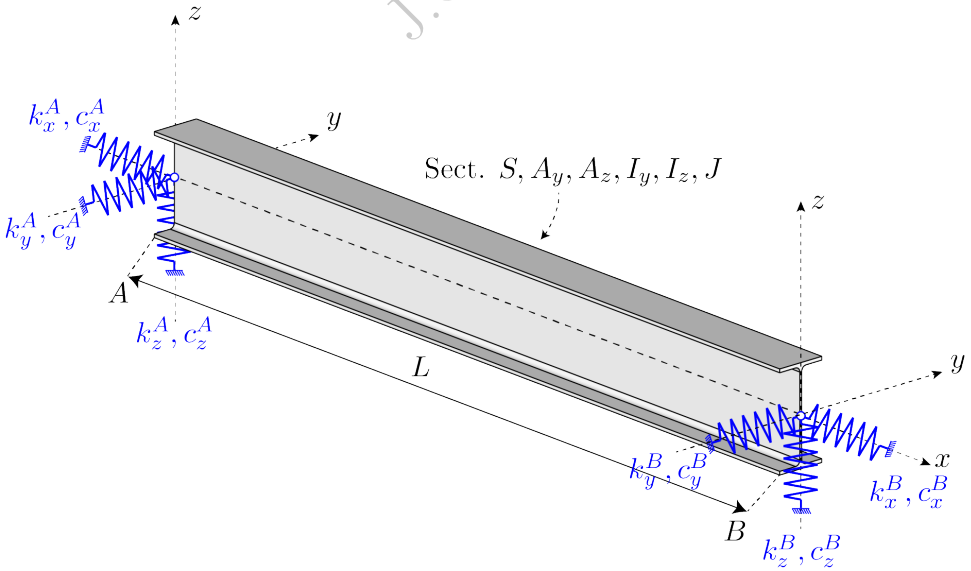

Figure 1: 1D beam model.

An arbitrary point M of the beam is therefore located by:

$$\overrightarrow{OM} = \overrightarrow{OG} + \overrightarrow{GM} = x\vec{e}_x + (y\vec{e}_y + z\vec{e}_z)\,, \tag{1}$$

where $x \in [0, L]$ is the axial coordinate and (y, z) are the cross-sectional coordinates. The cross-section Σ is fully characterized by the following geometric properties: the cross-sectional area S, the reduced shear areas A_y and A_z, the second moments of area I_{yy} and I_{zz}, and the torsion constant J_{xx}. The constitutive material is isotropic linear elastic steel, characterized by a YOUNG's modulus $E = 210\,\text{GPa}$ and a POISSON ratio $\nu = 0.33$, from which the shear modulus is derived as $G = \frac{E}{2(1+\nu)}$.

2.1.1 Kinematics

A TIMOSHENKO beam with uniform torsion is a beam element with six degrees of freedom per cross-section. The displacement of an arbitrary point M of the beam is written as:

$$\vec{u}(x, y, z) = \vec{u}_G(x) + \vec{\theta}(x) \times \overrightarrow{GM}, \tag{2}$$

where $\vec{u}_G = u_i^G \vec{e}_i$ is the centroidal displacement vector and $\vec{\theta} = \theta_i \vec{e}_i$ is the cross-section rotation vector about the principal centroidal axes. In decomposed form, this reads:

$$\vec{u}(x, y, z) = \begin{pmatrix} u_x^G(x) - y\theta_z(x) + z\theta_y(x) \\ u_y^G(x) - z\theta_x(x) \\ u_z^G(x) + y\theta_x(x) \end{pmatrix}. \tag{3}$$

The six independent kinematic fields are therefore the three centroidal displacements u_x^G, u_y^G, u_z^G and the three cross-section rotations θ_x, θ_y, θ_z, all functions of the axial coordinate x alone.

Euler-Bernoulli special case. Under the NAVIER-BERNOULLI hypothesis, cross-sections remain plane and perpendicular to the beam's neutral axis after deformation. This constrains the transverse rotations to:

$$\theta_z(x) = \frac{\mathrm{d}u_y^G}{\mathrm{d}x}, \qquad \theta_y(x) = -\frac{\mathrm{d}u_z^G}{\mathrm{d}x}, \tag{4}$$

reducing the element to four independent degrees of freedom. The TIMOSHENKO model relaxes this constraint by allowing independent rotation of the cross-section, thereby accounting for transverse shear deformation; an effect that becomes significant for deep sections and is retained throughout this work for generalization purposes.

Y. DURAZI
D. PELLEGRINO
J.C. TORRES

2.1.2 Beam model with elastic supports

Figure 2 illustrates the general elastic support model used at both beam ends.

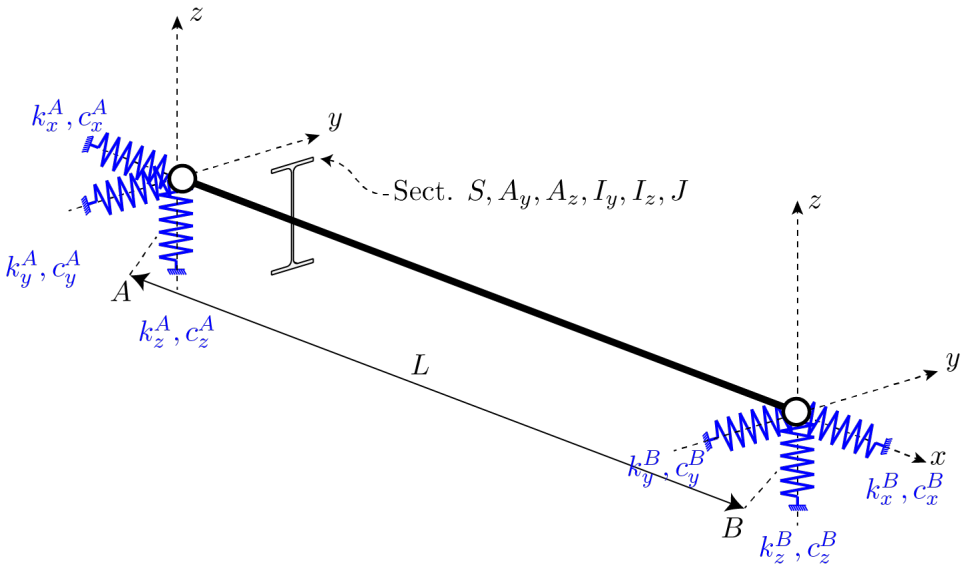

Figure 2: 1D beam model with elastic supports.

The mechanical model considered in this work is a TIMOSHENKO beam of length L and constant cross-section Σ, supported at both ends A and B by general elastic supports with six degrees of freedom each. The translational support stiffnesses are denoted $k_x^{A/B}$, $k_y^{A/B}$ and $k_z^{A/B}$, and the rotational support stiffnesses are denoted $c_x^{A/B}$, $c_y^{A/B}$ and $c_z^{A/B}$. These supports relate the nodal generalized displacements to their respective nodal generalized reactions through elementary HOOKE's law:

$$F_i^{A/B} = -k_i^{A/B} u_i^{A/B}, \qquad M_i^{A/B} = -c_i^{A/B} \theta_i^{A/B}, \tag{5}$$

where $i \in \{x, y, z\}$. This formulation encompasses the full range of boundary conditions encountered in classical structural analysis: free, pinned, roller, and fixed supports are recovered as limiting cases of vanishing or infinite stiffness, while intermediate values model the elastic restraint of adjacent structural elements.

Equilibrium equations. In the absence of distributed loads along the beam, the local equilibrium equations reduce to:

$$\frac{\mathrm{d}N}{\mathrm{d}x} = 0, \quad \frac{\mathrm{d}V_y}{\mathrm{d}x} = 0, \quad \frac{\mathrm{d}V_z}{\mathrm{d}x} = 0, \quad \frac{\mathrm{d}T}{\mathrm{d}x} = 0, \quad \frac{\mathrm{d}M_y}{\mathrm{d}x} - V_z = 0, \quad \frac{\mathrm{d}M_z}{\mathrm{d}x} + V_y = 0, \tag{6}$$

where N is the normal force, V_y and V_z are the transverse shear forces, T is the torsion moment, and M_y and M_z are the bending moments about the principal axes. These equations express that in the absence of distributed loads, the normal force, shear forces and torsion moment are constant along the beam, while the bending moments vary linearly.

Constitutive and kinematic equations. The local kinematic equations relate the generalized strains to the generalized displacements as follows:

$$\frac{\mathrm{d}u_x^G}{\mathrm{d}x} = \frac{N}{ES}, \qquad \frac{\mathrm{d}\theta_x}{\mathrm{d}x} = \frac{T}{GJ_{xx}}, \tag{7}$$

$$\frac{\mathrm{d}\theta_y}{\mathrm{d}x} = \frac{M_y}{EI_{yy}}, \qquad \frac{\mathrm{d}\theta_z}{\mathrm{d}x} = \frac{M_z}{EI_{zz}}, \tag{8}$$

$$\frac{\mathrm{d}u_y^G}{\mathrm{d}x} - \theta_z = \frac{V_y}{GA_y}, \qquad \frac{\mathrm{d}u_z^G}{\mathrm{d}x} + \theta_y = \frac{V_z}{GA_z}. \tag{9}$$

Y. DERRAZI
D. PELUFFO
J.C. TORRES

The last two equations are characteristic of the TIMOSHENKO model: the difference between the slope of the deflection curve and the cross-section rotation is proportional to the shear force, with the reduced shear areas A_y and A_z accounting for the non-uniform shear stress distribution over the cross-section. Combining the equilibrium and kinematic equations yields the governing differential equations for coupled bending in each principal plane:

$$EI_{zz} \frac{\mathrm{d}^2\theta_z}{\mathrm{d}x^2} + GA_y \left(\frac{\mathrm{d}u_y^G}{\mathrm{d}x} - \theta_z \right) = 0, \tag{10}$$

$$EI_{yy} \frac{\mathrm{d}^2\theta_y}{\mathrm{d}x^2} - GA_z \left(\frac{\mathrm{d}u_z^G}{\mathrm{d}x} + \theta_y \right) = 0, \tag{11}$$

whose integration constants are determined by the boundary conditions imposed by the elastic supports at A and B. The axial and torsional problems decouple entirely from the bending problems and from each other, and are governed by trivially integrable first-order equations given the equilibrium conditions above.

2.1.3 Load cases and superposition principle

Figure 3 shows the six elementary unit load cases considered in this work.

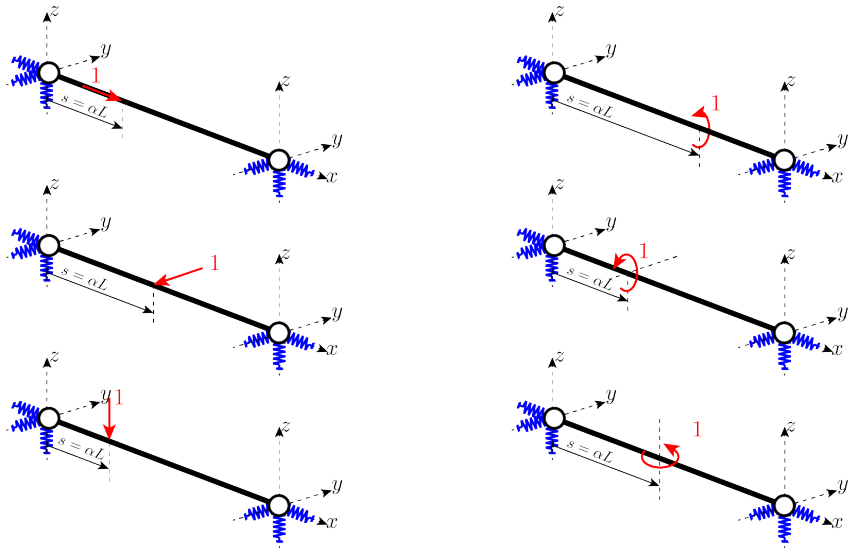

Figure 3: Elementary unit load cases.

Under the assumptions of geometric and material linearity, the principle of superposition holds, allowing any complex loading scenario to be decomposed into a combination of elementary load cases. Six elementary unit load cases are considered independently, corresponding to the three unit point loads $\vec{F} = \vec{e}_x$, $\vec{F} = \vec{e}_y$, $\vec{F} = \vec{e}_z$ and the three unit point moments $\vec{C} = \vec{e}_x$, $\vec{C} = \vec{e}_y$, $\vec{C} = \vec{e}_z$, each applied at an arbitrary relative position α along the beam axis.

Y. DERRAZI
D. PELUFFO
J.C. TORRES

The response of the beam to any generalized load per unit length $q(\alpha)$ can then be reconstructed by quadrature. If $Y(\alpha, x)$ denotes the beam's response to a unit excitation applied at relative position α, the response to an arbitrary distribution $q(\alpha)$ is given by:

$$y(x) = \int_0^1 q(\alpha)\, Y(\alpha, x)\, L\, \mathrm{d}\alpha. \tag{12}$$

This formulation would make the dataset directly usable as an influence function library, from which responses to arbitrary loading distributions can be reconstructed without additional solves.

Each elementary load case induces a specific and decoupled structural response by virtue of the equilibrium and kinematic equations established in the previous section:

- A unit axial load $\vec{F} = \vec{e}_x$ induces a state of tension/compression, with N and u_x^G as the non-trivial responses;
- A unit transverse load $\vec{F} = \vec{e}_y$ or a unit moment $\vec{C} = \vec{e}_z$ induces bending in the $(G, \vec{e}_y)$ plane, with V_y, M_z, u_y^G and θ_z as the non-trivial responses;
- A unit transverse load $\vec{F} = \vec{e}_z$ or a unit moment $\vec{C} = \vec{e}_y$ induces bending in the $(G, \vec{e}_z)$ plane, with V_z, M_y, u_z^G and θ_y as the non-trivial responses;
- A unit axial moment $\vec{C} = \vec{e}_x$ induces a state of torsion, with T and θ_x as the non-trivial responses.

2.1.4 Energy considerations

The total elastic strain energy stored in the beam under Timoshenko's beam theory writes:

$$U_D = \frac{1}{2} \int_0^L \left(\frac{N^2}{ES} + \frac{V_y^2}{GA_y} + \frac{V_z^2}{GA_z} + \frac{M_y^2}{EI_{yy}} + \frac{M_z^2}{EI_{zz}} + \frac{T^2}{GJ_{xx}} \right) \mathrm{d}x. \tag{13}$$

Accounting for the work performed by the elastic supports, the work-energy theorem reads:

$$2U_D = W_{ext} - W_k, \tag{14}$$

where W_k is the work stored in the elastic supports:

$$W_k = \sum_{j\in\{A,B\}} \sum_{i\in\{x,y,z\}} \left[k_i^j \left(u_i^j\right)^2 + c_i^j \left(\theta_i^j\right)^2 \right]. \tag{15}$$

The strain energy U_D will be included as an output quantity in the dataset and serves as a global scalar measure of the structural response, useful for energy-based validation and as a regression target for surrogate models.

2.1.5 Numerical finite element model

Let $\xi = x/L \in [0,1]$ denote the normalized axial coordinate along the beam. In accordance with the following nodal generalized displacement vector:

$$\boldsymbol{q}^T = \left[u_x^A\ u_x^B\ \theta_x^A\ \theta_x^B\ u_y^A\ u_y^B\ \theta_z^A\ \theta_z^B\ u_z^A\ u_z^B\ \theta_y^A\ \theta_y^B\right] \tag{16}$$

the degrees of freedom of said beam are in accordance with TIMOSHENKO's kinematic. The shape functions are introduced as follows:

$$\begin{aligned}
u_x^G(\xi) &= N_1^x(\xi)u_x^A + N_2^x(\xi)u_x^B && (17)\\
\theta_x(\xi) &= H_1^x(\xi)\theta_x^A + H_2^x(\xi)\theta_x^B && (18)\\
u_y^G(\xi) &= N_1^y(\xi)u_y^A + N_2^y(\xi)u_y^B + N_3^y(\xi)\theta_z^A + N_4^y(\xi)\theta_z^B && (19)\\
u_z^G(\xi) &= N_1^z(\xi)u_z^A + N_2^z(\xi)u_z^B + N_3^z(\xi)\theta_y^A + N_4^z(\xi)\theta_y^B && (20)\\
\theta_y(\xi) &= H_1^y(\xi)u_z^A + H_2^y(\xi)u_z^B + H_3^y(\xi)\theta_y^A + H_4^y(\xi)\theta_y^B && (21)\\
\theta_z(\xi) &= H_1^z(\xi)u_y^A + H_2^z(\xi)u_y^B + H_3^z(\xi)\theta_z^A + H_4^z(\xi)\theta_z^B && (22)
\end{aligned}$$

Y. DERRAZI
D. PELUHFO
J.C. TORRIS

Where:

$$N_1^x(\xi) = H_1^x(\xi) = 1 - \xi \tag{23}$$
$$N_2^x(\xi) = H_2^x(\xi) = \xi \tag{24}$$
$$N_1^y(\xi) = \frac{(1 - 3\xi^2 + 2\xi^3) + \Phi_y(1 - \xi)}{1 + \Phi_y} \tag{25}$$
$$N_2^y(\xi) = \frac{(3\xi^2 - 2\xi^3) + \Phi_y\xi}{1 + \Phi_y} \tag{26}$$
$$N_3^y(\xi) = \frac{L\xi(1 - \xi)(\Phi_y + 2 - 2\xi)}{2(1 + \Phi_y)} \tag{27}$$
$$N_4^y(\xi) = -\frac{L\xi(1 - \xi)(\Phi_y + 2\xi)}{2(1 + \Phi_y)} \tag{28}$$
$$H_1^z(\xi) = \frac{6\xi(\xi - 1)}{L(1 + \Phi_y)} \tag{29}$$
$$H_2^z(\xi) = -\frac{6\xi(\xi - 1)}{L(1 + \Phi_y)} \tag{30}$$
$$H_3^z(\xi) = \frac{(1 - \xi)(\Phi_y + 1 - 3\xi)}{1 + \Phi_y} \tag{31}$$
$$H_4^z(\xi) = \frac{\xi(\Phi_y + 3\xi - 2)}{1 + \Phi_y} \tag{32}$$
$$N_1^z(\xi) = \frac{(1 - 3\xi^2 + 2\xi^3) + \Phi_z(1 - \xi)}{1 + \Phi_z} \tag{33}$$
$$N_2^z(\xi) = \frac{(3\xi^2 - 2\xi^3) + \Phi_z\xi}{1 + \Phi_z} \tag{34}$$
$$N_3^z(\xi) = \frac{L\xi(1 - \xi)(\Phi_z + 2 - 2\xi)}{2(1 + \Phi_z)} \tag{35}$$
$$N_4^z(\xi) = \frac{L\xi(1 - \xi)(\Phi_z + 2\xi)}{2(1 + \Phi_z)} \tag{36}$$
$$H_1^y(\xi) = -\frac{6\xi(\xi - 1)}{L(1 + \Phi_z)} \tag{37}$$
$$H_2^y(\xi) = -\frac{6\xi(\xi - 1)}{L(1 + \Phi_z)} \tag{38}$$
$$H_3^y(\xi) = \frac{(1 - \xi)(\Phi_z + 1 - 3\xi)}{1 + \Phi_z} \tag{39}$$
$$H_4^y(\xi) = \frac{\xi(\Phi_z + 3\xi - 2)}{1 + \Phi_z} \tag{40}$$

Y. DERRIEN
D. FELIPPO
J. TORRES

In matrix-array form, we introduce the modal array $\boldsymbol{M}$ and therefore can write:

$$\vec{u}(x, y, z) = \boldsymbol{M}^T(y, z) \cdot \boldsymbol{d}(x), \tag{41}$$

where $\boldsymbol{d}$ is an array containing all the longitudinal degrees of freedom $(u_i^G(x), \theta_i(x))$ such as:

$$\boldsymbol{d}(x)^T = \left[u_x^G(x)\ \theta_x(x)\ u_y^G(x)\ \theta_z(x)\ u_z^G(x)\ \theta_y(x)\right] \tag{42}$$

Introducing shape functions array $\boldsymbol{N}$, we obtain:

$$\vec{u}(x, y, z) = \boldsymbol{M}^T(y, z)\boldsymbol{N}^T(x) \cdot \boldsymbol{q} \tag{43}$$

This allows to construct the stiffness matrix as follows using the virtual internal work:

$$W_{int}^* = -\boldsymbol{q}^{*T}\boldsymbol{K}\boldsymbol{q} \quad \Rightarrow \quad \boldsymbol{K} = \iiint_{M \in V} \boldsymbol{B}\boldsymbol{C}\boldsymbol{B}^T \mathrm{d}V, \tag{44}$$

where $\boldsymbol{B}$ is the geometric compatibility array, defined as follows:

$$\boldsymbol{B} = \boldsymbol{B}_L + \boldsymbol{B}_T, \tag{45}$$

where $\boldsymbol{B}_L$ and $\boldsymbol{B}_T$ are (respectively) the longitudinal and transverse compatibility arrays, found as follows:

$$\begin{aligned} \boldsymbol{B}_L(x,y,z) &= \boldsymbol{N}'(x)\boldsymbol{M}(y,z); & (46)\\ \boldsymbol{B}_T(x,y,z) &= \boldsymbol{N}(x)\boldsymbol{G}_T(y,z) & (47) \end{aligned}$$

where $\boldsymbol{G}_T$ is the generalized transverse gradient of the modal array written in a way that preserves compatibility conditions between displacements and strains.

