INTERNATIONAL TRANSACTIONS IN OPERATIONAL RESEARCH

WILEY

INTERNATIONAL TRANSACTIONS IN OPERATIONAL RESEARCH

Intl. Trans. in Op. Res. 0 (2025) 1–27
DOI: 10.1111/itor.70059

Standing on a common ground: a comparison of static stability approaches for pallet loading

Philipp G. Mazur[a,*], Frederick C. Gamer[a], António G. Ramos[b] and Detlef Schoder[a]

[a]*Cologne Institute for Information Systems, 50969 Cologne, Germany*
[b]*INESC TEC, ISEP, Polytechnic of Porto, 4249-015 Porto, Portugal*
E-mail: mazur@wim.uni-koeln.de[Mazur]; gamer@wim.uni-koeln.de[Gamer]; agr@isep.ipp.pt[Ramos]; schoder@wim.uni-koeln.de[Schoder]

Received 9 August 2024; received in revised form 17 March 2025; accepted 11 May 2025

Abstract

At the practical level, the static stability constraint is one of the most important constraints in practical pallet loading problems, such as air cargo palletizing. Approaches to modeling static stability, which range from base support and mechanical equilibrium calculations to physical simulation, differ in workflow, focus, and assumptions, so choosing the right static stability approach has a substantial impact on the quality of the solution and, ultimately, on loading security. To date, little research has investigated the structural differences between approaches. The aim of this paper is to integrate knowledge and shed light on the applicability of the different approaches for the practical scenario of air cargo palletizing. We tackle this problem through (1) a reformulation and extension of static stability toward loading stability, (2) a conceptual analysis of current approaches, and (3) benchmarking that employs an independent multibody simulation on multiple heterogeneous datasets. Our results show that all approaches are prone to structure errors and vary significantly in their premises and information usage. Further, full base support is revealed to be the most restrictive approach by far, while physical simulation achieves the greatest accuracy. Given the trade-off between accuracy and runtime, the mechanical equilibrium approach is a good choice, while partial base support performs best for lower support values.

Keywords: static stability; pallet loading problem; physical simulation; loading stability; air cargo

1. Introduction

Safely transporting cargo is a challenging task across industries and transportation modes (e.g., air, maritime, road), especially when cargo is arranged on pallets. For this pallet loading problem (PLP),

*Corresponding author.

important constraints must be addressed and regulations met to ensure safe transport (Bortfeldt and Wäscher, 2013; Lee et al., 2021). This article considers the PLP from an air cargo carrier's perspective. This problem can be classified as a (single) distributor's PLP (Hodgson, 1982) and entails the loading of three-dimensional cargo items (boxes) onto a single pallet with no lateral support. In this air cargo PLP, the assortment of items is highly heterogeneous (Brandt and Nickel, 2019; Lee et al., 2021), and solutions need to be calculated as soon as possible due to strict flight schedules (Mazur et al., 2020).

In this study, we specifically focus on the static (i.e., vertical) stability constraint. As one of the most important practical constraints, stability keeps (partly) stacked cargo from collapsing (Bischoff, 1991; Ramos et al., 2016a). An assortment of cargo that consists of both three-dimensional item placement positions and loading sequences is also called a *layout*. According to a common definition, a layout is *statically* stable if the loaded cargo items are able to "withstand the gravitational force acceleration over them" (Junqueira et al., 2012, p. 76) when the transport vehicle is not moving. In contrast, a layout is *dynamically* stable if the cargo remains stable during transport (Junqueira et al., 2012). Stability contributes to safe transport and is highly relevant for practical usage since unstable items that tip, slide, rotate, or fall can lead to personnel injuries or damaged cargo (Bischoff and Ratcliff, 1995; Junqueira and de Queiroz, 2022), and it is also very relevant in combination with vehicle routing problems where cargo is removed during the route (Ranck Júnior et al., 2019). Stability is also a strict flight safety regulation, as defined by the International Air Transport Association (IATA, 2019). Thus, it is of utmost importance to ensure proper cargo stability at all times. The static stability constraint has been tackled from various perspectives, including from the container loading problem (CLP) perspective, which describes when a rigid metal container is loaded and its walls provide lateral support for cargo. Since it is closely related to the PLP, we also consider findings related to the CLP.