Introducing the shear stiffness ratios:

$$\phi_y = 12\frac{EI_{zz}}{GA_yL^2}; \quad \phi_z = 12\frac{EI_{yy}}{GA_zL^2}; \tag{48}$$

the Timoshenko beam stiffness matrix takes the following block-diagonal form:

$$\boldsymbol{K} = \left[\begin{array}{c|c|c} \boldsymbol{K}_x & \boldsymbol{0} & \boldsymbol{0} \\ \hline \boldsymbol{0} & \boldsymbol{K}_y & \boldsymbol{0} \\ \hline \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{K}_z \end{array}\right] \tag{49}$$

where:

$$\boldsymbol{K}_x = \begin{array}{c} \begin{array}{cccc} u_x^A & u_x^B & \theta_x^A & \theta_x^B \end{array} \\ \left[\begin{array}{cccc} \frac{ES}{L} & -\frac{ES}{L} & 0 & 0 \\ -\frac{ES}{L} & \frac{ES}{L} & 0 & 0 \\ 0 & 0 & \frac{GJ_{xx}}{L} & -\frac{GJ_{xx}}{L} \\ 0 & 0 & -\frac{GJ_{xx}}{L} & \frac{GJ_{xx}}{L} \end{array}\right] \end{array} \begin{array}{c} \\ u_x^A \\ u_x^B \\ \theta_x^A \\ \theta_x^B \end{array} \tag{50}$$

$$\boldsymbol{K}_y = \frac{1}{1+\phi_y} \begin{array}{c} \begin{array}{cccc} u_y^A & u_y^B & \theta_z^A & \theta_z^B \end{array} \\ \left[\begin{array}{cccc} 12\frac{EI_{zz}}{L^3} & -12\frac{EI_{zz}}{L^3} & 6\frac{EI_{zz}}{L^2} & 6\frac{EI_{zz}}{L^2} \\ -12\frac{EI_{zz}}{L^3} & 12\frac{EI_{zz}}{L^3} & -6\frac{EI_{zz}}{L^2} & -6\frac{EI_{zz}}{L^2} \\ 6\frac{EI_{zz}}{L^2} & -6\frac{EI_{zz}}{L^2} & (4+\phi_y)\frac{EI_{zz}}{L} & (2-\phi_y)\frac{EI_{zz}}{L} \\ 6\frac{EI_{zz}}{L^2} & -6\frac{EI_{zz}}{L^2} & (2-\phi_y)\frac{EI_{zz}}{L} & (4+\phi_y)\frac{EI_{zz}}{L} \end{array}\right] \end{array} \begin{array}{c} \\ u_y^A \\ u_y^B \\ \theta_z^A \\ \theta_z^B \end{array} \tag{51}$$

$$\boldsymbol{K}_z = \frac{1}{1+\phi_z} \begin{array}{c} \begin{array}{cccc} u_z^A & u_z^B & \theta_y^A & \theta_y^B \end{array} \\ \left[\begin{array}{cccc} 12\frac{EI_{yy}}{L^3} & -12\frac{EI_{yy}}{L^3} & -6\frac{EI_{yy}}{L^2} & -6\frac{EI_{yy}}{L^2} \\ -12\frac{EI_{yy}}{L^3} & 12\frac{EI_{yy}}{L^3} & 6\frac{EI_{yy}}{L^2} & 6\frac{EI_{yy}}{L^2} \\ -6\frac{EI_{yy}}{L^2} & 6\frac{EI_{yy}}{L^2} & (4+\phi_z)\frac{EI_{yy}}{L} & (2-\phi_z)\frac{EI_{yy}}{L} \\ -6\frac{EI_{yy}}{L^2} & 6\frac{EI_{yy}}{L^2} & (2-\phi_z)\frac{EI_{yy}}{L} & (4+\phi_z)\frac{EI_{yy}}{L} \end{array}\right] \end{array} \begin{array}{c} \\ u_z^A \\ u_z^B \\ \theta_y^A \\ \theta_y^B \end{array} \tag{52}$$

The generalized external force array is defined such as:

$$W_{ext}^* = \boldsymbol{q}^{*T}\boldsymbol{F}_{ext} \quad \Rightarrow \quad \boldsymbol{F}_{ext} = \iiint_{M\in V} \boldsymbol{N}\boldsymbol{M}\cdot\vec{f}^T \mathrm{d}V. \tag{53}$$

Given the elastic supports on both ends, reaction components are assembled in a nodal array as follows:

$$\boldsymbol{R} = -\boldsymbol{K}_{ext}\cdot\boldsymbol{q} \tag{54}$$

where:

Y. DERRAZI
D. CELUFFO
J.C. TORRES

$$\boldsymbol{K}_{ext} = \begin{array}{c} \begin{array}{cccccccccccc} u_x^A & u_x^B & \theta_x^A & \theta_x^B & u_y^A & u_y^B & \theta_z^A & \theta_z^B & u_z^A & u_z^B & \theta_y^A & \theta_y^B \end{array} \\ \begin{bmatrix} k_x^A & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & k_x^B & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & c_x^A & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & c_x^B & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & k_y^A & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & k_y^B & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & c_z^A & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_z^B & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & k_z^A & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & k_z^B & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_y^A & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_y^B \end{bmatrix} \begin{array}{c} u_x^A \\ u_x^B \\ \theta_x^A \\ \theta_x^B \\ u_y^A \\ u_y^B \\ \theta_z^A \\ \theta_z^B \\ u_z^A \\ u_z^B \\ \theta_y^A \\ \theta_y^B \end{array} \end{array} \tag{55}$$

the system to solve becomes:

$$\boldsymbol{F}_{ext} = (\boldsymbol{K} + \boldsymbol{K}_{ext}) \cdot \boldsymbol{q}. \tag{56}$$

Since $\boldsymbol{K}$ is singular of order six by definition, the elastic support stiffnesses are chosen such that $\boldsymbol{K} + \boldsymbol{K}_{ext}$ is invertible, which is guaranteed by the sampling strategy described in the following sections.

2.2 Data Generation Pipeline

The data generation pipeline translates the mechanical model established in the previous sections into a fully automated and reproducible sequence of parametric structural solves. Each instance in the dataset corresponds to a unique beam problem, defined by a specific combination of geometric, cross-sectional, loading, and boundary condition parameters. The pipeline consists of four stages: the definition of the input feature space, the sampling strategy used to populate it, the automated resolution of each beam problem via Karamba3D within a Grasshopper parametric graph, and the structuring of the output feature space. Each stage is described in detail in the following subsections.

Y. Derradji
D. Peluffo
J.C. Torres

2.2.1 Input feature space

Each data instance is fully characterized by an input feature vector assembling all parameters that define a unique beam problem. The base geometric and boundary condition parameters are:

$$\boldsymbol{X} = [L, S, A_y, A_z, I_{yy}, I_{zz}, J_{xx}, k_x^A, k_y^A, k_z^A, k_x^B, k_y^B, k_z^B, c_x^A, c_y^A, c_z^A, c_x^B, c_y^B, c_z^B]. \tag{57}$$

where L is the beam length, α is the relative position of the applied unit load along the beam axis, S, A_y, A_z, I_{yy}, I_{zz}, J_{xx} are the cross-sectional geometric properties, and $k_i^{A/B}$, $c_i^{A/B}$ are the translational and rotational elastic support stiffnesses at ends A and B respectively. The feature vector is completed with the load case identifier $N_c \in \{1, \ldots, 6\}$, as well as cross-section metadata comprising a section identifier `SecID`, a family label `SecFamily` and a section name `SecName`, yielding the full input vector:

$$\boldsymbol{X} = [N_c, \texttt{SecID}, \texttt{SecFamily}, \texttt{SecName}, L, \alpha, S, A_y, A_z, I_{yy}, I_{zz}, J_{xx}, k_x^A, k_y^A, k_z^A, k_x^B, k_y^B, k_z^B, c_x^A, c_y^A, c_z^A, c_x^B, c_y^B, c_z^B], \tag{58}$$

The cross-section metadata fields are qualitative and serve identification and exploratory analysis purposes only; they are not intended as direct inputs to machine learning models. The dataset is restricted to IPE, HEA and HEB European steel profiles, whose geometric properties span a representative range of sections commonly used in structural engineering practice. The complete list of sections included is given in Table 1.

2.2.2 Sampling strategy

A total of $N = 60000$ beam instances were generated, with exactly 10000 instances assigned to each of the six elementary load cases. Sampling was performed in a normalized design space and subsequently mapped to physically meaningful quantities through bijective transformations, ensuring full traceability between the normalized samples and their physical counterparts.

Latin Hypercube Sampling. To ensure efficient and uniform coverage of the high-dimensional input space, Latin Hypercube Sampling (LHS) was adopted with a fixed seed (`seed = 2`), guaranteeing exact reproducibility of the dataset. LHS divides the normalized interval $[0, 1]$ into N equiprobable strata along each input dimension, ensuring that each stratum is sampled exactly once. Let $\boldsymbol{x} \in [0, 1]^d$ denote the d-dimensional normalized input vector, with $d = 15$ free parameters. Each component x_i is independently stratified and randomly permuted, yielding a space-filling design without the clustering typical of purely random Monte Carlo methods, with improved marginal uniformity and reduced variance.

Beam length. The beam length L is mapped from its normalized sample $x_L \in [0, 1]$ through linear scaling onto a discrete grid:

$$L = \mathcal{G}(x_L,\, L_{min},\, L_{max},\, \Delta L), \tag{59}$$

where $\mathcal{G}$ denotes the grid discretization operator, $L_{min} = 1\,\text{m}$, $L_{max} = 10\,\text{m}$, and $\Delta L = 0.05\,\text{m}$, yielding 181 admissible values.

Relative load position. The relative load position α is mapped analogously:

$$\alpha = \mathcal{G}(x_\alpha,\, \alpha_{min},\, \alpha_{max},\, \Delta\alpha), \tag{60}$$

with $\alpha_{min} = 0.05$, $\alpha_{max} = 0.95$, and $\Delta\alpha = 0.05$, yielding 19 admissible values. The bounds are chosen to avoid degenerate configurations where the load is applied at a beam endpoint.

Cross-section selection. Cross-sections are selected from a predefined catalogue of $n_s = 66$ European steel profiles spanning the IPE, HEA and HEB families. Each section is associated with a unique identifier `SecID` and a set of tabulated geometric properties $[S, A_y, A_z, I_{yy}, I_{zz}, J_{xx}]$. Selection is performed through index-based sampling:

$$\texttt{SecID} = \mathcal{C}\left(\lfloor x_s n_s \rfloor\right), \tag{61}$$

where $\mathcal{C}$ is the catalogue mapping from a zero-based index to a valid section identifier. HEAA profiles, which occupy indices 18 to 41 in KARAMBA3D's internal section library, are deliberately excluded to avoid over-representation of HEA-family sections in the dataset.

V. DERRAZI
D. DE CEFO
J.C. TORRES

Table 1: Cross-section catalogue used in the dataset.

Family	**IPE**	**HEA**	**HEB**
Count	18	24	24
Frequency (%)	27.28	36.36	36.36
First section	IPE80	HEA100	HEB100
Last section	IPE600	HEA1000	HEB1000

Above in Table 1 are all the cross sections that are included in the dataset : the disparity in section count between IPE and HEA/HEB families is reflected in the marginal distribution of cross-sectional properties.

Elastic support stiffness sampling. Translational and rotational support stiffnesses span several orders of magnitude in order to capture the full continuum between free, elastically restrained, and quasi-clamped boundary conditions. Given this wide dynamic range, stiffness parameters are sampled in logarithmic space through a three-step procedure.

First, a normalized exponent $m \in [m_{min}, m_{max}]$ is drawn for each stiffness degree of freedom via LHS:

$$m = m_{min} + x_m(m_{max} - m_{\min}), \tag{62}$$

with $m_{min} = -1$ and $m_{max} = 5$. The exponent is then transformed into a dimensionless stiffness multiplier:

$$\kappa = 10^m \quad \text{(translational)}, \qquad \zeta = 10^m \quad \text{(rotational)}, \tag{63}$$

spanning six decades from 10^{-1} to 10^{5}. Finally, the physical stiffnesses are constructed proportionally to the characteristic stiffness quantities of the beam:

$$k_x = \kappa_x \frac{ES}{L}, \qquad k_y = \kappa_y \frac{EI_{zz}}{L^3}, \qquad k_z = \kappa_z \frac{EI_{yy}}{L^3}, \tag{64}$$

$$c_x = \zeta_x \frac{GJ_{xx}}{L}, \qquad c_y = \zeta_y \frac{EI_{yy}}{L}, \qquad c_z = \zeta_z \frac{EI_{zz}}{L}. \tag{65}$$

This formulation ensures that boundary conditions are defined relative to the intrinsic stiffness of the member rather than in absolute terms, guaranteeing physical consistency across all sampled configurations. The dimensionless multipliers κ and ζ therefore control the ratio between support stiffness and member stiffness: $\kappa \ll 1$ corresponds to nearly free conditions, $\kappa \approx 1$ represents support stiffness comparable to member stiffness, and $\kappa \gg 1$ approximates clamped conditions. The same interpretation holds for ζ. The stiffness multipliers κ and ζ are stored as input features in the dataset, with the physical stiffnesses k and c computed internally within the KARAMBA3D pipeline as described in the following subsection.

2.2.3 Grasshopper/Karamba3D pipeline

Each beam problem defined by an input feature vector $\boldsymbol{X}$ is solved numerically using KARAMBA3D, a finite element plug-in for the GRASSHOPPER parametric modelling environment within RHINOCEROS 3D. The resolution of all $N = 60000$ instances is handled by a fully automated parametric graph.

Input parsing and stiffness computation. A pre-generated input CSV file, produced by the Python sampling script described in Section 2.2.2, is loaded into the graph at initialization. A row parser component unpacks each input row into its constituent parameters: load case identifier N_c, section identifier `SecID`, beam length L, relative load position α, and the twelve dimensionless stiffness multipliers $\kappa_i^{A/B}$ and $\zeta_i^{A/B}$. A dedicated stiffness computation component retrieves the cross-sectional geometric properties associated with `SecID` from KARAMBA3D's internal section library. Since KARAMBA3D returns cross-sectional areas and second moments of area in cm^2 and cm^4 respectively, these are converted to m^2 and m^4 prior to stiffness computation. Material properties are set to $E = 210\,\mathrm{GPa}$ and $\nu = 0.33$, from which the shear modulus is derived. The physical support stiffnesses $k_i^{A/B}$ and $c_i^{A/B}$ are then computed according to the scaling relations defined in Section 2.2.2.

Structural model assembly. For each input row, the structural model is assembled through the following sequence of native KARAMBA3D components:

1. *Geometry*: a line element of length L is created along the beam axis;
2. *Cross-section*: the section corresponding to `SecID` is selected from the European profile library;
3. *Material*: standard steel with $E = 210\,\mathrm{GPa}$ and $\nu = 0.33$ is assigned by default;
4. *Boundary conditions*: elastic supports with the computed stiffnesses $k_i^{A/B}$ and $c_i^{A/B}$ are applied at both ends A and B;
5. *Loading*: a unit point load or unit point moment corresponding to load case N_c is applied at relative position α;
6. *Assembly and analysis*: the model is assembled and solved via KARAMBA3D's linear static solver, yielding the nodal displacement vector $\boldsymbol{q}$ as the solution to equation 56.

Result retrieval. Kinematic quantities $u_i^G(x)$ and $\theta_i(x)$ and internal force quantities $N(x)$, $V_y(x)$, $V_z(x)$, $T(x)$, $M_y(x)$, $M_z(x)$ are retrieved at 21 equally spaced positions $\xi \in \{0.00, 0.05, \ldots, 1.00\}$ using KARAMBA3D's results-at-point query components, one for kinematic quantities and one for internal force quantities. The strain energy U_D is retrieved as a scalar output of the analysis component.

Y. Dimitrov
D. Pitchforth
J. C. Torres

Automation and reproducibility. The iteration over all $N = 60000$ input rows is handled by a self-driving index driver component operating over a configurable range $[i_{\text{start}}, i_{\text{end}}]$, allowing the pipeline to be run over arbitrary sub-ranges of the dataset and resumed after interruption without data loss. The component maintains a persistent row counter in Rhinoceros 3D's application-level sticky dictionary and explicitly handles three states: running, paused, and done. At each iteration, the counter is incremented and the next solution is scheduled via a callback:

```
def callback(d):
    ghenv.Component.ExpireSolution(False)
doc.ScheduleSolution(1, callback)
```

causing the component to expire and recompute autonomously after each solve. A reset flag allows the sticky state to be cleared for a clean restart. Combined with the fixed LHS seed and deterministic parameter mappings, this mechanism guarantees that the dataset can be regenerated identically from the provided input CSV and Grasshopper definition.

Result structuring and output. After each solve, a result structuring component assembles the retrieved quantities into a structured output row indexed by `row_id`. Alongside the structural response quantities, the physical cross-sectional properties S, A_y, A_z, I_{yy}, I_{zz}, J_{xx} and the physical support stiffnesses $k_i^{A/B}$, $c_i^{A/B}$ are written into the output row, making the output file self-contained with respect to the mechanical parameters of each beam problem. A deduplication mechanism based on a persistent set of processed indices prevents duplicate entries in the event of spurious component recomputations. Rather than accumulating all instances in memory before writing, rows are buffered and automatically flushed to disk in chunks of 5000 instances. After the first flush, subsequent writes operate in append mode, ensuring that results are incrementally preserved and that no data is lost in the event of an interruption. The CSV header is written exactly once on the first flush. Output columns follow a consistent naming convention of the form $\{\texttt{quantity}\}\{\xi\}$, where `quantity` is one of `ux`, `uy`, `uz`, `rx`, `ry`, `rz`, `Nx`, `Vy`, `Vz`, `Tx`, `My`, `Mz`, and $\xi \in \{0.00, 0.05, \ldots, 1.00\}$ is the normalized sampling position formatted to two decimal places. The strain energy U_D is stored as a scalar column `Ud`.

V. Derrazi
D. Peluffo
J. C. Torres

2.2.4 Output feature space

For each beam problem defined by an input feature vector $\boldsymbol{X}$, the output file assembles the physical parameters computed internally by the pipeline together with all quantities of interest resulting from the linear static analysis. The physical cross-sectional properties S, A_y, A_z, I_{yy}, I_{zz}, J_{xx} and the physical support stiffnesses $k_i^{A/B}$, $c_i^{A/B}$ are stored alongside the structural response quantities, making the output file self-contained with respect to the mechanical parameters required to fully define the beam problem.

The structural response comprises the six kinematic fields and six internal force fields, each sampled at $n_\xi = 21$ equally spaced normalized positions $\xi \in \{0.00, 0.05, \ldots, 1.00\}$, together with the scalar strain energy U_D:

$$\begin{aligned}\boldsymbol{Y} = \Big[& U_D,\ u_x^{0.00}, \ldots, u_x^{1.00},\ u_y^{0.00}, \ldots, u_y^{1.00},\ u_z^{0.00}, \ldots, u_z^{1.00},\ \theta_x^{0.00}, \ldots, \theta_x^{1.00},\ \theta_y^{0.00}, \ldots, \theta_y^{1.00}, \\ & \theta_z^{0.00}, \ldots, \theta_z^{1.00},\ N^{0.00}, \ldots, N^{1.00},\ V_y^{0.00}, \ldots, V_y^{1.00},\ V_z^{0.00}, \ldots, V_z^{1.00},\ T^{0.00}, \ldots, T^{1.00}, \\ & M_y^{0.00}, \ldots, M_y^{1.00},\ M_z^{0.00}, \ldots, M_z^{1.00} \Big],\end{aligned} \tag{66}$$

yielding a response vector of length $12 \times 21 + 1 = 253$. A summary of the output quantities, their physical interpretation, units and column name prefixes is given in Table 2.

Table 2: Summary of output quantities.

Quantity	Description	Unit	Column prefix
U_D	Elastic strain energy	kN·m	`Ud`
u_x^G	Axial displacement	m	`ux`
u_y^G	Transverse displacement (y)	m	`uy`
u_z^G	Transverse displacement (z)	m	`uz`
θ_x	Torsional rotation	rad	`rx`
θ_y	Bending rotation (y-plane)	rad	`ry`
θ_z	Bending rotation (z-plane)	rad	`rz`
N	Normal force	kN	`Nx`
V_y	Shear force (y-direction)	kN	`Vy`
V_z	Shear force (z-direction)	kN	`Vz`
T	Torsion moment	kN·m	`Tx`
M_y	Bending moment (y-axis)	kN·m	`My`
M_z	Bending moment (z-axis)	kN·m	`Mz`

By virtue of the constitutive decoupling, each load case N_c activates only a specific subset of output quantities, with the remaining fields being identically zero. This sparsity structure is an intrinsic property of the dataset and should be accounted for in any machine learning model trained on it.

3 Data Records

The dataset is distributed as two CSV files, joinable on a common identifier column `ID` = `row_id`.

Input file (`TimoDS_input_data_VX.csv`). This file contains $N = 60{,}000$ rows and the following columns: `ID`, `Lcase`, `SecID`, `SecFamily`, `SecName`, `L`, `a`, κ_x^A, κ_y^A, κ_z^A, κ_x^B, κ_y^B, κ_z^B, ζ_x^A, ζ_y^A, ζ_z^A, ζ_x^B, ζ_y^B, ζ_z^B. This file is fully determined by the Python sampling script and the fixed Latin Hypercube Sampling seed, and can be regenerated exactly from the provided code.

Y. DERRAZI
D. PINEDO
J.C. TORRES

Output file (`TimoDS_beam_results_VX.csv`). This file contains $N = 60000$ rows and 272 columns, comprising the instance identifier `row_id`, the six physical cross-sectional properties S, A_y, A_z, I_{yy}, I_{zz}, J_{xx}, the twelve physical support stiffnesses $k_i^{A/B}$ and $c_i^{A/B}$, the scalar strain energy `Ud`, and the 252 sampled response columns described in Section 2.2.4. Units are metres and kilonewtons throughout, consistent with the material and geometric parameters defined in Section 2.2.3.

Analysis dataframe. For the purposes of technical validation and exploratory data analysis, a merged analysis dataframe is constructed by joining the input and output files on `ID` = `row_id`, and enriching the result with the following derived quantities: load case identifier, section metadata (`SecID`, `SecFamily`, `SecName`), beam length L, relative load position α, physical cross-sectional properties and support stiffnesses, all sampled response fields, per-instance minimum, maximum and maximum absolute values of each response quantity, and the external work W_{ext}. This dataframe is used exclusively for validation and characterization purposes and is not part of the distributed dataset.

Data repository. Both files are made publicly available at https://doi.org/10.5281/zenodo.19324944. The Python sampling script and the Grasshopper pipeline definition are made available at https://github.com/YDERRAZI/TimoDS_pipeline .

4 Data Overview and technical validation

The following section provides a comprehensive characterization of the dataset from both statistical and physical standpoints. The input space is first examined to verify that the sampling strategy produces the intended coverage and decorrelation properties.

The output space is then characterized to establish the range and distribution of structural responses across the dataset. Finally, physical consistency is verified at the level of individual instances and at the level of global energy balance, the latter leading to the strain energy correction procedure.

4.1 Statistical characterization of the input space

4.1.1 Section family distribution

The dataset comprises 66 European steel profiles drawn from three families: 18 IPE, 24 HEA and 24 HEB sections. Since each section is sampled with equal probability within its family, and the three families contain different numbers of profiles, the resulting instance counts are not equal across families.

Over $N = 60000$ instances, HEA and HEB each account for approximately 21600 instances (36.36%) while IPE accounts for approximately 16400 instances (27.28%), as shown in Figure 4. This disparity is an expected consequence of the catalogue composition.

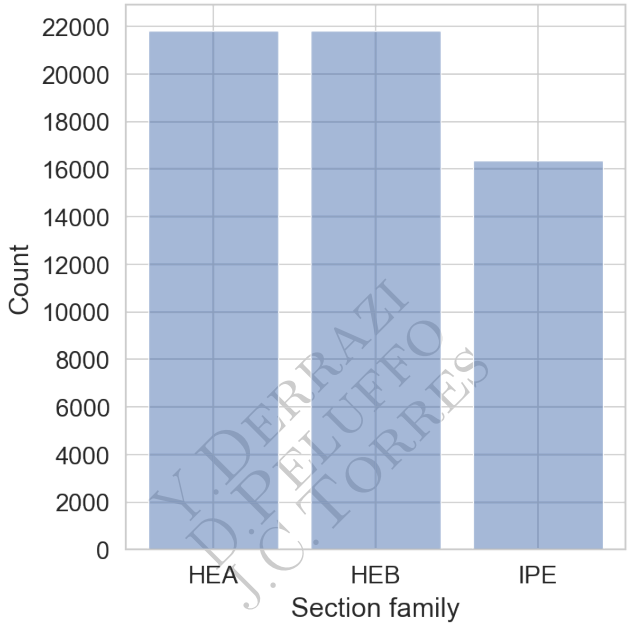

Figure 4: Section family distribution.

4.1.2 Decorrelation of sampling parameters

A fundamental requirement of the dataset is that the geometric, loading and boundary condition parameters are mutually independent, so that no spurious correlations are introduced between input dimensions. Figures 5 and 6 show the distributions of beam length L and relative load position α stratified by section family. The section family proportion is identical at every length bin and every α bin, confirming that section type, beam length and load position are mutually independent. Figure 7 further confirms the decorrelation between L and α — every length bin contains an identical uniform distribution of load positions, with all 19 admissible α values equally represented. The slight excess at $L = 10$m is a grid boundary artifact with no physical significance.

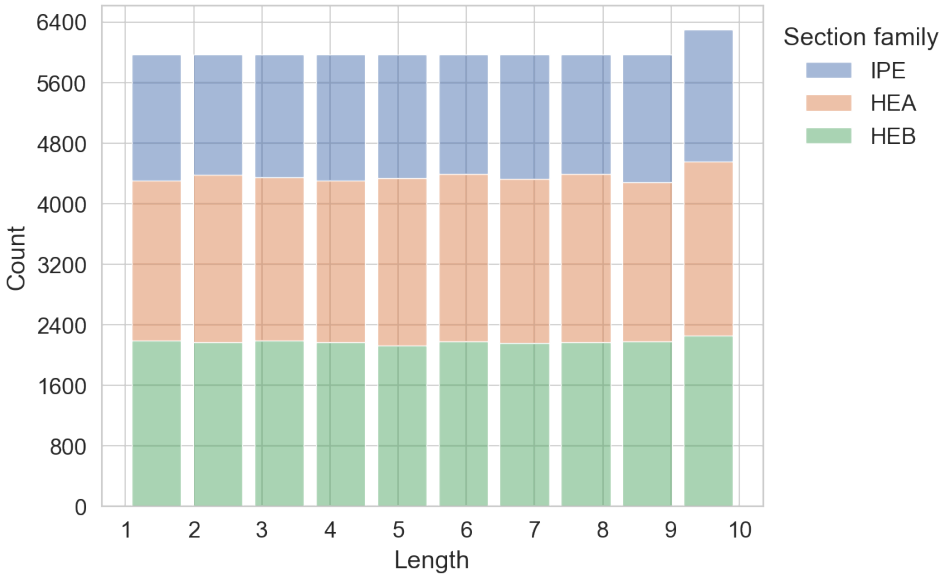

Figure 5: Length distribution hued by section family.

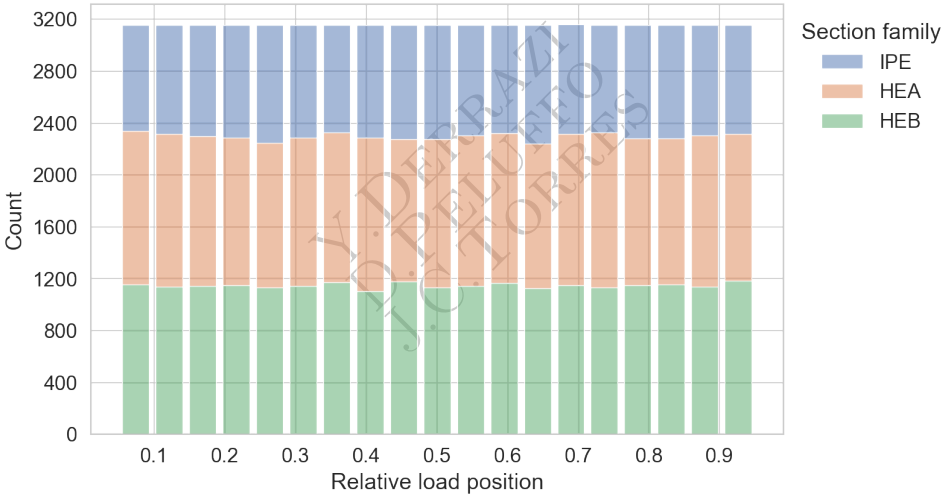

Figure 6: Load position distribution hued by section family.

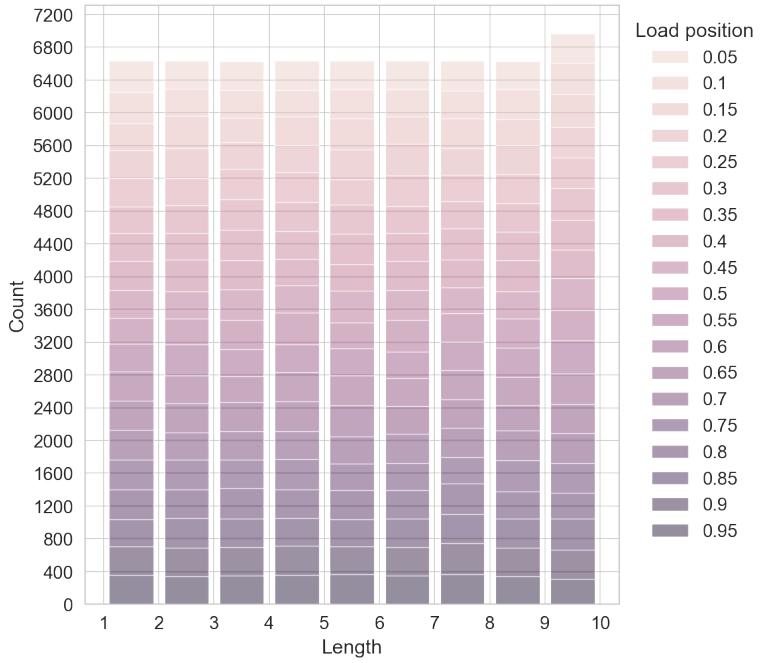

Figure 7: Length distribution hued by load position.

4.1.3 Physical stiffness distributions

The physical support stiffnesses $k_i^{A/B}$ and $c_i^{A/B}$ result from the product of the dimensionless LHS-sampled multipliers κ and ζ with the characteristic stiffness quantities of each beam instance. Their distributions are shown in Figures 8, 9, 10, 11, 12 and 13. All six stiffness pairs exhibit bell-shaped distributions on the log scale, consistent with uniform sampling of the exponent $\ell \in [-1, 5]$ followed by the 10^ℓ transformation and multiplication by the characteristic stiffness. The A and B end distributions overlap nearly perfectly for all stiffness types, confirming that the two ends are sampled independently with no systematic asymmetry.

The absolute ranges of the physical stiffnesses differ significantly across stiffness types, reflecting the different characteristic scalings. The axial translational stiffness $k_x^{A/B}$, governed by ES/L, spans the widest absolute range ($\sim 10^3$ to 10^{12} kN/m) due to the large axial rigidity of steel sections. The transverse bending stiffnesses $k_y^{A/B}$ and $k_z^{A/B}$, governed by EI_{zz}/L^3 and EI_{yy}/L^3 respectively, are shifted several decades lower owing to the cubic length effect, spanning roughly 10^{-2} to 10^8 kN/m. The rotational stiffnesses $c_i^{A/B}$ occupy intermediate ranges consistent with their GJ_{xx}/L and EI/L scalings. These differences are entirely by design and ensure that boundary conditions are sampled in physically meaningful proportion to the beam's intrinsic stiffness in each degree of freedom.

The mild positive correlation between k_i^A and k_i^B observed in the correlation heatmap (Figure 14) is explained by their shared characteristic stiffness scaling; for a given beam instance, both end stiffnesses are multiplied by the same EI/L^3 or ES/L factor, introducing a correlation that is entirely attributable to the beam geometry rather than to any dependence between the sampled multipliers κ_i^A and κ_i^B.

V. DERRAZI
D. PELUFFO
J. C. TORRES

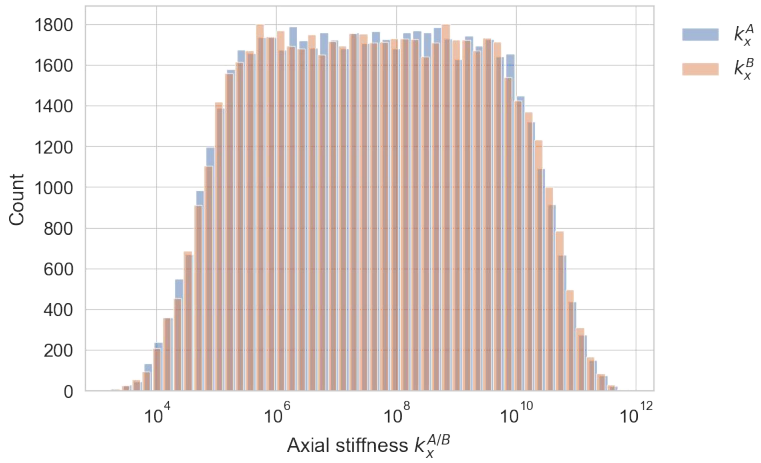

Figure 8: $k_x^{A/B}$ overlapped histogram.

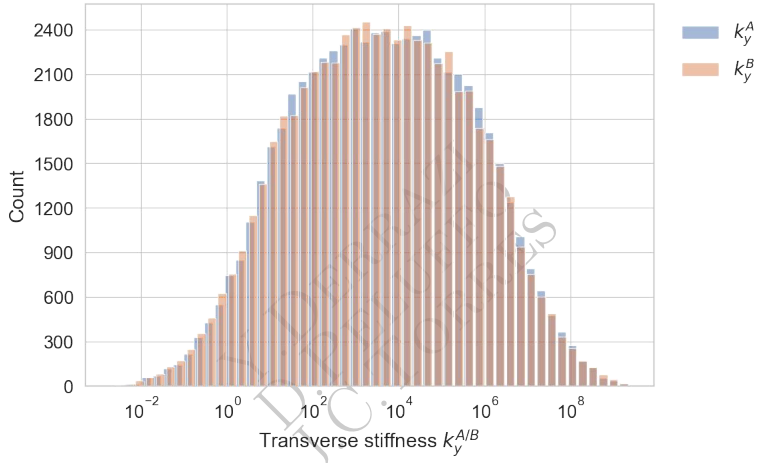

Figure 9: $k_y^{A/B}$ overlapped histogram.

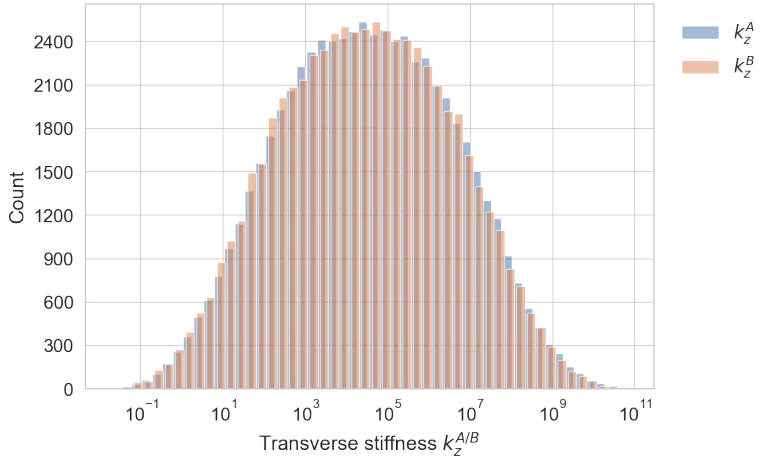

Figure 10: $k_z^{A/B}$ overlapped histogram.

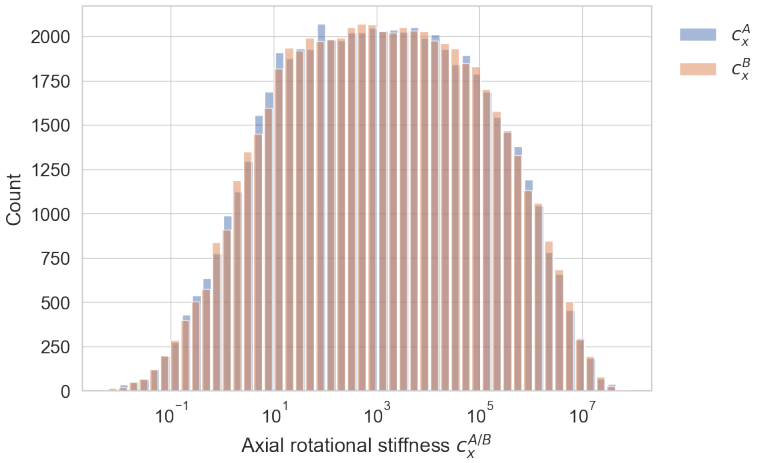

Figure 11: $c_x^{A/B}$ overlapped histogram.

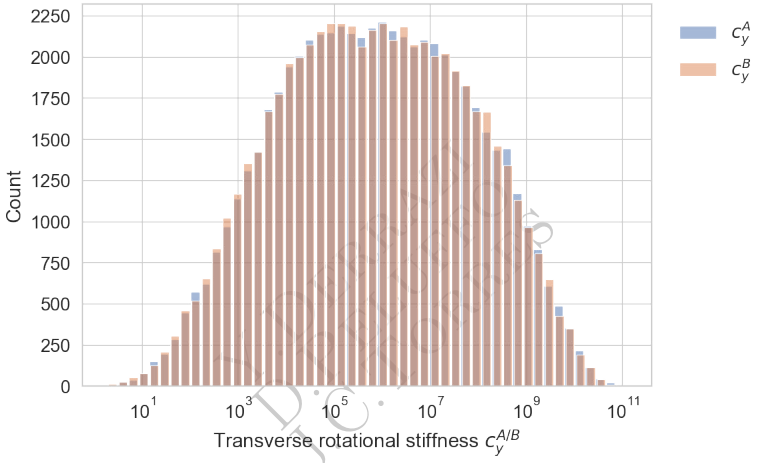

Figure 12: $c_y^{A/B}$ overlapped histogram.

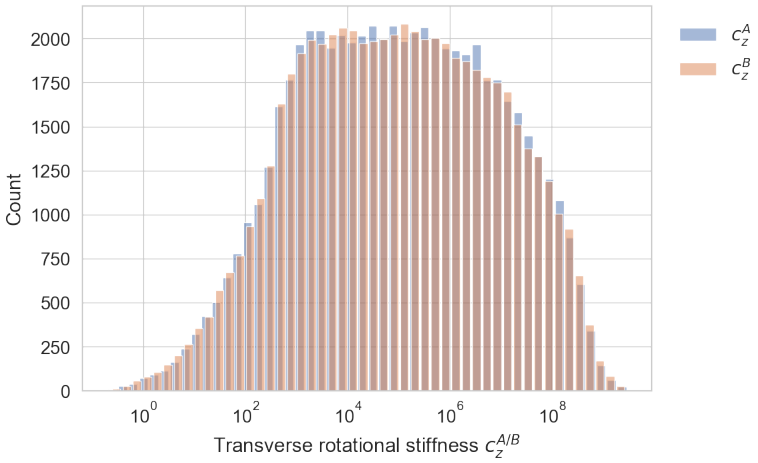

Figure 13: $c_z^{A/B}$ overlapped histogram.

4.1.4 Correlation structure of physical inputs

Figure 14 shows the Pearson correlation matrix of all physical input parameters: L, α, the six cross-sectional geometric properties S, A_y, A_z, I_{yy}, I_{zz}, J_{xx}, and the twelve physical support stiffnesses $k_i^{A/B}$, $c_i^{A/B}$. Several distinct correlation blocks are visible and admit precise physical explanations.

The beam length L and relative load position α are effectively uncorrelated with each other and with all other parameters, confirming the independence of the LHS sampling in these dimensions. The cross-sectional geometric properties form a tightly correlated block, with pairwise correlations exceeding 0.8 in most cases. This is not a sampling artifact but a structural property of the European steel profile catalogue: within each profile family, section dimensions scale proportionally, so that larger sections have larger area, larger shear areas, larger second moments of area and larger torsion constants simultaneously.

A moderate negative correlation of approximately -0.2 is observed between L and all physical stiffness parameters. This is a direct consequence of the relative stiffness sampling formulation: since $k_x \propto ES/L$, $k_y \propto EI_{zz}/L^3$ and so on, longer beams systematically produce smaller physical stiffnesses for the same dimensionless multiplier. The effect is more pronounced for the translational bending stiffnesses k_y and k_z (governed by $1/L^3$) than for k_x (governed by $1/L$), though the difference is partially masked by the section property variation.

The physical stiffnesses show a moderate positive correlation of approximately 0.1–0.3 with the cross-sectional geometric properties, again a direct consequence of the scaling formulation. Rotational stiffnesses c_y and c_z correlate more strongly with I_{yy} and I_{zz} than translational stiffnesses do with S, because the EI/L scaling of rotational stiffnesses is less diluted by the length effect than the EI/L^3 scaling of bending translational stiffnesses. Finally, the A/B end stiffness pairs show a mild mutual correlation of approximately 0.1–0.2, attributable entirely to their shared characteristic stiffness factor as discussed above.

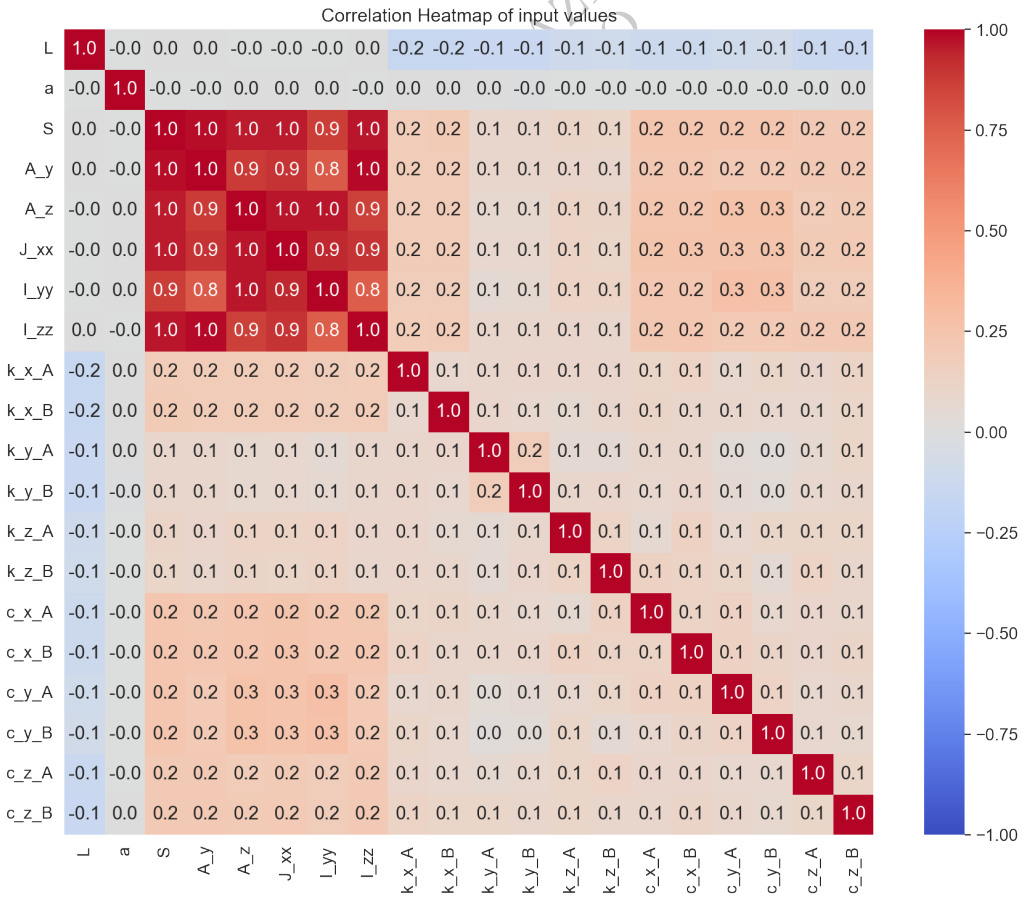

Figure 14: Input heatmap.

4.2 Statistical characterization of the output space

The output space is characterized separately for kinematic quantities, internal force quantities, and strain energy. For each quantity, the maximum absolute value over the 21 sampling positions is used as a scalar summary of the response magnitude. All distributions are examined on a logarithmic scale given the wide dynamic range of the responses across the dataset. The figures presented in this section are computed on the per-load-case dataframes, since each load case activates a distinct and decoupled subset of output quantities.

4.2.1 Kinematic response distributions

Figures 15, 16, 17, 18, 19, 20, 21, 22, 23 and 24 show the distributions of maximum absolute kinematic quantities for each relevant load case, stratified by section family.

Axial displacement, Load case 1. The distribution of $\max|u_x^G|$ spans roughly 10^{-8} to 10^{-4} m, making it the narrowest and smallest-magnitude kinematic response in the dataset. This is consistent with the very high axial rigidity of steel sections: the axial stiffness ES/L is several orders of magnitude larger than the bending stiffness EI/L^3, so unit axial loads produce very small displacements. The section family stratification is proportional across all response bins, with no family dominating a particular magnitude regime.

Transverse displacements, Load cases 2, 3, 5, 6. The distributions of $\max|u_y^G|$ and $\max|u_z^G|$ span a significantly wider range than $\max|u_x^G|$, from roughly 10^{-7} to 10^1 m, reflecting the much softer bending response of beam elements. Within each bending plane, the force-activated load cases (LC2 for u_y, LC3 for u_z) produce systematically larger peak displacements than their moment-activated counterparts (LC6 for u_y, LC5 for u_z), since a unit point moment generates smaller deflections than a unit point force on average. The IPE family consistently extends further into the high-displacement tail compared to HEA and HEB, reflecting the lower second moments of area of IPE profiles (particularly I_{zz}) which governs bending in the $(G, \vec{e}_y)$ plane and is the smallest of the three families for equivalent profile heights.

Torsional rotation, Load case 4. The distribution of $\max|\theta_x|$ spans roughly 10^{-3} to 10^1 rad and exhibits the most pronounced section family separation of all kinematic quantities. IPE profiles extend significantly further into the high-rotation tail compared to HEA and HEB, a direct consequence of their substantially lower torsion constant J_{xx}. Open thin-walled sections such as IPE profiles are well known to be torsionally flexible, and this property is clearly captured by the dataset.

Bending rotations, Load cases 2, 3, 5, 6. The distributions of $\max|\theta_y|$ and $\max|\theta_z|$ mirror the structure of the transverse displacement distributions, with force-activated load cases producing larger peak rotations than moment-activated ones. The y-plane bending rotations (θ_y, LC3 and LC5) are systematically smaller than the z-plane bending rotations (θ_z, LC2 and LC6) for the same load case type, consistent with $I_{yy} > I_{zz}$ for all three profile families ; bending about the strong axis produces smaller rotations than bending about the weak axis.