Multiple approaches in CLP and PLP research have been proposed to quantify and evaluate static stability, including full base support (FBS), partial base support (PBS), static mechanical equilibrium (SME), and physical simulation (PS). These approaches, which differ in terms of performance and modeling characteristics (i.e., they cover relevant forces, consider multiple shapes, or address nonuniform mass distributions), have not been used consistently in recent related studies. Specifically, in a review, we found that FBS/PBS (Olsson et al., 2020; Rajaei et al., 2024; Yang et al., 2024), SME (Ramos et al., 2016a), PS (Bracht et al., 2016; Mazur et al., 2022), a combination of approaches (Krebs and Ehmke, 2021; Oliveira et al., 2021), and none of these approaches (Araya et al., 2020) have been used in related studies. Both the unclear strengths and limitations of current static stability approaches and their inconsistent use in related studies speak to the need for a structured comparison to find out which static stability approach most accurately predicts the safe loading of cargo on pallets and why. Previous comparisons of different static stability approaches have either explored the impact of support value levels on static stability (Junqueira and de Queiroz, 2022) or have investigated the impact of a static stability approach on the volume utilization of the resulting optimization (Junqueira et al., 2012; Ramos et al., 2016a; Oliveira et al., 2021) while neglecting any direct comparison of the approach against a benchmark. Such a benchmark helps researchers and practitioners gain clearer insights into the strengths and weaknesses of different approaches and thus select the best approach for their respective use cases, which, in our case, is the air cargo PLP.

Our aim is to make the benchmark as realistic as possible, so we integrate insights from a practical cooperation from the air cargo industry. Instead of only focusing on idealized mechanical

conditions (e.g., only the gravitational force acts as external force), we formulate this loading stability as the requirement to safely and consecutively load cargo given a realistic air cargo–based loading setting that includes item collisions and nonuniform mass distributions and prevents corner situations that would not be feasible to load in practice. Consequently, we reformulate and extend the classic static stability problem toward a dynamic loading stability problem, introduce a novel metric to capture loading stability, and design a benchmark simulation that virtually and consecutively loads cargo layouts on pallets. We also evaluate cargo items' ability to indeed withstand the gravitational force acceleration over them (Junqueira et al., 2012) by measuring the number of items that tip, slide, rotate, and fall (Bischoff and Ratcliff, 1995; Bortfeldt and Wäscher, 2013; Junqueira and de Queiroz, 2022). We compare the different static stability algorithms against this loading stability benchmark and evaluate whether the different approaches correctly predict the instability of cargo items.

To fulfill these goals, this paper is organized as follows. The next section presents a literature review covering relevant basics about the different stability approaches and previous comparisons. Afterward, we formulate the problem. We then elaborate on the conceptual differences among the different approaches, followed by a presentation of our benchmark results, a discussion, and a brief concluding section.

2. Related literature

2.1. Base support

The base support approach has traditionally been the choice for ensuring static stability. It can be subdivided into the FBS and the PBS. FBS and PBS require a minimum amount of the lower face of each cargo item to be supported from below. An item can be supported through one or multiple item top faces or the pallet. FBS requires the entire base of an item to be supported (Bortfeldt and Wäscher, 2013). The support factor $\alpha \in [0, 1]$ captures the minimal amount of static stability required (Junqueira et al., 2012). For FBS, every cargo item has a support value of $\alpha = 1$. FBS ensures that the conditions for static stability are met and that stable layouts are created (de Queiroz and Miyazawa, 2014; Ramos et al., 2016a). Relaxing this requirement can produce unstable layouts (De Castro Silva et al., 2003). However, FBS is not only restrictive since no overhanging items are permitted but also disadvantageous for highly heterogeneous problem cases (Ramos et al., 2016a). It can negatively influence the load factor, especially for strongly heterogeneous cases, and implies lower overall solution quality (Ramos et al., 2016a). Due to its simplicity, numerous studies have employed FBS (e.g., Gehring and Bortfeldt, 1997; Moura and Bortfeldt, 2017; Rajaei et al., 2024; Yang et al., 2024). To overcome the disadvantages of FBS, many studies have adopted a support value of $\alpha < 1$ (Junqueira and de Queiroz, 2022), which is termed PBS. PBS is less restrictive than FBS as it allows overhanging items and bridging item placements (Bortfeldt and Wäscher, 2013). In contrast to FBS, PBS requires only a percentage α of support for the base area. This α level differs from study to study. Typically, α ranges from 0.55 upward. Multiple studies have employed PBS, with α beginning at different levels: 0.55 (Mack et al., 2004), 0.7 (Gehring and Bortfeldt, 1997), 0.75 (Krebs and Ehmke, 2021), and 0.8 (Olsson et al., 2020; Christensen and Rousøe, 2009). In a comparison study, Junqueira and de Queiroz (2022) tested different α levels for the constrained