V. DERRAZI
D. PELUFFO
J.C. TORRES

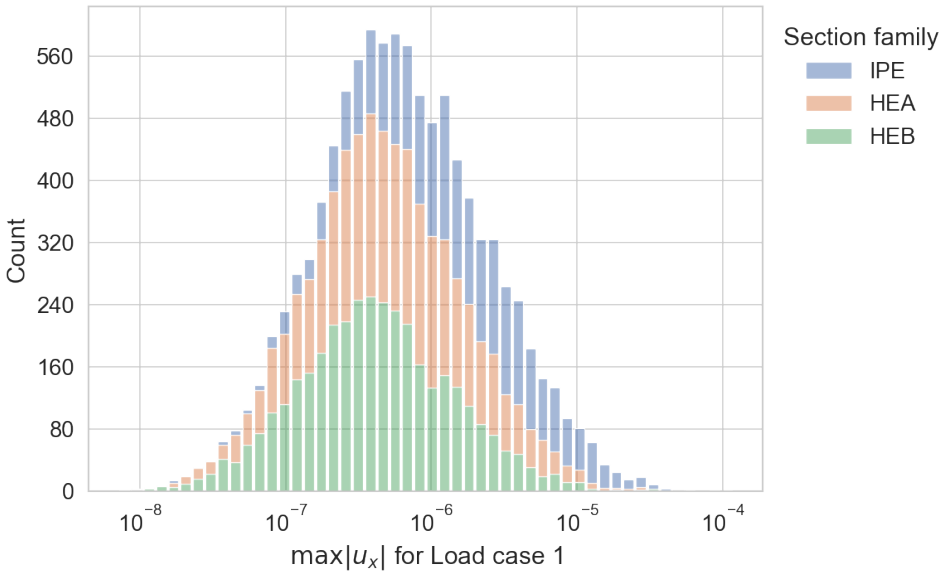

Figure 15: max $|u_x|$ LC1 hued by section family

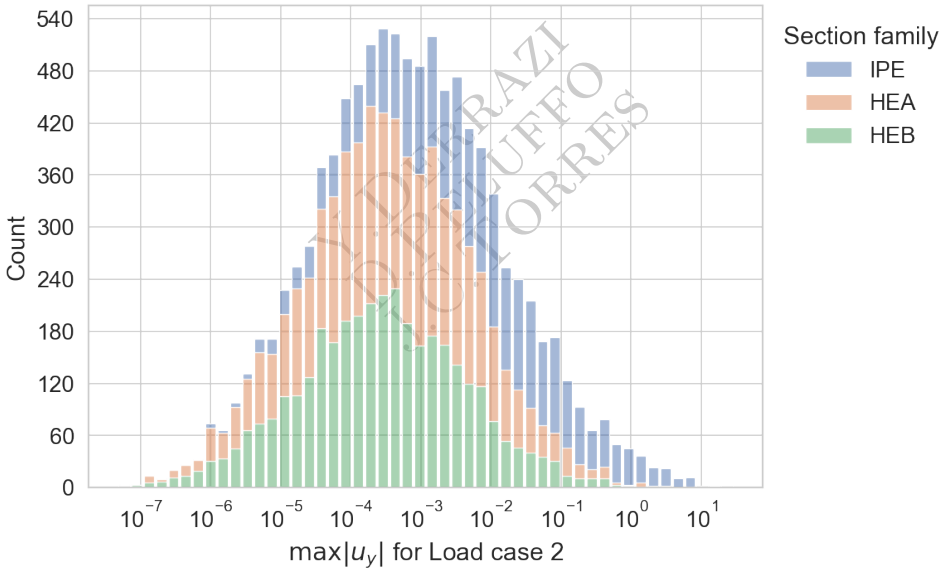

Figure 16: max $|u_y|$ LC2 hued by section family

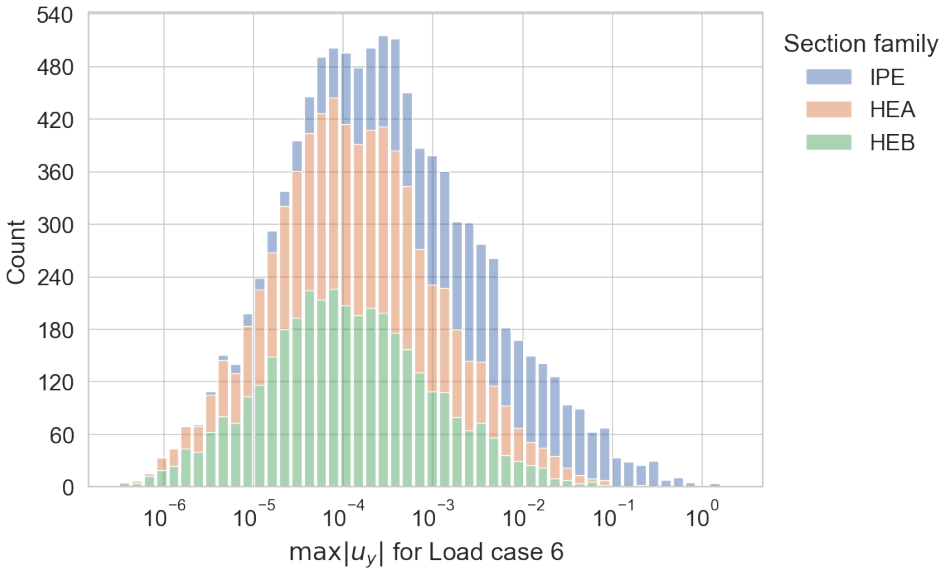

Figure 17: max $|u_y|$ LC6 hued by section family

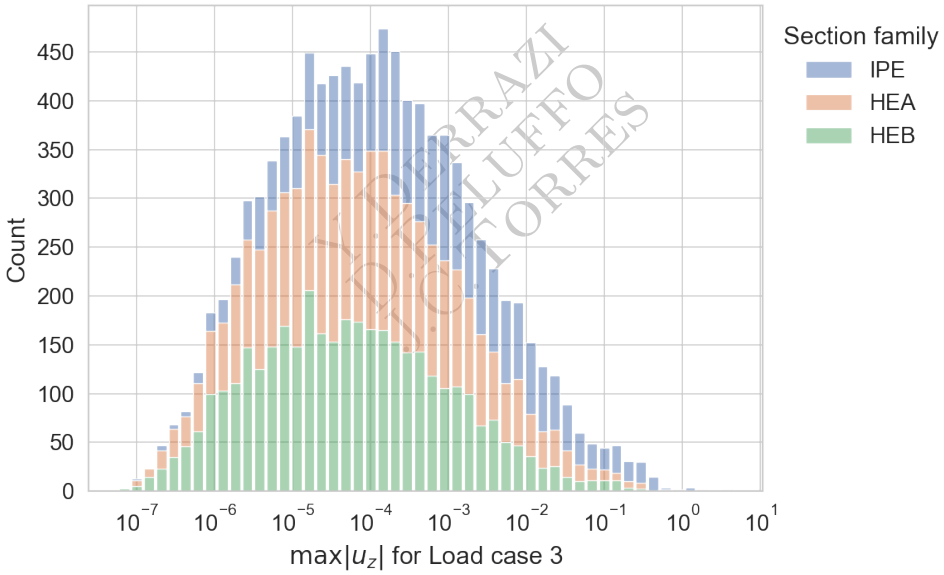

Figure 18: max $|u_z|$ LC3 hued by section family

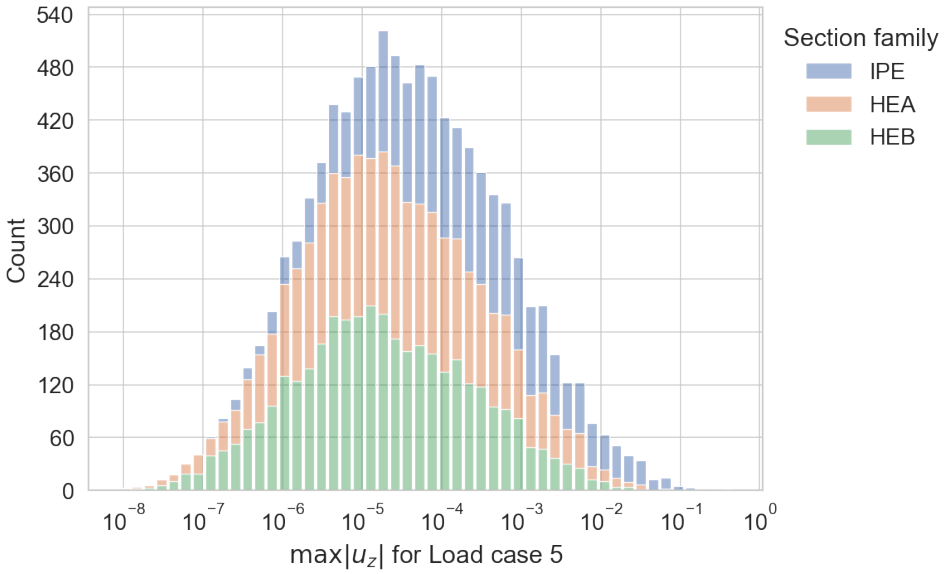

Figure 19: max $|u_z|$ LC5 hued by section family

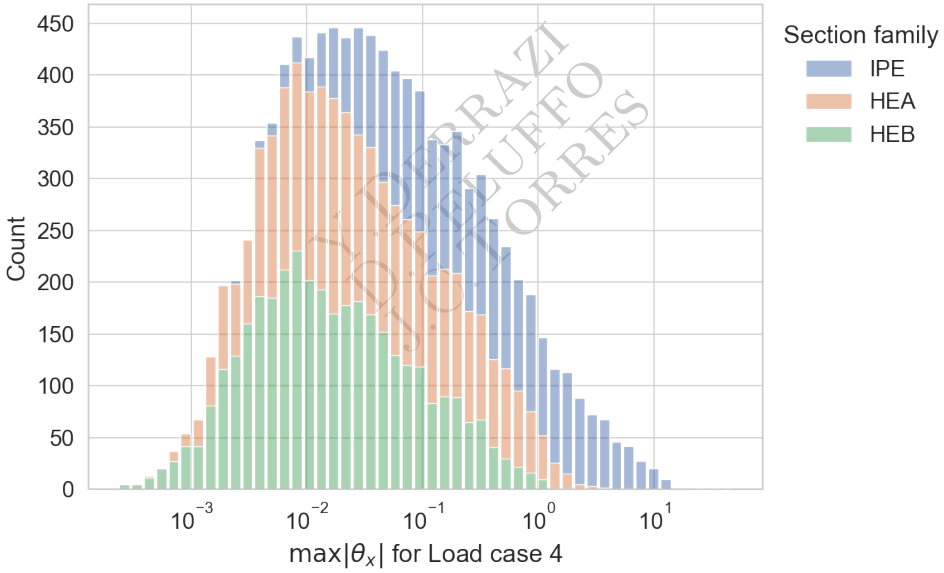

Figure 20: max $|\theta_x|$ LC4 hued by section family

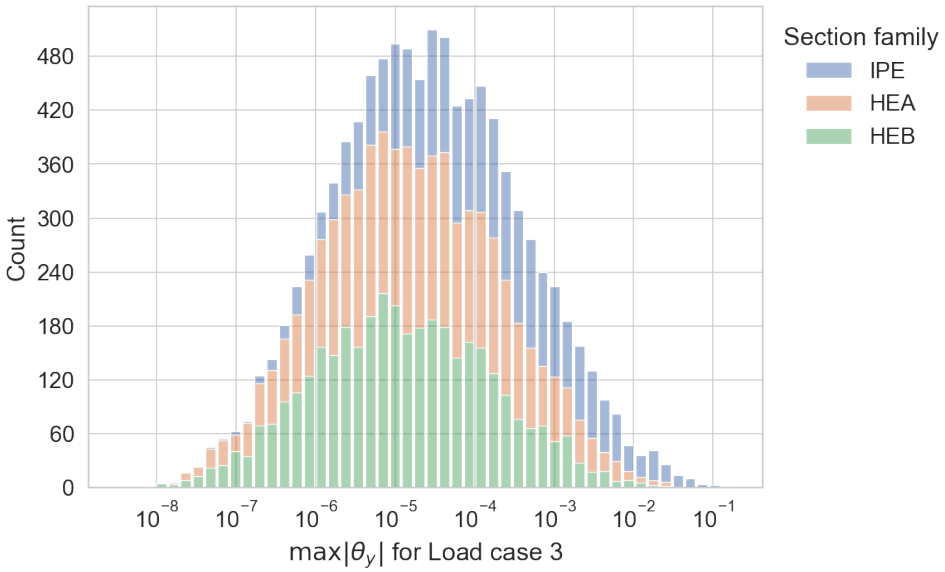

Figure 21: max $|\theta_y|$ LC3 hued by section family

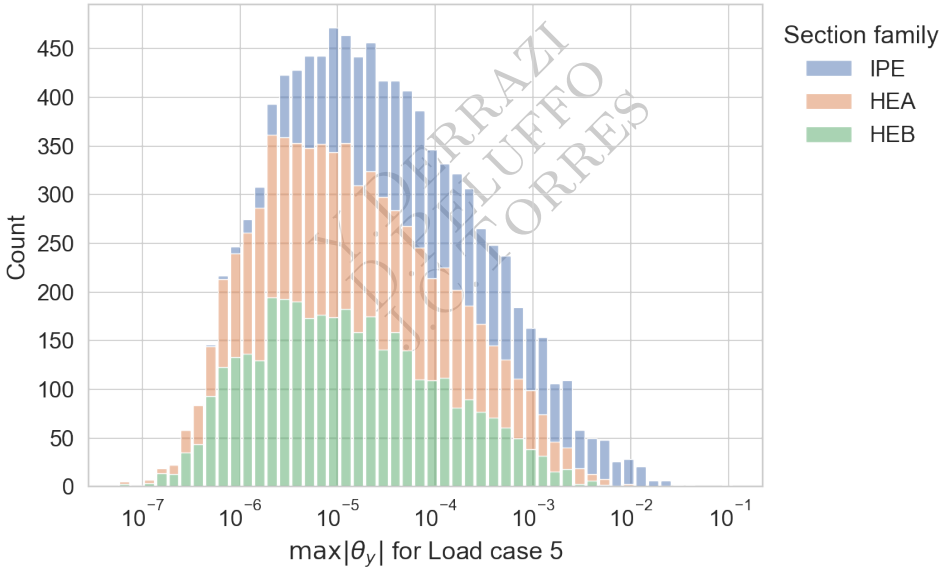

Figure 22: max $|\theta_y|$ LC5 hued by section family

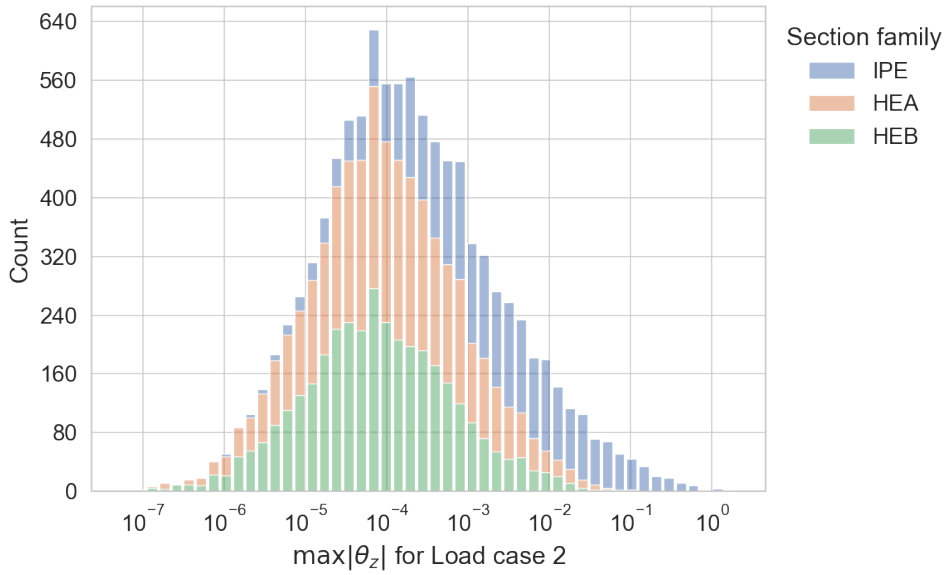

Figure 23: max $|\theta_z|$ LC2 hued by section family

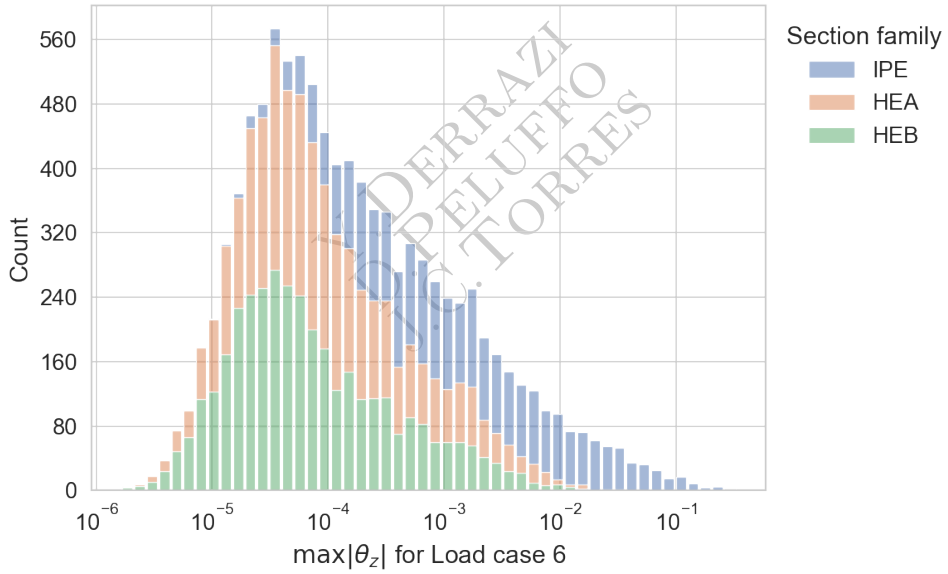

Figure 24: max $|\theta_z|$ LC6 hued by section family

4.2.2 Internal force distributions

The internal force distributions reveal distinct and physically meaningful patterns that differ fundamentally between force-activated and moment-activated load cases. All figures in this subsection are stratified by relative load position α.

Normal force, Load case 1. The distribution of max $|N|$ in Figure 25 exhibits a highly structured comb-like pattern on a linear scale, with narrow spikes at discrete values and all 19 load positions contributing equally to each spike. This is a direct numerical consequence of the equilibrium equation $\mathrm{d}N/\mathrm{d}x = 0$: in the absence of distributed loads, the normal force is constant along the beam and depends only on the support reaction split, which is governed by the stiffness sampling rather than the load position. The perfect load position decorrelation confirms that α has no influence on N for a unit axial load, as expected from the equilibrium equations.

Torsion moment, Load case 4. The distribution of $\max|T|$ in Figure 26 exhibits the same comb-like structure as $\max|N|$, for the same reason: $\mathrm{d}T/\mathrm{d}x = 0$ in the absence of distributed torsional loads, so T is constant along the beam and independent of α. The identical load position stratification across all spikes confirms this analytically.

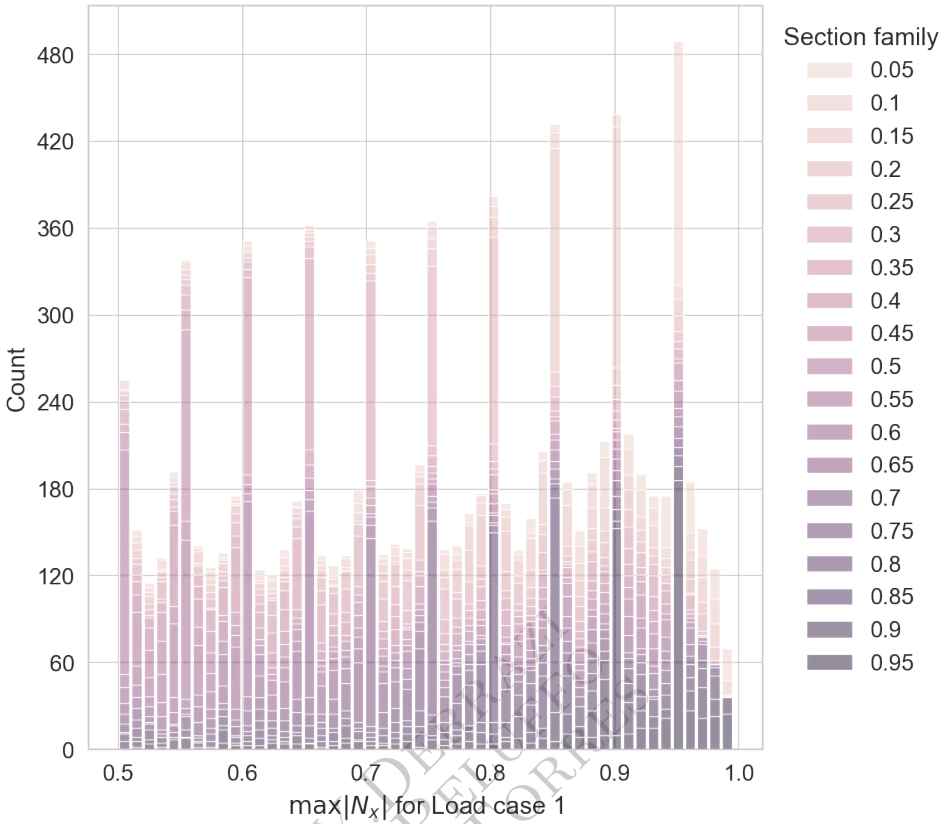

Figure 25: $\max|N|$ hued by load position.

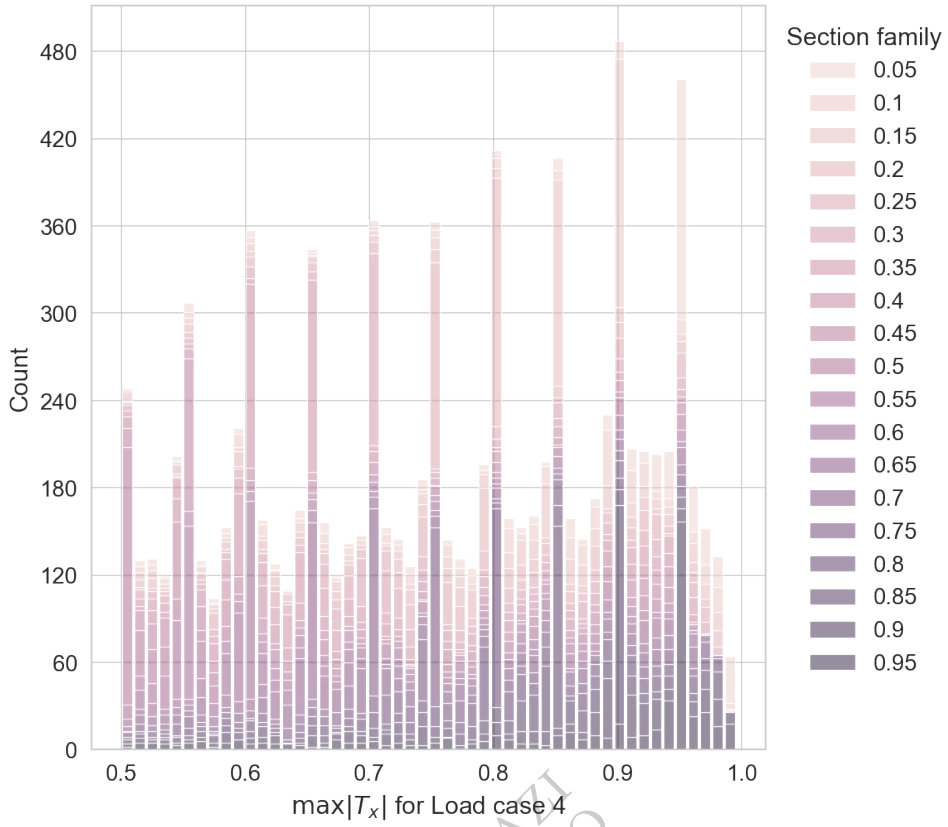

Figure 26: max $|T_x|$ hued by load position.

Shear forces, Force-activated load cases 2 and 3. The distributions of max $|V_y|$ (LC2) and max $|V_z|$ (LC3) in Figures 27 and 29 are strongly right-skewed with a pronounced spike at 1.0 kN; the magnitude of the applied unit load. This upper bound is exact: for a unit point load, the maximum shear force cannot exceed the applied load, and approaches it when the load is applied very close to one end. The load position stratification is clearly visible in the right tail; lighter colors corresponding to α near 0.05 or 0.95 concentrate near 1.0 kN, while darker colors corresponding to α near 0.5 spread toward lower values. This directly reflects the reaction distribution for a unit transverse load, modulated by the elastic support conditions.

Shear forces, Moment-activated load cases 5 and 6. The distributions of max $|V_y|$ (LC6) and max $|V_z|$ (LC5) in Figures 28 and 30 are dramatically different from their force-activated counterparts. This is physically exact: a unit point moment $\vec{C} = \vec{e}_z$ or $\vec{C} = \vec{e}_y$ produces shear forces of order $1/L$ from moment equilibrium, independently of α. The distribution of max $|V_y|$ for LC6 therefore mirrors the distribution of $1/L$ in the dataset ; a right-skewed distribution concentrated near zero consistent with L being sampled uniformly between 1 and 10 m. The load position decorrelation is perfect, confirming the α-independence of the shear force for moment loading.

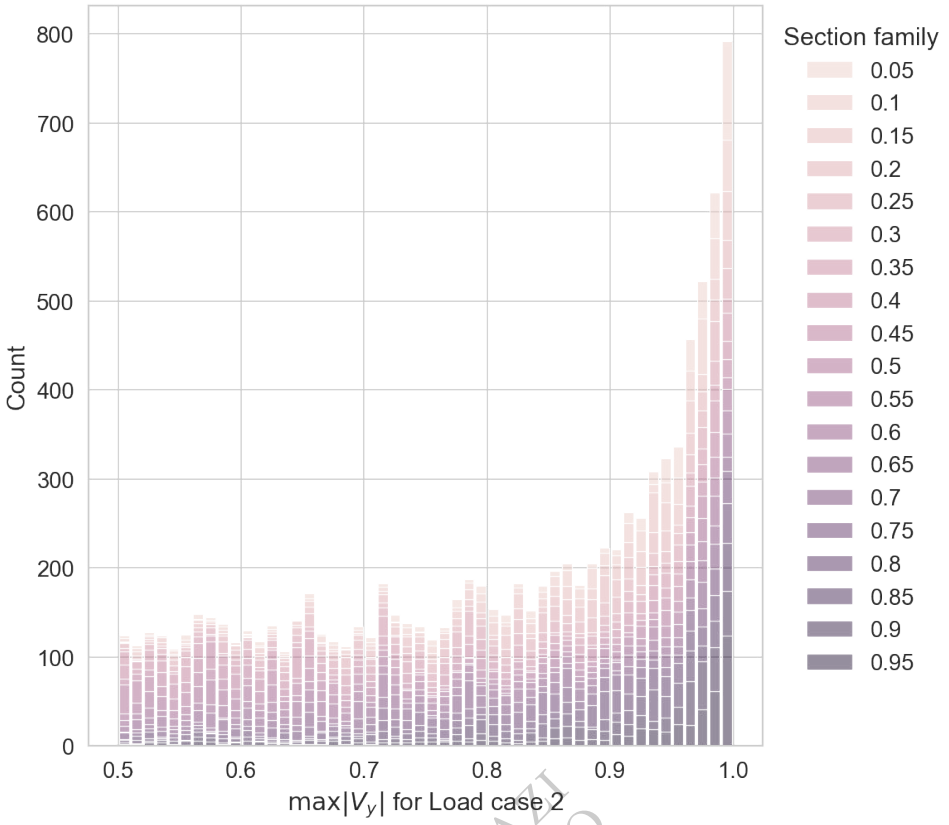

Figure 27: max $|V_y|$ LC2 hued by load position.

Y. DERRAZI
D. PELUFFO
J.C. TORRES

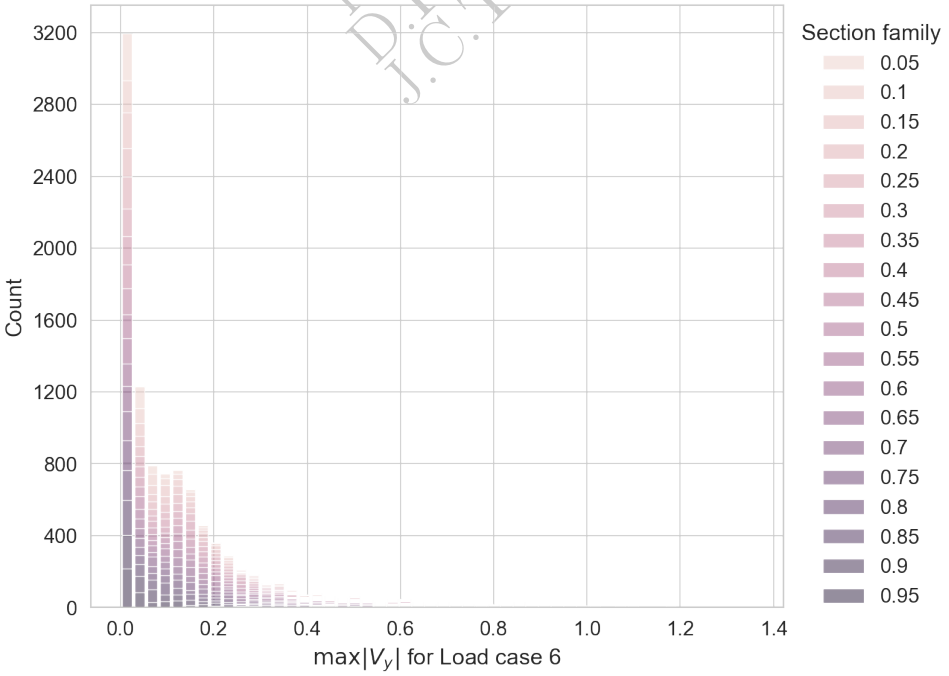

Figure 28: max $|V_y|$ LC6 hued by load position.

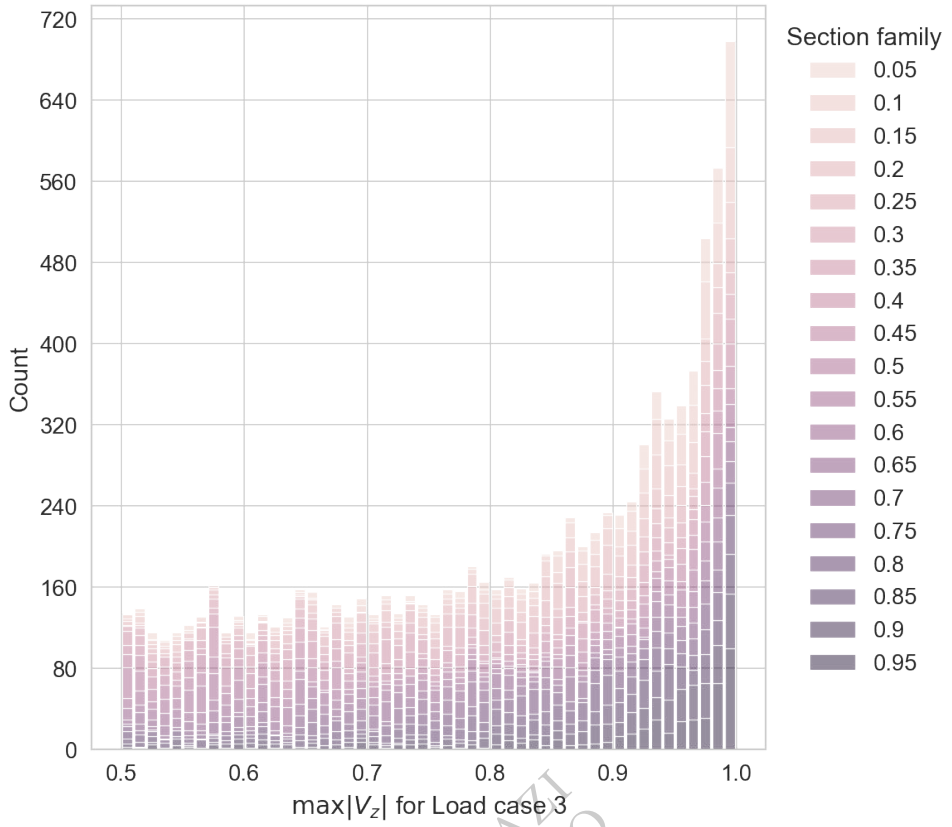

Figure 29: $\max|V_z|$ LC3 hued by load position.

Y. DERRAZI
D. PELUFFO
J.C. TORRES

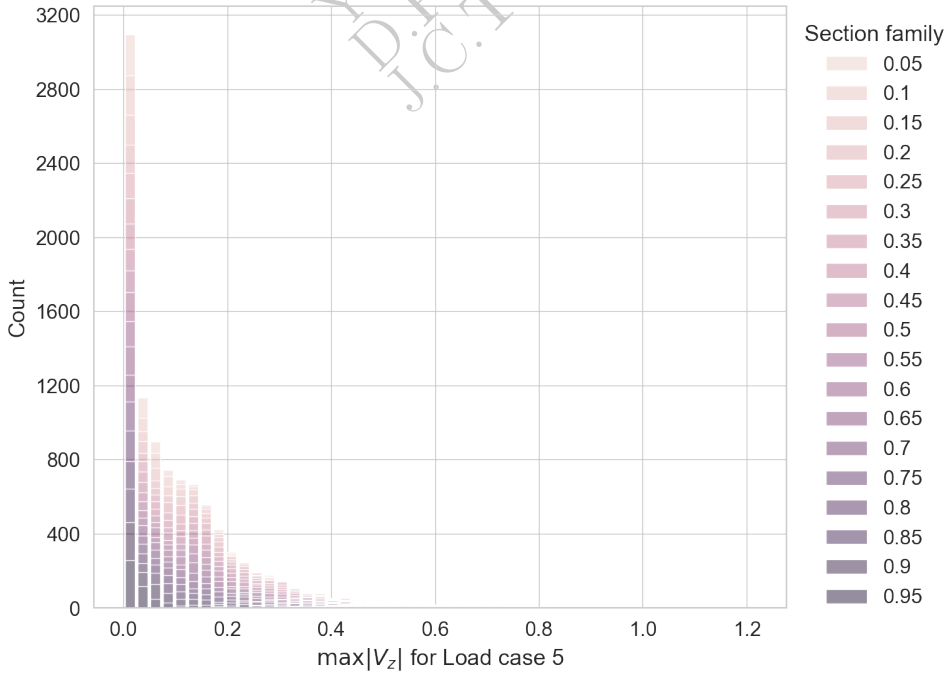

Figure 30: $\max|V_z|$ LC5 hued by load position.

Bending moments, Force-activated load cases 2 and 3. The distributions of $\max|M_z|$ (LC2) and $\max|M_y|$ (LC3) in Figures 31 and 33 are right-skewed on a linear scale, spanning 0 to approximately 7 kN·m.

The load position stratification is visible : load positions near the beam midspan ($\alpha \approx 0.5$) produce larger bending moments and extend further into the high-moment tail, consistent with the classical influence line for bending moment which is maximized by midspan loading. Load positions near the beam endpoints produce smaller moments and concentrate at the lower end of the distribution.

Bending moments, Moment-activated load cases 5 and 6. The distributions of $\max|M_y|$ (LC5) and $\max|M_z|$ (LC6) in Figures 34 and 32 are bounded between 0.5 and 1.0 kN·m and exhibit a multi-modal structure driven by the discrete α grid. This is a known property of unit moment loading: the maximum bending moment is bounded by the applied moment magnitude (1 kN·m) and its distribution along the beam creates a characteristic pattern that depends on α and the support conditions. The discrete load position values are clearly visible as separate modes in the distribution, confirming that the bending moment response to a unit moment is sensitive to load position in a structured and physically predictable way.

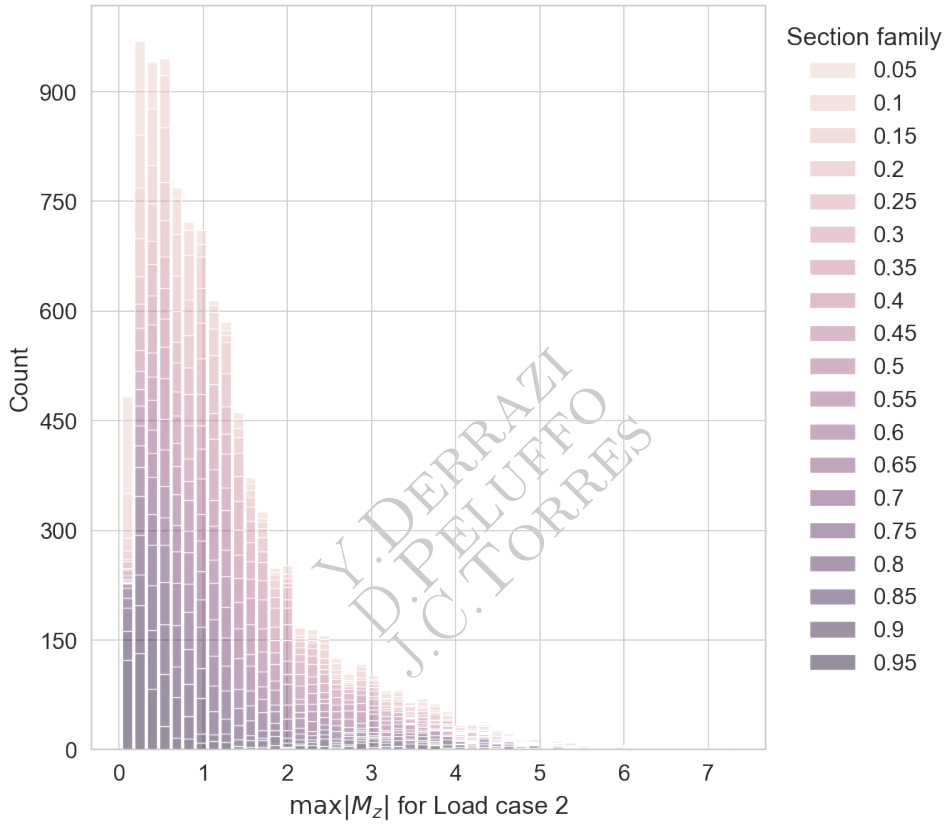

Figure 31: $\max|M_z|$ LC2 hued by load position.

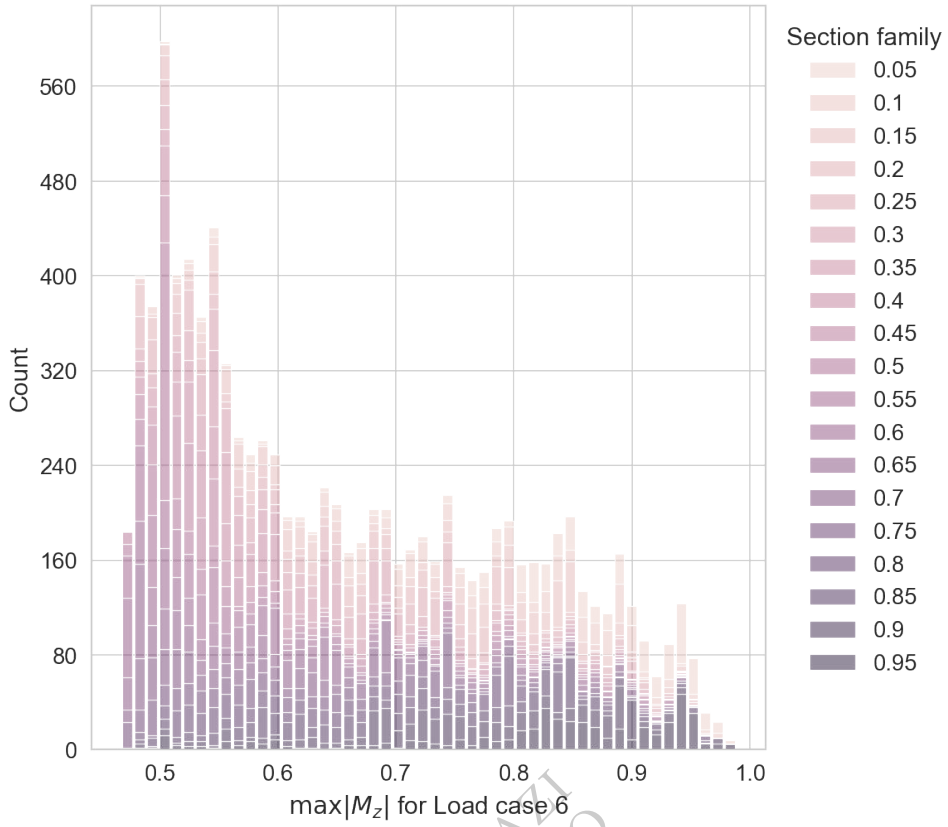

Figure 32: max $|M_z|$ LC6 hued by load position.

Y. DERRAZI
D. PELUFFO
J.C. TORRES

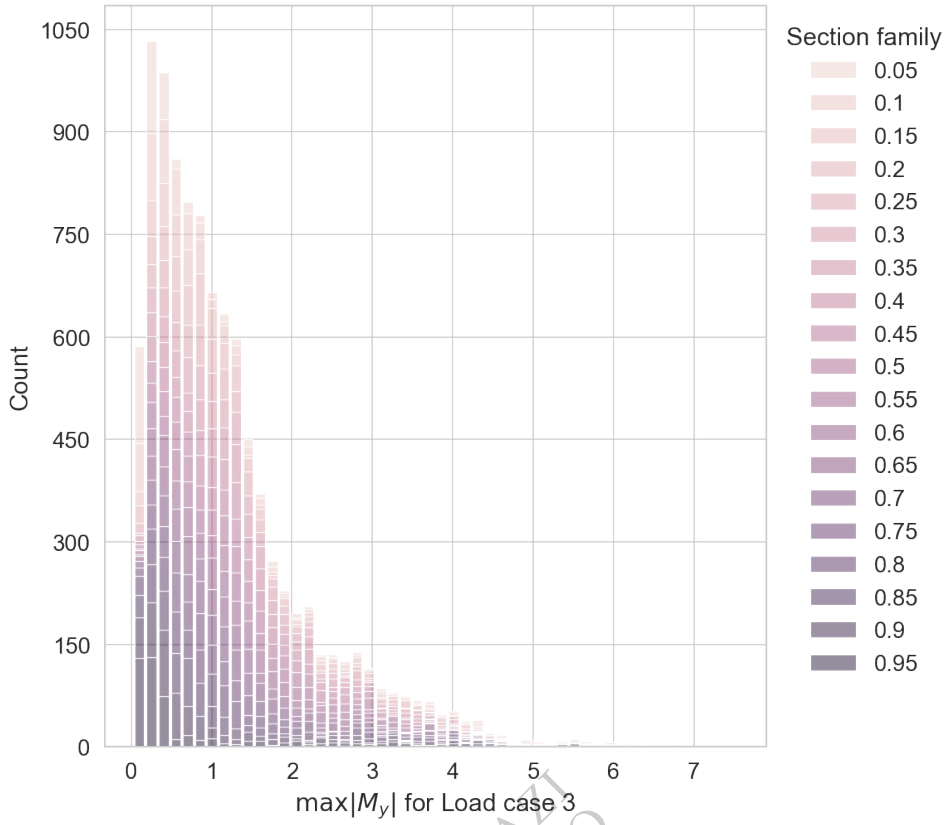

Figure 33: max $|M_y|$ LC3 hued by load position.

Y. DERRAZI
D. PELUFFO
J.C. TORRES

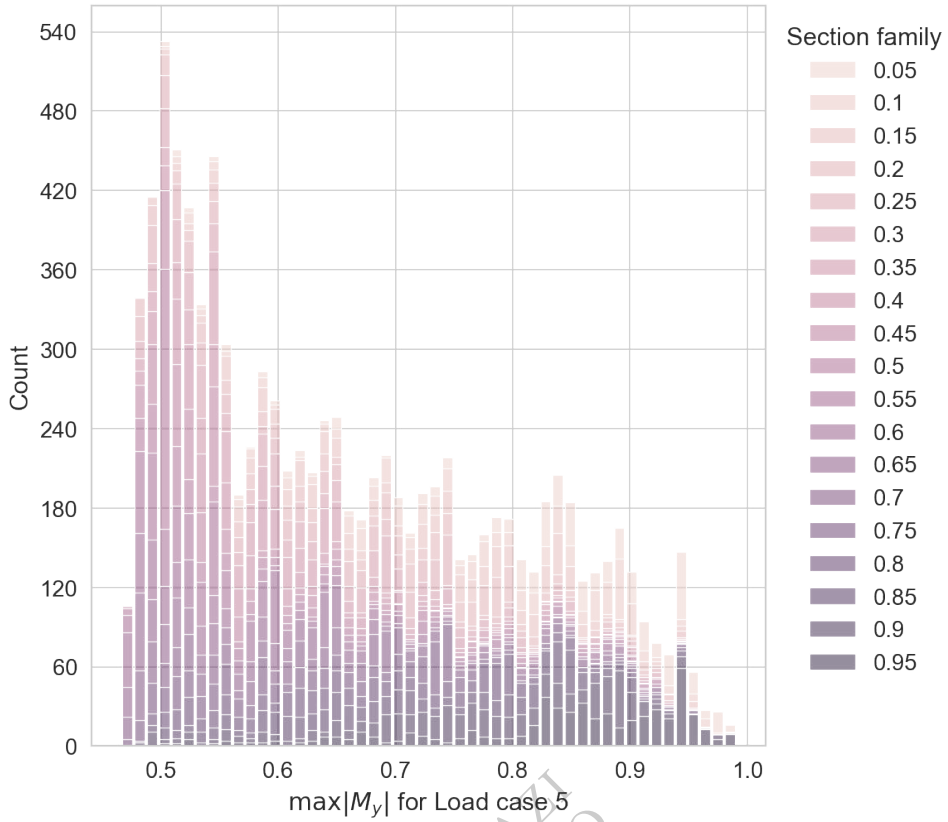

Figure 34: max $|M_y|$ LC5 hued by load position.

Y. DERRAZI
D. PELUFFO
J.C. TORRES

4.2.3 Strain energy distribution

Figures 35 and 36 show the distribution of the KARAMBA3D strain energy output `Ud` across all 60000 instances, stratified by load position and section family respectively. The distribution is bell-shaped on the log scale, spanning roughly 10^{-11} to 10^{1} kN·m, reflecting the enormous variability in structural response across the sampled parameter space.

Figure 35 shows perfect load position decorrelation: all 19 α values contribute identically at every energy bin, confirming that the strain energy is not systematically biased by load position. Figure 36 shows a clear section family separation: IPE profiles extend further into both the low and high energy tails compared to HEA and HEB, consistent with their greater flexibility. HEB profiles are the most concentrated around the median, reflecting their greater stiffness relative to equivalent profile heights.

It should be noted that the `Ud` column shown here corresponds to KARAMBA3D's internal energy output, whose limitations are discussed in detail later on. The corrected strain energy $\hat{U}_D$, recomputed from the internal force fields, is stored as `Ud_rebuilt` in the distributed dataset.

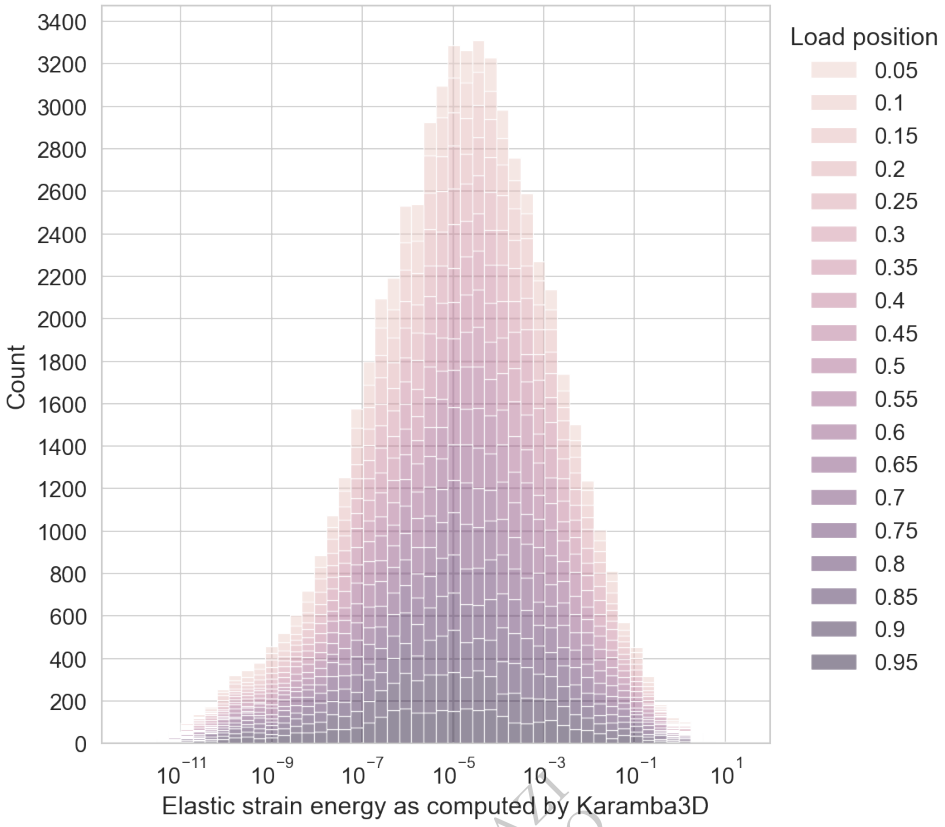

Figure 35: U_D distribution hued by load position.

Y. DIERRAZI
D. PELUFFO
J.C. TORRES

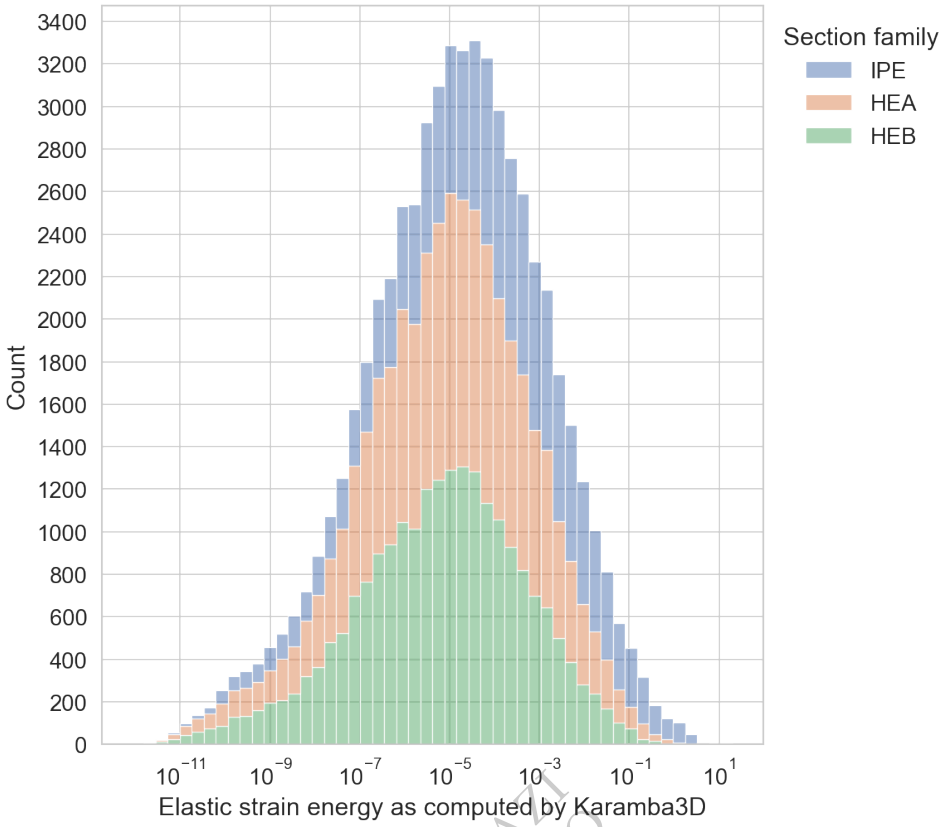

Figure 36: U_D distribution hued by section family.