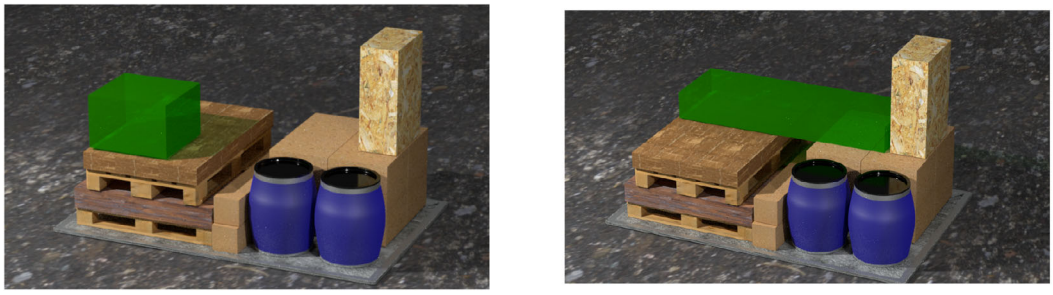

Fig. 1. Exemplary FBS and PBS air cargo item-loading situations. The focal, transparent green item is placed such that its entire bottom face (FBS) or a minimum percentage α (PBS) is supported from below. Note that PBS allows for overhanging items.

two-dimensional packing problem to determine the minimal support levels necessary to guarantee static stability. Their results revealed that a support factor of $\alpha = 0.42$ is sufficient to ensure a statically stable solution with a probability of at least 90%. In Fig. 1, we depict two illustrative item-loading situations.

Other studies have combined the base support approach with other factors. For example, Paquay et al. (2016) introduced the requirement of at least three supported corners of the lower box face. Thereby, the focus shifts from supported areas to supported vertices. Olsson et al. (2020) presented similar approaches, sometimes in combination with a minimal required supported area. Further, Zhu et al. (2024) proposed a heuristic of dividing the lower face of a box into 2×2 regions and ensuring that at least three out of the four regions are supported.

2.2. Static mechanical equilibrium

The static mechanical equilibrium approach assumes that (1) only the gravitational force operates on the objects and (2) that they lie at rest o a constant speed, so we can examine static stability from the perspective of static material bodies (Junqueira and de Queiroz, 2022). The equilibrium calculations are derived from the first and third of Newton's laws of motion. Formally, an equilibrium of bodies is achieved if both (1) the sum of external forces $\sum \vec{F} = \vec{0}$ and (2) moments acting on the item $\sum \vec{M}_O = \sum(\vec{r} \times \vec{F}) = \vec{0}$ are zero. $\vec{r}$ represents the vector from point O to the force vector $\vec{F}$, and $\vec{M}_O$ represents the moment of force $\vec{F}$ on point O. Let O be the origin of the coordinate system. The conditions ensure that the body neither translates nor rotates (Hibbeler, 2010). Further, if only the gravitational force acts as an external force and if the body has a uniform density since a system's mass can be treated as point mass, the weight force can be summarized as acting downward on the item's geometric center.

For an exemplary item b, there are three types of relevant forces involved: weight forces ($\vec{W}$), action forces ($\vec{A}$), and reaction forces ($\vec{R}$). Weight forces act downward at the item's center of gravity (G). Action forces are exerted on item b from above by the supported item k. Reaction forces act on item b from below by the support of item j. Since an item can support multiple items and be supported by multiple items, there can be multiple action and reaction forces. The overall action and reaction forces are then the sum of all action and reaction forces obtained as $\vec{A} = \sum_{k=1}^{K} \vec{A}_k$

and $\vec{R} = \sum_{j=1}^{J} \vec{R}_j$. The overall moments from action and reaction forces about point O, applied to points M and Q, are consequently given by the following: $\vec{M}_A = \vec{A} \times O\vec{M} = \sum_{k=1}^{K} (\vec{A}_k \times O\vec{M}_k)$ and $\vec{M}_R = \vec{R} \times O\vec{Q} = \sum_{j=1}^{J} (\vec{R}_j \times O\vec{Q}_j)$