I. DIRRAZI
D. PELUFFO
J.C. TORRES

4.2.4 Response scaling with section properties

Figures 37, 38 and 39 show the log-log scatter plots of maximum absolute displacements against their governing section property for load cases 1, 2 and 3 respectively.

All three figures reveal a clear negative correlation on the log-log scale; larger section properties systematically produce smaller peak displacements, as expected from the governing stiffness relations $u_x \sim 1/(ES)$, $u_y \sim L^3/(EI_{zz})$ and $u_z \sim L^3/(EI_{yy})$. The discrete section catalogue manifests as a vertical striping pattern; each stripe corresponds to a single section in the catalogue, with the spread within each stripe driven entirely by the length and stiffness sampling. The width of the vertical spread, spanning roughly 6 to 8 decades within a single section stripe, is a direct visual confirmation that the support stiffness sampling is effective: for a given section and length, the peak displacement can vary enormously depending on the boundary conditions, from near-rigid clamped conditions to nearly free configurations.

The upper envelope of each cloud, traced by the most flexible configurations (nearly free supports), has a slope of approximately -1 in the $\ln(\max|u_x|)$ vs $\ln(S)$ and $\ln(\max|u_z|)$ vs $\ln(I_{yy})$ plots, consistent with the expected power-law scalings. The $\ln(\max|u_y|)$ vs $\ln(I_{zz})$ plot shows slightly more scatter in the upper envelope, attributable to the simultaneous influence of I_{zz}, A_y and L on the bending response in the $(G, \vec{e}_y)$ plane.

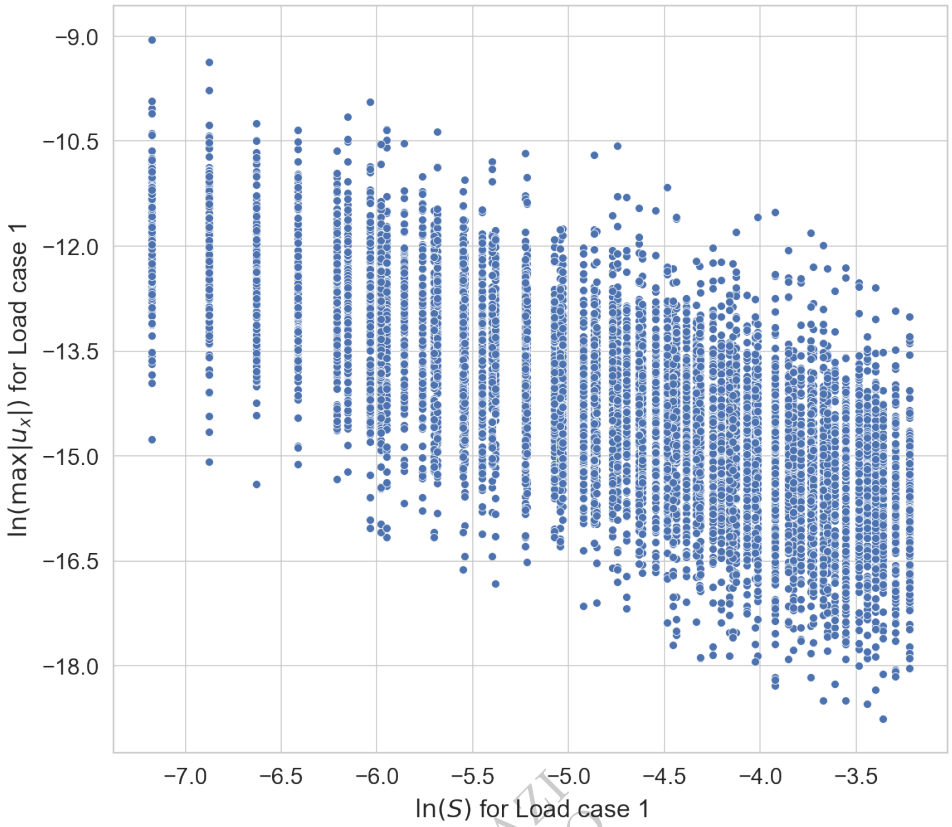

Figure 37: $\ln(\max|u_x|)$ vs $\ln(S)$ scatter.

Y. DERRAZI
D. PELUFFO
J.C. TORRES

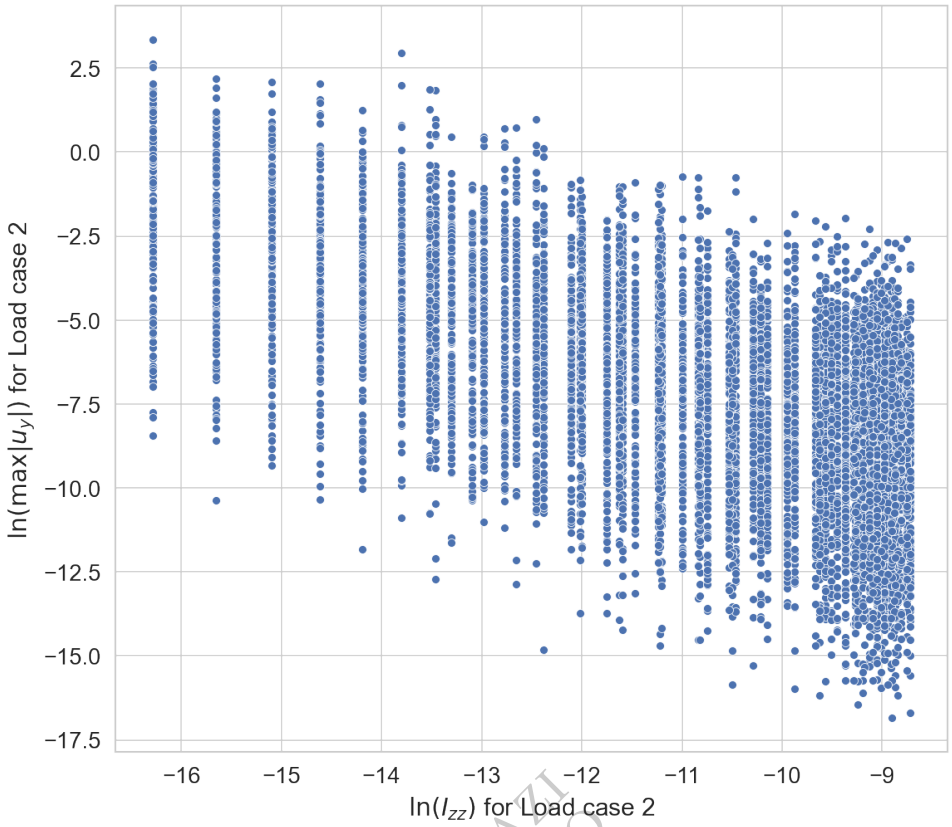

Figure 38: $\ln(\max|u_y|)$ vs $\ln(I_{zz})$ scatter.

Y. DERRAZI
D. PELUFFO
J.C. TORRES

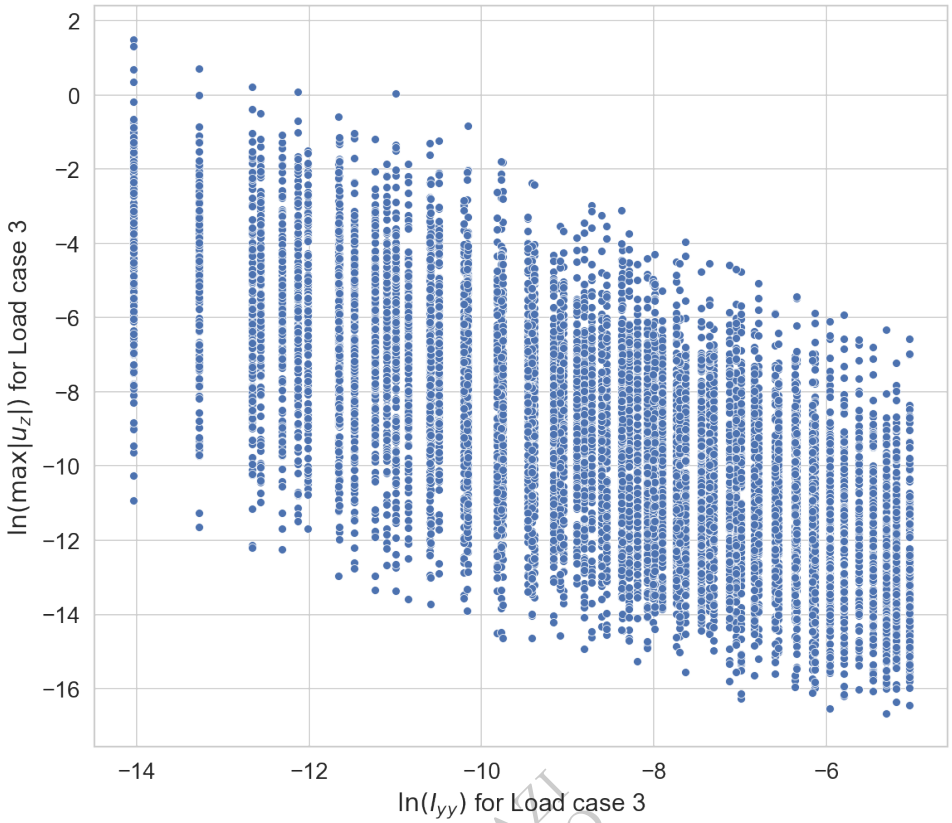

Figure 39: $\ln(\max|u_z|)$ vs $\ln(I_{yy})$ scatter.

Y. DERRAZI
D. PELUFFO
I.C. TORRES

4.3 Physical consistency

4.3.1 Equilibrium verification

The static equilibrium of each beam instance is verified by checking that the sum of all generalized forces and moments acting on the system vanishes identically. For each instance, the nodal reactions at ends A and B are computed from the elastic support laws:

$$R_i^{A/B} = -k_i^{A/B} u_i^{A/B}, \qquad M_i^{A/B} = -c_i^{A/B} \theta_i^{A/B}, \tag{67}$$

where $u_i^{A/B}$ and $\theta_i^{A/B}$ are retrieved from the sampled kinematic fields at $\xi = 0.00$ and $\xi = 1.00$ respectively. The six equilibrium residuals are then defined as:

$$r_{F,i} = F_i^{\text{applied}} + R_i^A + R_i^B, \qquad i \in \{x, y, z\}, \tag{68}$$

$$r_{M,i} = C_i^{\text{applied}} + M_i^A + M_i^B + (\text{moment arm contributions}), \qquad i \in \{x, y, z\}, \tag{69}$$

where F_i^{applied} and C_i^{applied} are the components of the applied unit load or unit moment for the relevant load case, and zero otherwise. Figures 40, 41, 42, 43, 44, 45, 46, 47, 48 and 49 show the distributions of $|r_{F,i}|$ and $|r_{M,i}|$ across all $N = 60000$ instances on a logarithmic scale. Force equilibrium residuals and the torsional moment residual satisfy equilibrium to machine precision ($\sim 10^{-16}$ kN), while bending moment residuals reach at most 2.66×10^{-15} kN·m; approximately one order of magnitude above machine epsilon, attributable to the multiplication by L in the moment arm computation.

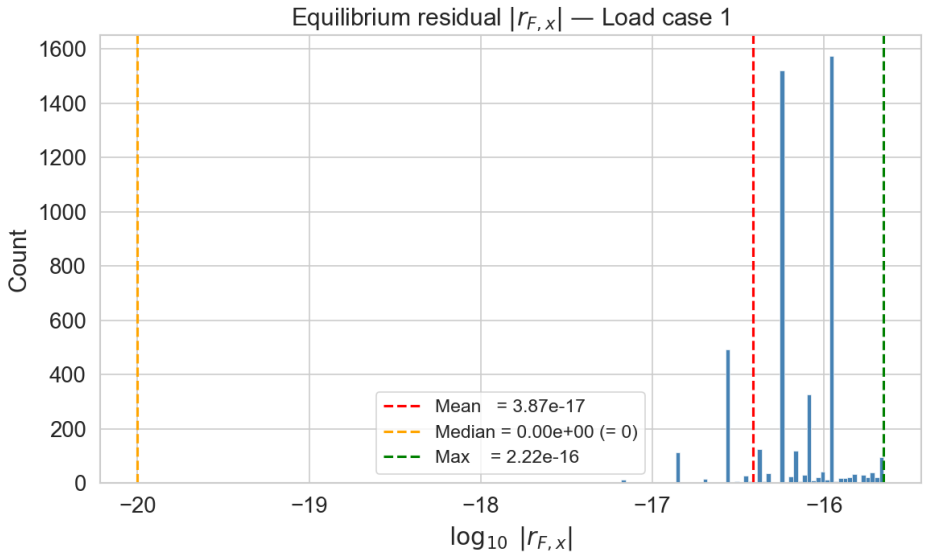

Figure 40: Axial equilibrium residual histogram for LC1.

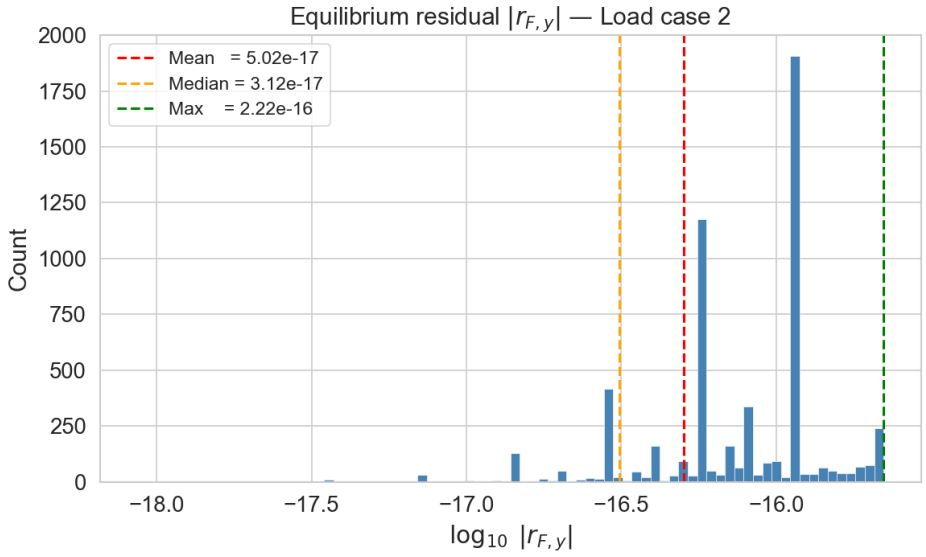

Figure 41: y-transverse equilibrium residual histogram for LC2.

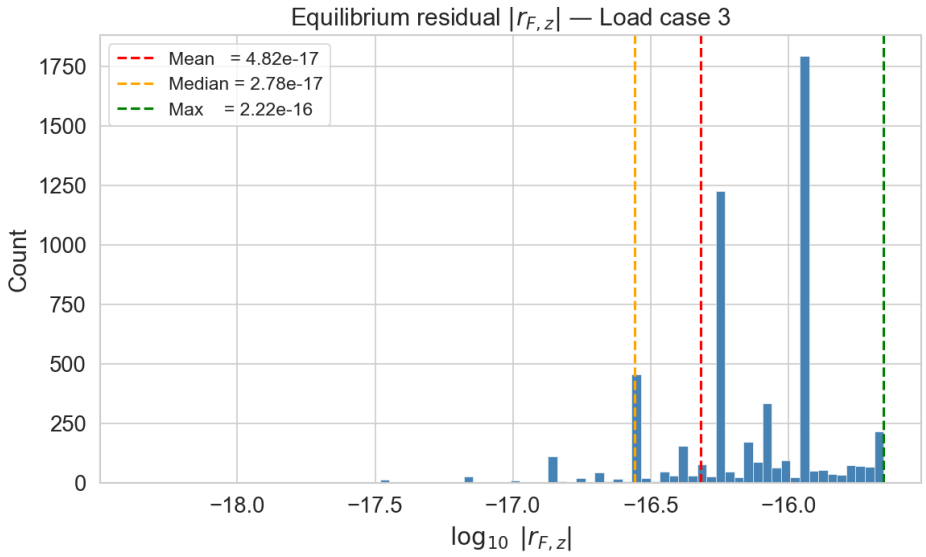

Figure 42: z-transverse equilibrium residual histogram for LC3.

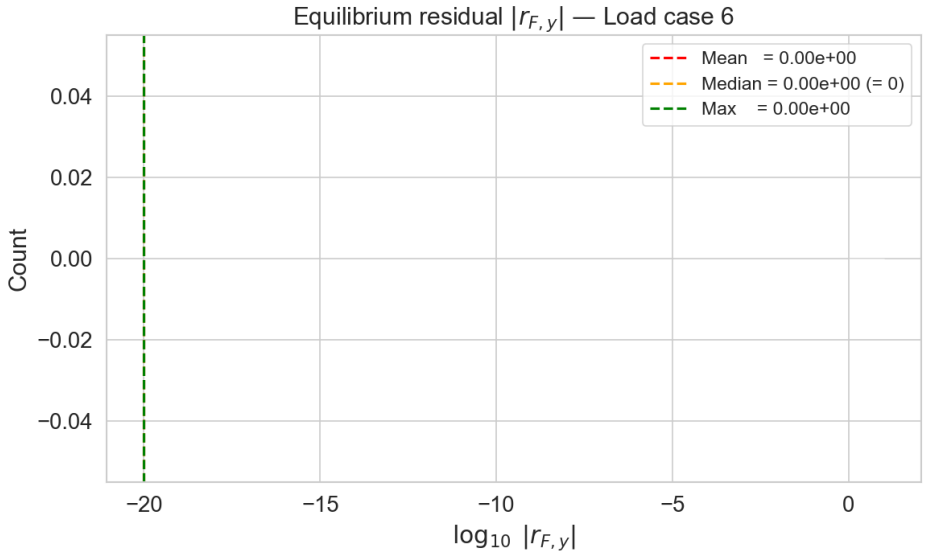

Figure 43: y-transverse equilibrium residual histogram for LC6.

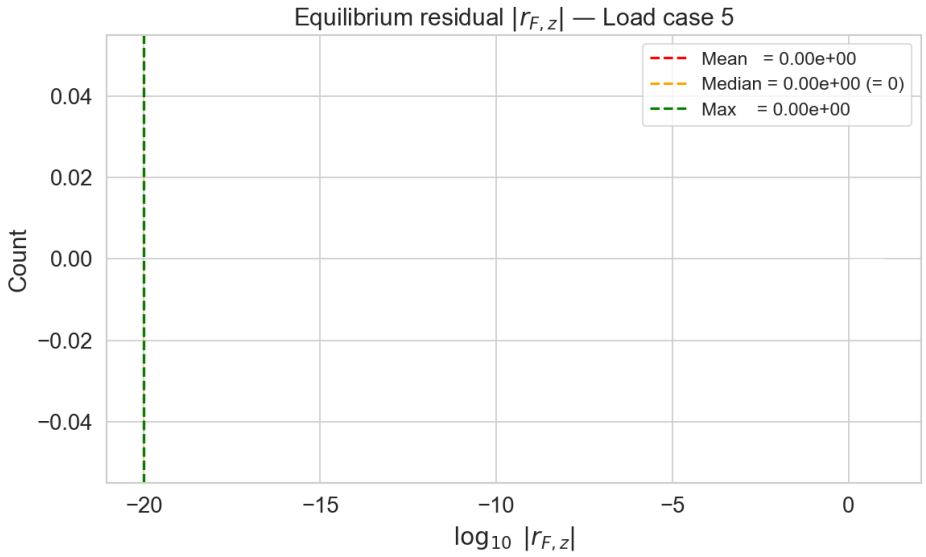

Figure 44: z-transverse equilibrium residual histogram for LC5.

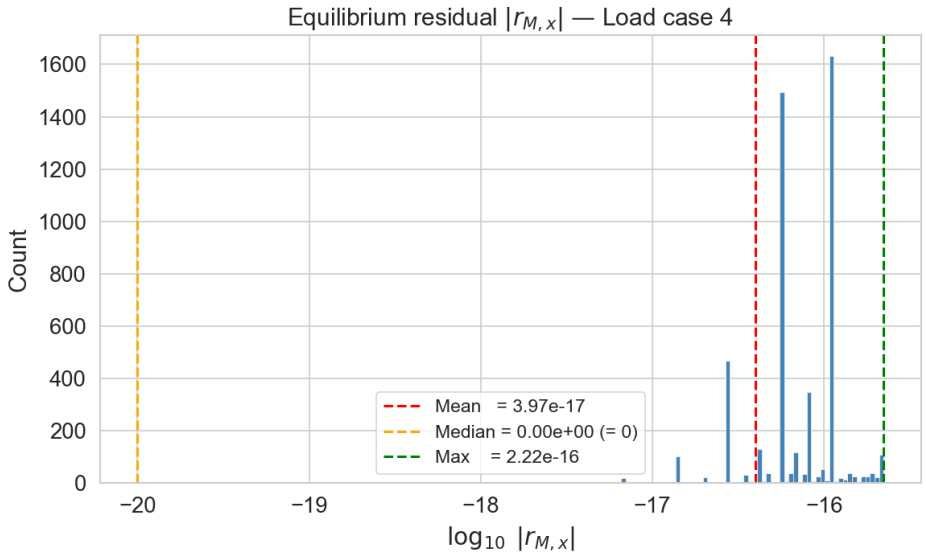

Figure 45: Axial moment equilibrium residual histogram for LC4.

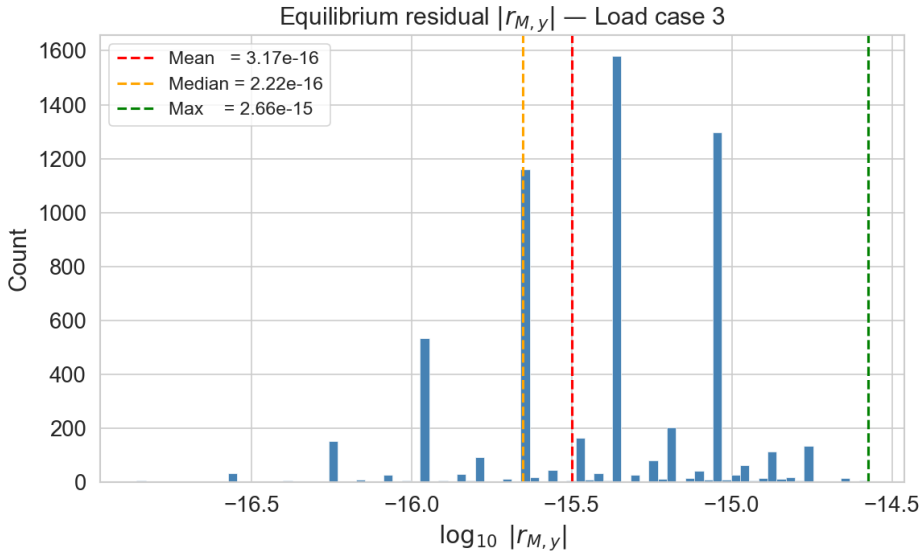

Figure 46: y-transverse moment equilibrium residual histogram for LC3 .

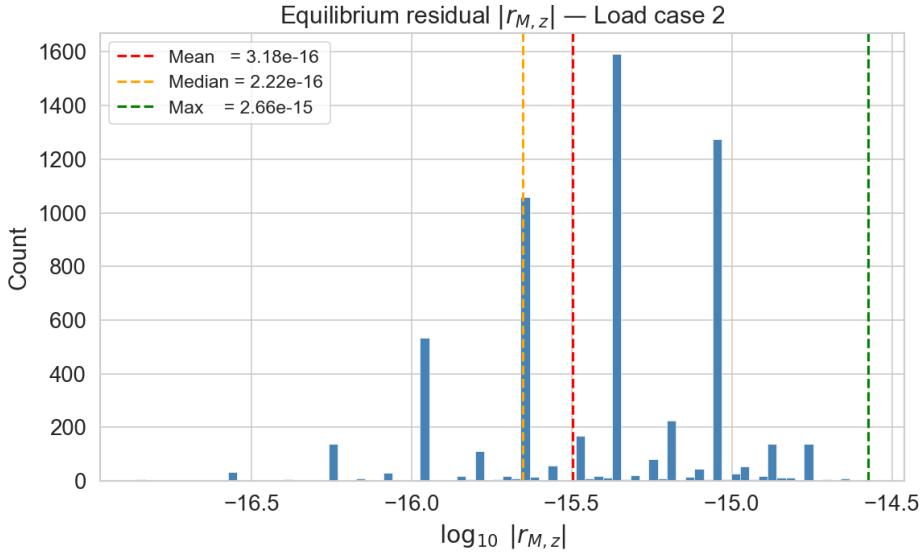

Figure 47: z-transverse moment equilibrium residual histogram for LC2.

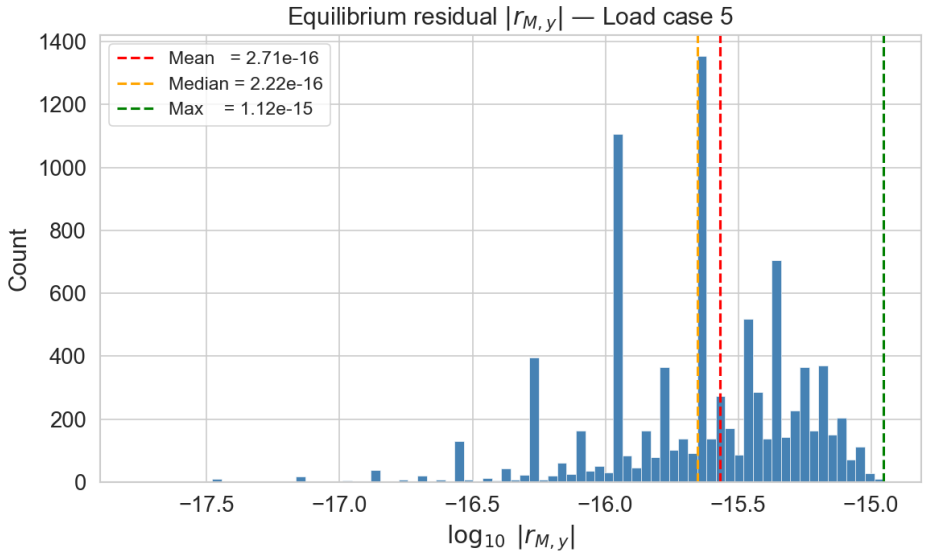

Figure 48: y-transverse moment equilibrium residual histogram for LC5.

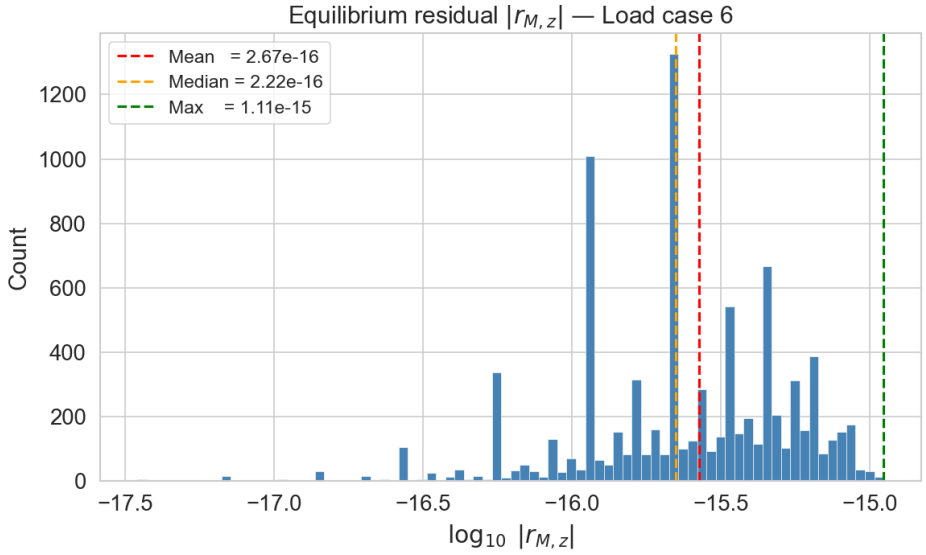

Figure 49: z-transverse moment equilibrium residual histogram for LC6.

This result confirms that the KARAMBA3D solver satisfies global equilibrium to machine precision for all instances in the dataset, regardless of the support stiffness regime.

4.3.2 Individual instance response profiles

Beyond global statistical checks, the physical correctness of the dataset is further verified by examining the full response profiles of individual instances. Figures 50 and 51 show the complete set of active response quantities for two representative instances; one from load case 5 (unit point moment $\vec{C} = \vec{e}_y$, ID = 45006) and one from load case 2 (unit transverse load $\vec{F} = \vec{e}_y$, ID = 13080). The red dashed line indicates the load application position αL in each figure.

Load case 5 — ID = 45006. The shear force V_z is perfectly constant at approximately 0.21 kN across the entire beam length, with no discontinuity at the load application point. This is physically exact: a unit point moment does not introduce a shear discontinuity, and the constant shear value corresponds to $1/L$ from moment equilibrium, consistent with the beam length of approximately 2.75 m readable from the axis. The bending moment M_y shows a piecewise linear profile with a sharp discontinuity of exactly 1 kN·m at αL, the magnitude of the applied unit moment, with equal slopes on both sides confirming the constant shear force. The bending rotation θ_y exhibits a tent-shaped profile peaking at αL, with near-zero values at both ends consistent with stiff rotational supports at A and B. The transverse displacement u_z shows a smooth S-shaped profile changing sign near αL, consistent with the moment discontinuity imposing a slope change in the deflection curve.

Load case 2 — ID = 13080. The shear force V_y is piecewise constant with a sharp discontinuity of exactly 1 kN at the load application point (the magnitude of the applied unit transverse load) that confirms force equilibrium at the load point. The values on each side, approximately 0.73 kN and -0.27 kN, sum to 1 kN and reflect the reaction split at ends A and B for this particular instance. The bending moment M_z shows the expected piecewise linear profile with an angle point at αL, its peak value consistent with the shear distribution. The transverse displacement u_y is smooth and peaks slightly to the right of αL, consistent with the asymmetric loading and elastic support conditions.

In both instances, the decoupling structure established earlier is exactly satisfied: only the response quantities activated by the relevant load case are non-trivial, and all inactive quantities are identically zero.

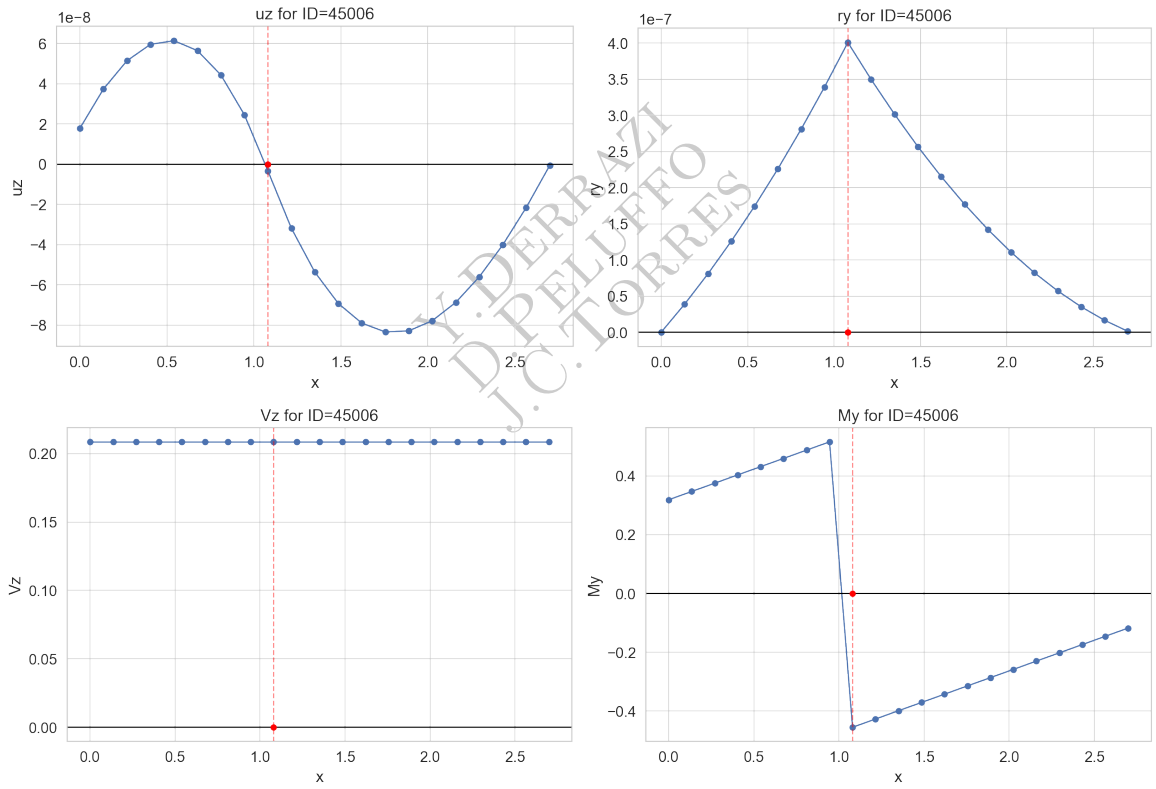

Figure 50: Four panels for ID=45006, LC5.

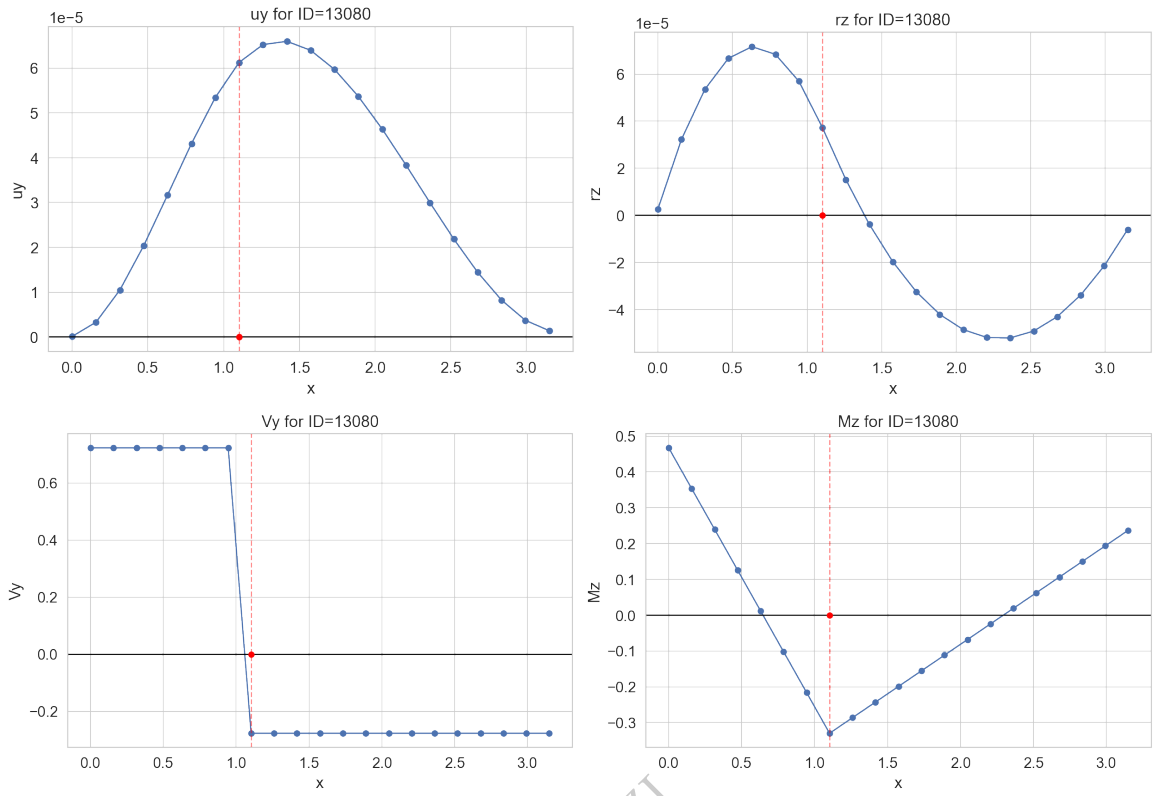

Figure 51: Four panels for ID=13080, LC2.

Y. DERRAZI
D. PELUFFO
J.C. TORRES

4.4 Strain energy consistency and correction

The strain energy U_D is a fundamental scalar output of the dataset, serving as a global measure of the structural response and a natural regression target for surrogate models. Its correct computation is therefore essential to the physical consistency of the dataset. This subsection establishes the theoretical energy balance for a Timoshenko beam with elastic supports, identifies a systematic difference in Karamba3D's internal energy output due to numerical considerations, and describes the correction procedure applied to produce a physically consistent strain energy estimate.

4.4.1 Karamba3D's internal energy output

Karamba3D returns a scalar energy output `Ud` computed internally as:

$$U_{D\mathrm{K3D}} = \frac{1}{2}\boldsymbol{q}^T\boldsymbol{K}\boldsymbol{q}, \tag{70}$$

where $\boldsymbol{q}$ is the nodal displacement vector and $\boldsymbol{K}$ is the beam stiffness matrix . This quantity represents the strain energy of the beam element expressed in terms of nodal displacements alone, and does not account for the elastic energy stored in the support springs.

The limitation of this formulation becomes apparent in the limiting case of fully rigid supports, where $k_i^{A/B} \to \infty$ and $c_i^{A/B} \to \infty$. In this regime, the nodal displacement vector $\boldsymbol{q} \to \boldsymbol{0}$ by definition, and Karamba3D therefore returns $U_{D\mathrm{K3D}} = 0$ regardless of the applied load. This is numerically consistent but physically incorrect : a fully clamped beam subjected to a point load deforms locally, develops non-trivial internal force distributions, and stores a strictly positive strain energy. The vanishing of $\boldsymbol{q}$ reflects the constraint imposed by the supports on the nodal degrees of freedom, not the absence of internal deformation.

4.4.2 Numerical recomputation from internal force fields

Since the internal force distributions $N(\xi)$, $V_y(\xi)$, $V_z(\xi)$, $T(\xi)$, $M_y(\xi)$, $M_z(\xi)$ are retrieved at 21 equally spaced positions $\xi \in \{0.00, 0.05, \ldots, 1.00\}$ for every instance, the beam strain energy can be recomputed directly from

these fields via numerical integration, bypassing KARAMBA3D's internal energy output entirely. Defining the strain energy density:

$$f(\xi) = \frac{1}{2}\left(\frac{N(\xi)^2}{ES} + \frac{V_y(\xi)^2}{GA_y} + \frac{V_z(\xi)^2}{GA_z} + \frac{M_y(\xi)^2}{EI_{yy}} + \frac{M_z(\xi)^2}{EI_{zz}} + \frac{T(\xi)^2}{GJ_{xx}}\right), \tag{71}$$

the strain energy is approximated via the composite trapezoidal rule:

$$U_D \approx \frac{L}{2}\sum_{k=0}^{n_\xi - 2}(\xi_{k+1} - \xi_k)\left[f(\xi_k) + f(\xi_{k+1})\right], \tag{72}$$

where $n_\xi = 21$ and $\xi_{k+1} - \xi_k = 0.05$ uniformly. This expression is exact for piecewise linear $f(\xi)$ and converges to the true integral as $n_\xi \to \infty$. For the TIMOSHENKO beam element considered here, the internal force distributions are piecewise polynomial between the load application point α and the beam endpoints, with at most a single discontinuity in V_y or V_z at $\xi = \alpha$. The trapezoidal approximation therefore introduces a truncation error that is localized near $\xi = \alpha$ and is of order $O(\Delta\xi^2)$ elsewhere, which is acceptable given the sampling resolution $\Delta\xi = 0.05$.

The recomputed strain energy $\hat{U}_D$ does not replace `Ud` in the distributed dataset but can complete it with a column named `Ud_rebuilt`. The raw KARAMBA3D output `Ud` is retained in the analysis dataframe.

4.4.3 Validation of the corrected strain energy

The corrected strain energy $\hat{U}_D$ is validated against the theoretical energy balance. For each instance, the external work W_{ext} is computed as the displacement at the load application point α in the direction of the applied unit load, retrieved directly from the sampled kinematic fields. The elastic work W_k is computed from the nodal displacements and rotations at A and B, which correspond to the sampled values at $\xi = 0.00$ and $\xi = 1.00$ respectively:

$$W_k = \sum_{j\in\{A,B\}}\sum_{i\in\{x,y,z\}}\left[k_i^j\left(u_i^j\right)^2 + c_i^j\left(\theta_i^j\right)^2\right]. \tag{73}$$

The energy balance residual is then defined as:

$$\varepsilon = \left|2\hat{U}_D - (W_{\text{ext}} - W_k)\right|, \tag{74}$$

and computed for all $N = 60000$ instances. Figure 52 shows $\ln(\hat{U}_D)$ against $\ln\left(\frac{1}{2}(W_{ext} - W_k)\right)$ across all instances. The vast majority of points lie on the identity line across the full range of 22 decades, confirming that the recomputed $\hat{U}_D$ satisfies the energy balance to high precision. The scatter visible in the intermediate energy range is attributable to the trapezoidal integration error localized near $\xi = \alpha$, where the shear force discontinuity introduces a truncation error that is most significant when the shear contribution to U_D is large relative to the bending contribution. The one-element discretization adopted in this work is exact for prismatic beams under point loads in the absence of distributed loads, as the internal force distributions are piecewise polynomial and exactly represented by the TIMOSHENKO shape functions.

F. DERRAZI
D. PELFFO
J.C. TORRES

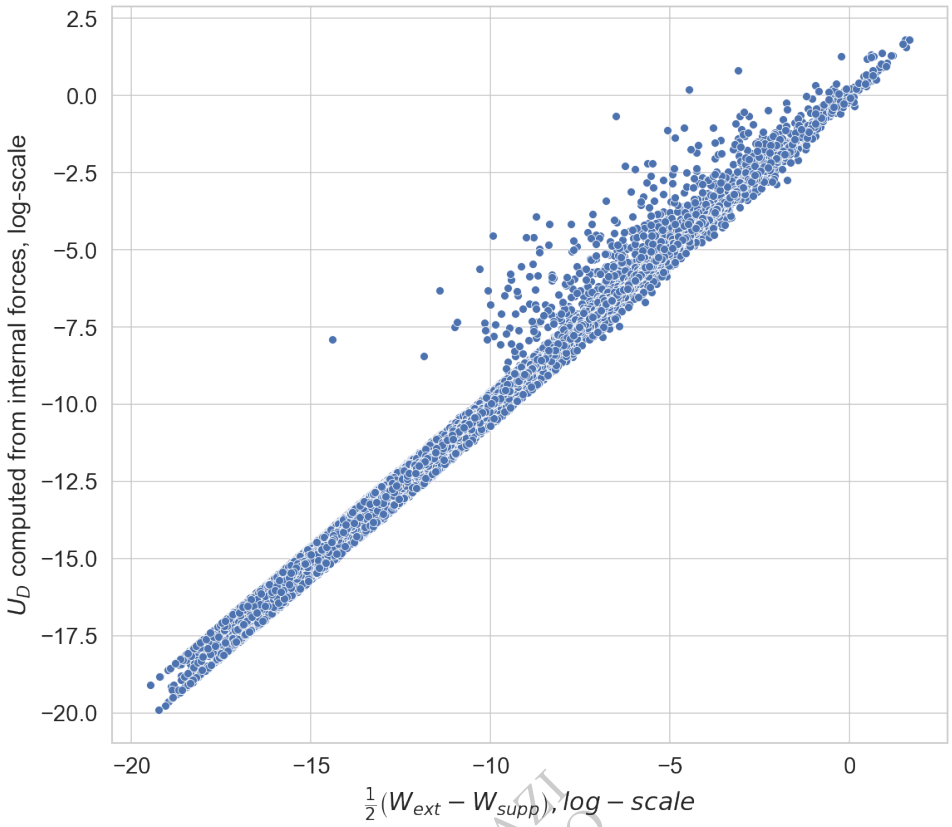

Figure 52: scatter plot of $\hat{U}_D$ vs $\frac{1}{2}(W_{ext} - W_k)$ across all 60000 instances, in log-scale.

V. DERRAZI
D. PELUFFO
J.C. TORRES

5 Baseline machine learning experiment

To demonstrate the utility of the dataset for data-driven structural mechanics, a baseline feedforward neural network surrogate is trained on load case 2, corresponding to a unit transverse point load $\vec{F} = \vec{e}_y$. By virtue of the constitutive decoupling established earlier, this load case activates only the bending response in the $(G, \vec{e}_y)$ plane, with u_y^G, θ_z, V_y and M_z as the non-trivial output fields. The scalar strain energy $\hat{U}_D$, recomputed from the internal force fields via the trapezoidal procedure described in the energy reconstruction section, is included as an additional regression target, yielding a total output dimension of 85.

5.1 Example protocol

The input feature vector is reduced to the eight quantities that govern the LC2 response: the normalised beam length $L/L_{\max}$, the relative load position α, the normalised cross-sectional properties $I_{zz}/I_{zz,\max}$ and $A_y/A_{y,\max}$, and the four dimensionless support stiffness multipliers κ_y^A, κ_y^B, ζ_z^A, ζ_z^B mapped to $[0, 1]$ via $(\log_{10}(\cdot) - \ell_{\min})/(\ell_{\max} - \ell_{\min})$ with $\ell_{\min} = -1$ and $\ell_{\max} = 5$. Output quantities are adimensionalised per instance prior to training: transverse displacements by $L^3/(EI_{zz})$, rotations by $L^2/(EI_{zz})$, bending moments by $1/L$, and strain energy by $L^3/(EI_{zz})$. Shear forces are already of order unity for a unit applied load and require no rescaling.

The dataset is partitioned into a training set of 8000 instances and a test set of 2000 instances using a stratified split by section family, ensuring proportional representation of IPE, HEA and HEB profiles in both subsets. The surrogate model is a fully-connected network with three hidden layers of 256 units each and SiLU activations, trained for 300 epochs using the Adam optimiser with an initial learning rate of 10^{-3}, weight decay 10^{-5}, and a cosine annealing schedule. The mean squared error over all 85 adimensional outputs is used as the training objective.

5.2 Results

Table 3 reports the test set performance per output group in terms of the weighted R^2 score, the root mean squared error in adimensional units, and the median and 95th percentile of the instance-wise relative error $\|\hat{y} - y\|/\|y\|$. The model achieves strong predictive performance on the internal force fields V_y and M_z, which are governed by simple equilibrium relations and exhibit limited variability for a given load position. Kinematic fields u_y^G and θ_z are predicted with good accuracy in the elastically restrained regime, where deflections are large relative to the adimensionalisation scale; prediction quality degrades toward the quasi-clamped boundary, where u_y^G becomes small and the signal-to-noise ratio of the network output decreases. This behaviour is an expected consequence of the wide dynamic range of the transverse displacement distribution documented in the exploratory data analysis, and motivates the use of sign-preserving output transformations or dedicated per-field architectures in future work.

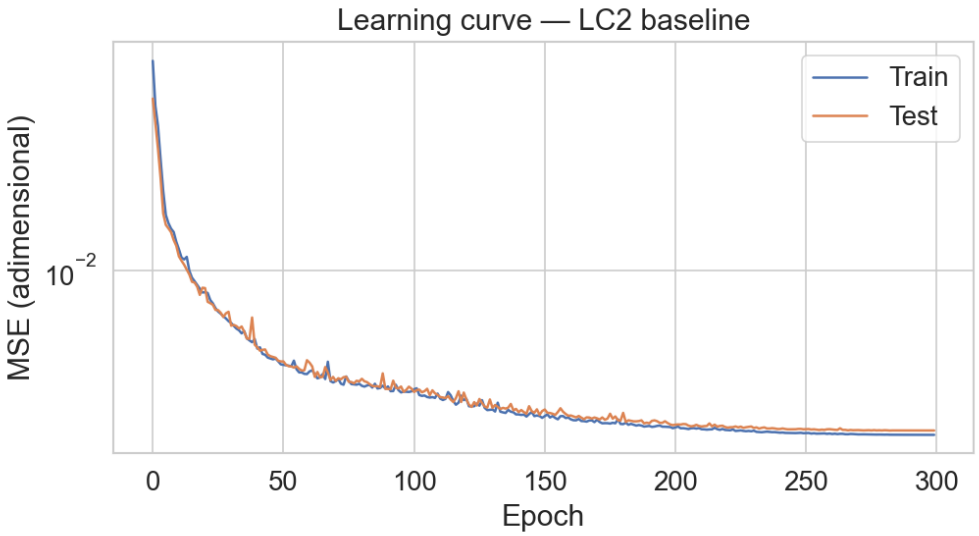

Figure 53: Learning curve MSE for both training set and validation set.