The SME approach calculates the resultant force $\vec{F}_R$ and the resultant force point P. The magnitude of $\vec{F}_R$ is obtained as $\vec{F}_R = \vec{W} + \vec{A}$ and combines the magnitude of weight and action forces. The moment of $\vec{F}_R$ about point O, $\vec{M}_{F_R}$ combines the moments of $\vec{W}$ ($\vec{M}_W$) about point O and $\vec{A}$ ($\vec{M}_A$) about point O, so $\vec{M}_{F_R} = \vec{M}_W + \vec{M}_A$ about O. The moment $\vec{M}_W = \vec{W} \times O\vec{G}$ acts on the item's center of gravity G. The moment $\vec{M}_A = \vec{A} \times O\vec{M}$ acts on the point M where action forces are applied. The point where the resultant force acts can be calculated using the relationship $\vec{M}_{F_R} = \vec{F}_R \times O\vec{P}$, so $\vec{F}_R \times O\vec{P} = \vec{W} \times O\vec{G} + \vec{A} \times O\vec{M}$. Given the resultant force point P, SME assumes an item to be in a mechanical equilibrium if one of the following three conditions is fulfilled, which require an item's center of mass to be positioned in a stable region (De Castro Silva et al., 2003; de Queiroz and Miyazawa, 2014; Ramos et al., 2016a): (1) the item lies on a pallet or floor, (2) the orthogonal projection of the item's center of gravity to its base lies directly on another item's top surface, or (3) a virtual plane exists that connects two nonadjacent items and goes through the orthogonal projection of the center of gravity of an item to its base (De Castro Silva et al., 2003).

Multiple SME algorithms currently exist. For the two-dimensional case, de Queiroz and Miyazawa (2014) proposed calculating the forces for any item given there are items placed on top of it. An item is stable if its center of mass lies on a stable region (i.e., on top of one of its supporting items, between two of its supporting items, or on the floor). They proposed an iterative approach moving from top items to bottom items and calculating if the center of mass is in a stable region. For the three-dimensional case, the difference between different algorithms lies in the forces considered and the equilibrium assessment mechanisms used. De Castro Silva et al. (2003) published the first algorithm based on equilibrium conditions that evaluate the physical stability of a layout. Their algorithm inserts items into a bin in a feasible (stable) position that generates the least wasted space. Ramos et al. (2016a) argued that there is a lack of method to check if point P is supported by a vertical plane (the third condition above) and the need to control the equilibrium of moments. Therefore, they introduced a support polygon, which is given by the convex hull of all contact points on the item base. For stability, the resultant force point must lie within the support polygon. If a new item is placed, every subset of items $A_r \subseteq A$ previously loaded must be evaluated to see if the new subset is stable. Combining gravitational and action forces, the authors calculated both a resultant force and a resultant force point. They did not, however, present any details on their calculation. To control the third condition, the authors proposed testing whether the resultant point lies on a support polygon that comprises the convex hull of all intersection points of the items with the items below. To ensure static stability during loading, each placement requires an entire stability assessment for every item directly or indirectly affected. Oliveira et al. (2021) proposed two static stability approaches based on normal forces in dynamics simulations of rigid bodies and the equilibrium of buildings. The authors proposed employing the conditions for correct normal forces on contact points: the normal forces are equal to the reaction forces; they restrict the interpenetration of objects, so they must be positive and must not change their sign; they push but do not pull; they act only on contact points; and they are continuous in time. Krebs and Ehmke (2021) assessed stability only for a fully loaded pallet but required an item's point of gravitational force

to be supported at every level below the item. The authors combined the SME approach with a minimal support base factor of 75%. They included the minimal support base factor for the following reason: The SME approach only checks if point P is supported, ignoring how much area around this point is actually supported. Therefore, an item would be rated as stable even if it is supported by only a small box so long as point P is supported. Any horizontal forces could tilt the item immediately. Horizontal forces can occur, for example, if a packer unintentionally touches the loading arrangement slightly while loading the next item. This algorithm also uses the center of mass but without the need to determine the action forces nor the need to test the stability after every placement. In a recent study, Le Jean et al. (2024) considered the stability problem for items that have an edge reduction, meaning the upper surface is smaller than the lower surface and item weights and sequences are not known when calculating the layout. The authors introduced a novel stability constraint for their use case, which is called recursive stability, based on the conditions of having full support, all corners supported, or no void under the center of gravity.