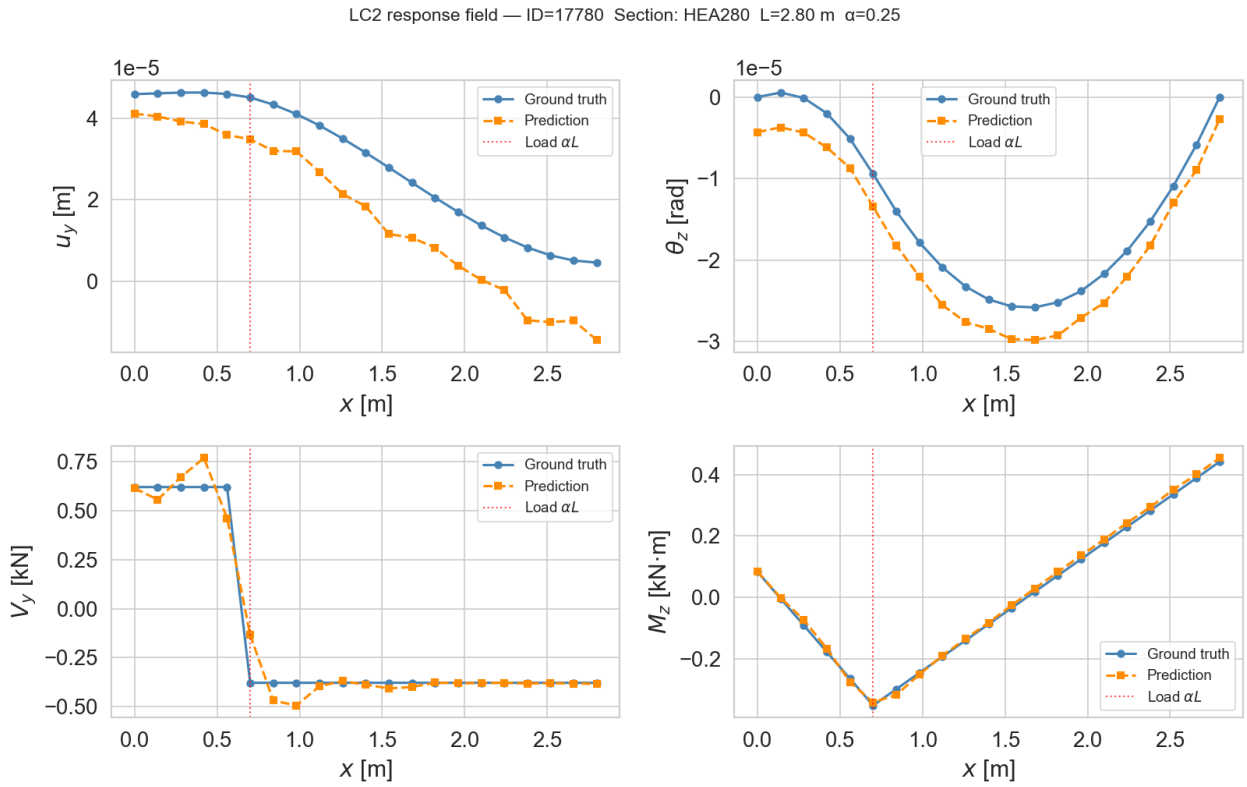

Figure 54: Response reconstruction for an arbitrary case.

Table 3: Test set metrics for the LC2 baseline surrogate. R^2 is the variance-weighted coefficient of determination; rel. error is $\|\hat{y} - y\|/\|y\|$ computed per instance.

Output	R^2	RMSE (adim.)	Rel. error median	Rel. error p95
u_y^G	0.9945	$2.239 \cdot 10^{-2}$	$2.061 \cdot 10^{-1}$	2.555
θ_z	0.9689	$1.541 \cdot 10^{-2}$	$1.609 \cdot 10^{-1}$	1.456
V_y	0.9601	$8.741 \cdot 10^{-2}$	$1.744 \cdot 10^{-1}$	$3.002 \cdot 10^{-1}$
M_z	0.9886	$1.373 \cdot 10^{-2}$	$9.075 \cdot 10^{-2}$	$3.995 \cdot 10^{-1}$

5.3 Discussion

The learning curve shown in Figure 53 confirms that the model converges smoothly over 300 epochs, with training and test losses tracking each other closely throughout and no sign of overfitting.

The per-field performance is however heterogeneous, as summarised in Table 3. The internal force fields V_y and M_z are predicted with high accuracy, achieving R^2 scores of 0.9601 and 0.9886 respectively and median relative errors below 1.8%. This is consistent with the physically constrained structure of these fields: V_y is piecewise constant with a single discontinuity of unit magnitude at α, and M_z is piecewise linear, so both are essentially low-complexity functions of the input parameters that a shallow network can represent efficiently. The bending rotation θ_z is predicted with good accuracy in shape, achieving $R^2 = 0.9689$, though a mild amplitude bias is visible in individual instance reconstructions such as Figure 54. The transverse displacement u_y^G is the most challenging output, with $R^2 = 0.9945$ and a median relative error of 2.06%, and exhibits the most pronounced discrepancy between prediction and ground truth in individual instances.

This performance hierarchy is a direct consequence of the wide dynamic range of u_y^G across the dataset, which spans approximately eight decades. The adimensionalisation by $L^3/(EI_{zz})$ reduces but does not eliminate this range, since the elastic support conditions introduce additional variability that is not captured by the section and length scaling alone. In the quasi-clamped regime, u_y^G becomes small relative to the adimensionalisation scale

and the network's absolute prediction error, while small in adimensional terms, translates into a large relative error upon redimensionalisation. A sign-preserving output transformation such as the inverse hyperbolic sine, or a dedicated architecture per field group, would likely improve performance in this regime and is left as a direction for future work.

6 Usage Notes

6.1 Superposition and generalization

The dataset is designed around the principle of superposition, which constitutes its primary usage mechanism. Each instance provides the complete structural response to a unit excitation (either a unit point load or a unit point moment) applied at a specific relative position α along the beam axis. By virtue of the linearity assumptions underlying the TIMOSHENKO beam model, the response of the same beam to any arbitrary loading distribution $q(\alpha)$ can be reconstructed by quadrature:

$$y(x) = \int_0^1 q(\alpha)\, Y(\alpha, x)\, L\, \mathrm{d}\alpha, \tag{75}$$

where $Y(\alpha, x)$ is the response field retrieved from the dataset for a unit excitation at position α, and $y(x)$ is the response to the generalized load distribution. In discrete form, this becomes a weighted sum over the 19 admissible load positions $\alpha \in \{0.05, 0.10, \ldots, 0.95\}$:

$$y(x) \approx L \sum_{k=1}^{19} q(\alpha_k)\, Y(\alpha_k, x)\, \Delta\alpha, \tag{76}$$

with $\Delta\alpha = 0.05$. This makes the dataset directly usable as a discrete influence function library for any beam instance in the dataset, without requiring additional solves.

The six load cases are designed to be complete in the sense that any generalized point load or point moment can be decomposed into a linear combination of the six elementary unit cases. For a general point load $\vec{F} = F_x\vec{e}_x + F_y\vec{e}_y + F_z\vec{e}_z$ and point moment $\vec{C} = C_x\vec{e}_x + C_y\vec{e}_y + C_z\vec{e}_z$ applied at position α, the full response is obtained by scaling and summing the responses of the six load cases:

$$y(x) = F_x Y_1(\alpha, x) + F_y Y_2(\alpha, x) + F_z Y_3(\alpha, x) + C_x Y_4(\alpha, x) + C_y Y_5(\alpha, x) + C_z Y_6(\alpha, x), \tag{77}$$

where $Y_k(\alpha, x)$ denotes the response field for load case k. Combined with the quadrature formula above, this allows the reconstruction of responses to fully general three-dimensional distributed loading scenarios from the dataset alone.

It should be noted that the superposition principle applies strictly under the geometric and material linearity assumptions of the model. Extensions to nonlinear regimes would require separate datasets or correction procedures and are beyond the scope of the present work.

6.1.1 Illustrative example

To illustrate this principle concretely, Figure 55 shows the response fields predicted by the trained LC2 surrogate for a combined loading scenario consisting of two simultaneous unit transverse point loads of magnitudes F_1 and F_2 applied at relative positions α_1 and α_2 respectively, on an IPE300 beam of length L with elastic supports characterised by dimensionless multipliers κ and ζ.

The combined response is obtained by querying the network independently at α_1 and α_2 and forming the linear combination $\hat{Y}_{\text{super}} = F_1\hat{Y}(\alpha_1) + F_2\hat{Y}(\alpha_2)$. The predicted fields are compared against an independent finite element solve performed in KARAMBA3D under the identical combined loading, confirming that the superposition principle is satisfied to high accuracy by the trained surrogate. This demonstrates that a model trained exclusively on elementary unit load cases can be deployed as an influence function library, reconstructing responses to arbitrary multi-point loading scenarios without retraining or additional solves.

Y. DERRAZI
D. PELUFFO
J.C. TORRES

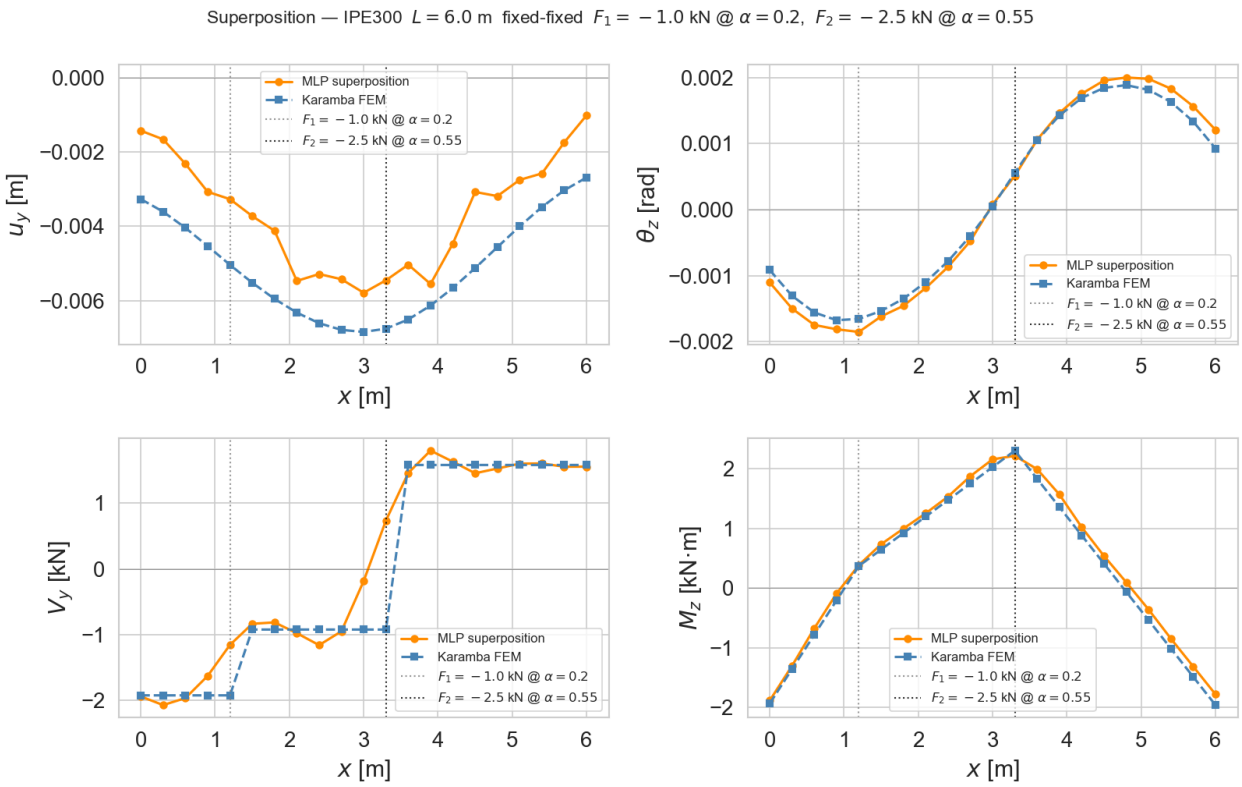

Figure 55: Superposition principle with an already trained network.

L. DERRA, R. PELUFFO, L. TORRES

6.1.2 Discussion

The superposition experiment shown in Figure 55 provides a complementary assessment of the model's generalisation capability. A combined loading scenario consisting of two simultaneous transverse point loads of magnitudes $F_1 = 1kN$ and $F_2 = 2.5kN$ applied at relative positions $\alpha_1 = 0.2$ and $\alpha_2 = 0.55$ on an IPE300 beam of length $L = 6.0m$ with dimensionless support stiffness multipliers $\kappa = 100$ and $\zeta = 10$ is reconstructed by querying the network independently at α_1 and α_2 and forming the linear combination $\hat{Y}_{\text{super}} = F_1\hat{Y}(\alpha_1) + F_2\hat{Y}(\alpha_2)$. This geometry and loading combination lie entirely outside the training set, as the dataset contains only elementary unit load cases. The predicted fields are compared against an independent finite element solve performed in Karamba3D under the identical combined loading. The internal force fields and bending rotation are reproduced with excellent agreement, while the transverse displacement exhibits the same amplitude bias documented above. These results confirm that the superposition principle is satisfied by the trained surrogate to the same level of accuracy as the single-load predictions, and that the model can serve as an influence function library for arbitrary multi-point loading scenarios within the elastically restrained regime covered by the dataset.

6.2 Potential applications

The dataset is intended to support the development and benchmarking of data-driven models for structural mechanics. Several application categories are directly enabled by its design.

Surrogate modelling. The dataset provides a large-scale, high-dimensional input-output mapping between beam parameters and structural responses, making it well-suited for training regression-based surrogate models. The full spatial distribution of responses rather than scalar outputs enables the training of function-valued surrogates, including neural operators and Gaussian process regression models that operate on spatially distributed fields. The superposition structure further allows a trained surrogate to generalize to arbitrary loading distributions without retraining.

Physics-informed neural networks. The dataset is particularly well-suited for benchmarking PINNs applied to beam problems. The governing equations are fully specified, providing the physical constraints required to construct the PINNs loss function. The wide range of boundary conditions sampled through the elastic

support formulation allows systematic evaluation of PINNs generalization across the full BC continuum, from nearly free to quasi-clamped configurations. The known failure mode of PINNs when the residual cannot be reduced to zero (where squaring the residual in the loss function kills gradient information) can be studied directly against the dataset's ground truth.

Graph neural networks. The beam instances in the dataset can be naturally represented as graphs, with the 21 sampling positions along the beam axis as nodes and the beam element as the connecting edge carrying the geometric and material attributes. This makes the dataset suitable for training and benchmarking GNNs for structural response prediction, particularly for architectures that aim to generalize across varying boundary conditions and loading scenarios.

Operator learning. The influence function structure of the dataset (where each instance maps a load position to a full response field) is a natural fit for neural operator architectures, which are designed to learn mappings between function spaces. The dataset provides a clean and physically grounded benchmark for evaluating such architectures on a structural mechanics problem with known analytical structure.

Non-dimensionalization. For machine learning applications, it is recommended to non-dimensionalize the input and output quantities before training. Natural non-dimensional groups include the shear stiffness ratios ϕ_y and ϕ_z defined earlier, the dimensionless support stiffness multipliers κ and ζ already stored in the input file, and response quantities normalized by characteristic scales such as $L^3/(EI)$ for displacements and $L/(EI)$ for rotations. The decoupling structure of the dataset (where each load case activates a distinct subset of output quantities) should also be exploited by training separate models per load case rather than a single model across all six cases simultaneously.

6.3 Limitations and future extensions

Several limitations of the present dataset should be kept in mind when using it for training or benchmarking.

Y. DERRAZI
D. PELUFFO
J.C. TORRES

Material and geometric linearity. The dataset is generated under the assumptions of geometric and material linearity. Nonlinear effects such as large displacements, plasticity, buckling, and contact are not represented. The dataset is therefore not suitable for training models intended to capture nonlinear structural behaviour without significant extension or correction.

Single-span prismatic beams. Each instance corresponds to a single-span beam with a constant cross-section along its entire length. Multi-span beams, tapered members, haunched sections, and beams with intermediate supports or releases are not represented. Extensions to these configurations would require a more general parametric model and are left for future work.

Steel material only. The constitutive material is fixed to standard steel with $E = 210$ GPa and $\nu = 0.33$ throughout the dataset. The effect of material variability is not captured. Since the governing equations are linear, responses for other isotropic elastic materials can in principle be recovered by rescaling, provided the shear stiffness ratios ϕ_y and ϕ_z are preserved.

European steel profiles only. The cross-section catalogue is restricted to IPE, HEA and HEB profiles from the European standard. Other profile families are not represented. Since the dataset stores the full set of geometric properties for each section, the results are transferable to any prismatic section with known S, A_y, A_z, I_{yy}, I_{zz} and J_{xx}, provided the section falls within the geometric range covered by the catalogue.

Warping effects neglected. The torsion model used in this work is the uniform Saint-Venant torsion model, which neglects warping. For open thin-walled sections such as IPE profiles under torsional loading, warping can contribute significantly to the torsional response, particularly for short beams or beams with restrained warping at the supports. The Vlasov torsion model, which accounts for warping through the warping constant I_w, would be required for a more accurate representation of torsional behaviour in such cases.

Sampling resolution. The response fields are sampled at 21 equally spaced positions $\Delta\xi = 0.05$ along the beam axis. While this resolution is sufficient for smooth fields, it introduces a truncation error at the load application point α where shear forces are discontinuous. This error is localized and of order $O(\Delta\xi^2)$, but users requiring higher resolution near the load application point should account for this limitation.

Future extensions. Natural extensions of this dataset include the incorporation of additional profile families and custom sections, the introduction of material variability as a sampled parameter, the extension to multi-span beams and frame structures, the inclusion of distributed loading scenarios directly rather than through superposition, and the incorporation of warping effects through the VLASOV torsion model or any other enriched beam models. The reproducible pipeline described in this paper provides a direct foundation for such extensions.

Data Availability

The TimoDS dataset [22] is publicly available on Zenodo under the Creative Commons Attribution 4.0 International (CC BY 4.0) license: https://doi.org/10.5281/zenodo.19324944 The repository contains two CSV files: `input_data_TimoDS.csv` and `beam_results_TimoDS.csv`, joinable on the common identifier column `ID` = `row_id`.

Code Availability

The Python input generation script and the GRASSHOPPER pipeline definition used to produce TimoDS are publicly available at: https://github.com/YDERRAZI/TimoDS_pipeline under the MIT License.

Y. DERRAZI
D. PELUFFO
J.C. TORRES

References

[1] Adeli, H. & Yeh, C. Perceptron learning in engineering design. *Computer-Aided Civil and Infrastructure Engineering* **4**, 247–256 (1989).

[2] Adeli, H. Neural networks in civil engineering: 1989–2000. *Computer-Aided Civil and Infrastructure Engineering* **16**, 126–142 (2001).

[3] Thai, H.-T. Machine learning for structural engineering: A state-of-the-art review. *Structures* **38**, 448–491 (2022).

[4] Sun, H., Burton, H. V. & Huang, H. Machine learning applications for building structural design and performance assessment: State-of-the-art review. *Journal of Building Engineering* **33**, 101816 (2021).

[5] Málaga-Chuquitaype, C. Machine learning in structural design: an opinionated review. *Frontiers in Built Environment* **8**, 815717 (2022).

[6] Zheng, H., Moosavi, V. & Akbarzadeh, M. Machine learning assisted evaluations in structural design and construction. *Automation in Construction* **119**, 103346 (2020).

[7] Zabin, A., González, V. A., Zou, Y. & Amor, R. Applications of machine learning to BIM: A systematic literature review. *Advanced Engineering Informatics* **51**, 101474 (2022).

[8] Feng, D. & Feng, M. Q. Computer vision for SHM of civil infrastructure: From dynamic response measurement to damage detection – A review. *Engineering Structures* **156**, 105–117 (2018).

[9] Spencer Jr, B. F., Hoskere, V. & Narazaki, Y. Advances in computer vision-based civil infrastructure inspection and monitoring. *Engineering* **5**, 199–222 (2019).

[10] Le-Duc, T., Nguyen-Xuan, H. & Lee, J. A finite-element-informed neural network for parametric simulation in structural mechanics. *Finite Elements in Analysis and Design* **217**, 103904 (2023).

[11] Liang, R., Liu, W., Fu, Y. & Ma, M. Physics-informed deep learning for structural dynamics under moving load. *International Journal of Mechanical Sciences* **284**, 109766 (2024).

[12] Qin, S. et al. Intelligent design for component size generation in reinforced concrete frame structures using heterogeneous graph neural networks. *Automation in Construction* **171**, 105967 (2025).

[13] Liao, W., Lu, X., Huang, Y., Zheng, Z. & Lin, Y. Automated structural design of shear wall residential buildings using generative adversarial networks. *Automation in Construction* **132**, 103931 (2021).

[14] Liao, W., Lu, X., Fei, Y., Gu, Y. & Huang, Y. Generative AI design for building structures. *Automation in Construction* **157**, 105187 (2024).

[15] Rabie, M., Almutairi, F. F., Tsavdaridis, K. D. & Shaaban, I. G. Machine learning-driven capacity design and embodied carbon reduction optimization in composite reduced web section (RWS) connections. *Advances in Engineering Software* **217**, 104145 (2026).

[16] Tsavdaridis, K. D. & Papadopoulos, T. A FE parametric study of RWS beam-to-column bolted connections with cellular beams. *Journal of Constructional Steel Research* **116**, 92–113 (2016).

[17] Gongora, A. E. et al. Accelerating the design of lattice structures using machine learning. *Scientific Reports* **14**, 13703 (2024).

[18] Rasouligandomani, M. et al. Dataset of finite element models of normal and deformed thoracolumbar spine. *Scientific Data* **11**, 549 (2024).

[19] Preisinger, C. & Heimrath, M. KARAMBA – a toolkit for parametric structural design. *Structural Engineering International* **24**, 217–221 (2014).

[20] Conde-Rodríguez, F., García-Fernández, Á. L. & Torres, J. C. Modeling the internal architecture of composites. *Computer-Aided Design* **129**, 102930 (2020).

[21] López, L., Torres, J. C., Arroyo, G., Cano, P. & Martín, D. An efficient GPU approach for designing 3D cultural heritage information systems. *Journal of Cultural Heritage* **41**, 142–151 (2020).

[22] Derrazi, Y., Peluffo, D. H. & Torres, J. C. TimoDS: A comprehensive and synthetically generated dataset of TIMOSHENKO steel beams with general elastic supports and responses to elementary unit load cases. Zenodo https://doi.org/10.5281/zenodo.19324944 (2025).

Y. DERRAZI
D. PELUFFO
J.C. TORRES

Author contributions

Youssef Derrazi : Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing - original draft, visualization. **Diego Hernan Peluffo Ordonez :** methodology, validation, resources, writing - review & editing, supervision, project administration, funding acquisition. **Juan Carlos Torres :** methodology, validation, writing - review & editing, supervision, project administration.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was carried out within the SDAS Research Group and the Universidad de Granada. The authors thank Abdelkrim Alahyane for the discussions and support around exploratory data analysis. Structural analysis was performed using Karamba3D within the Grasshopper parametric environment for Rhinoceros 3D; the authors acknowledge the use of these tools in the generation of the dataset. No external funding was received in support of this work.

Y. Derrazi
D. Peluffo
J.C. Torres