2.3. Physical simulation

A few studies have employed PS for static and dynamic stability calculations. PS employs real-time physics engines, such as Bullet or PhysX, to simulate cargo responses to a multitude of forces. Real-time physics engines can compute objects' spatial positions in a plausible approximative but performant way. They offer a reasonable amount of precision to obtain realistic physical feedback for rigid body scenarios. They build up a virtual representation of a physical scene with physical relationships and events equivalent to the real world. At the core of physics engines are dynamic simulations, which represent physical systems within mathematical models that are solved numerically in computer programs (Witkin and Baraff, 1997). Physics engines orchestrate and manage a variety of tasks to simulate the basic elements of Newtonian mechanics for rigid body dynamics, soft body dynamics, and fluids (Martinez-Franco and Alvarez-Martinez, 2018).

Physical simulations for static and dynamic stability use predefined amounts of time t_{max} and compare items' positions at t_{max} to their initial states at t_0, usually measured in Euclidean distance. From this, a number of metrics can be calculated, such as the number of fallen boxes or the number of items within the damage boundary curve.

Multiple studies have employed PS, mostly in the domain of dynamic stability. Ramos et al. (2017) proposed a standalone simulation software prototype (StableCargo) for evaluating the dynamic stability of container cargo that builds on the physics engine Bullet. They benchmarked their physics engine–based standalone tool against a high-precision engineering simulator (Autodesk Inventor[1]) and analytical solutions. They modeled different forces and velocities that represent different movements during transport, and their tool is able to approximate analytical and high-precision solutions.

For static stability, Bracht et al. (2016) performed two stability tests: one local stability test that is executed for every single box in a layout and verifies whether a box's center of gravity is supported by the supporting convex hull of previously packed boxes and one static and dynamic stability verification test using simulation. In a similar study, Mazur et al. (2020) demonstrated the integration of PS in an optimization heuristic for pallet loading. Their prototype employs a physics engine

[1]https://www.autodesk.com/de/products/inventor/

Table 1
Overview of comparative studies of static stability

Study	Problem type	Stability benchmark	Compared approaches			Cargo mass distribution	
			Base support	SME	PS	Uniform	Non-uniform
Junqueira et al. (2012)	CLP		√			√	
Ramos et al. (2016a)	CLP		√	√		√	
Oliveira et al. (2021)	CLP		√	√		√	
Junqueira and de Queiroz (2022)	2CPP[a]		√	√		√	
This study	PLP	√	√	√	√	√	√

[a]Constrained two-dimensional packing problem.

to simulate static stability for entire layouts or item by item, both in combination with a genetic algorithm. Mazur et al. (2022) focused on the acceleration of integrated algorithms using a graphical processing unit (GPU)–accelerated physics engine to evaluate multiple layouts at the same time using GPU parallelization capabilities. A detailed algorithmic description of the simulation-based approach for static stability can be found in Mazur et al. (2020). Although PS promises high accuracy, there is some controversy in the literature regarding the high computational time required for simulation and its integration into optimization algorithms (Ramos et al., 2015; Oliveira et al., 2021).

2.4. Structured static stability comparisons

Studies comparing different static stability approaches are scarce. Most studies have explored the impact of a static stability approach on the optimization's volume utilization but have neglected a direct comparison of static stability approaches. We provide relevant studies that have compared multiple static stability approaches in Table 1. We detail each study, the problem type, whether or not the study employed a benchmark to directly compare static stability approaches, the compared static stability approaches, and whether or not the study included mass distribution information. In one comparison study, Oliveira et al. (2021) introduced two SME approaches derived from the dynamics of rigid bodies and the equilibrium of buildings. The authors measured the efficiency of the approaches based on a container's fill rate and distinguished between the obtained support factors. In a different study, Ramos et al. (2016a) compared different optimization algorithms with and without SME and for supported and unsupported cases. Junqueira et al. (2012) included base support and load-bearing strength in their optimization. Finally, Junqueira and de Queiroz (2022) investigated the relationship between minimal support factors and static stability, which they operationalized using an SME evaluation.

3. Problem statement

3.1. Air cargo PLP

We collaborated with a major German air cargo carrier to conduct our research. Over several months, our research team conducted more than 10 joint workshops with two experts from the

Fig. 2. Air cargo PLP: Strong heterogeneity of items during buildup in an air cargo hub. Most items are boxes or box-shaped, and some are attached to a wooden subpallet. The layout contour can take various forms, for example, cuboidal when placed in the middle of the aircraft main deck or with cutouts when placed next to the rounded aircraft sides.

field of air cargo operations, met three palletizing workers, and made multiple visits to palletizing operations in a large German cargo hub and aircraft loading operations. In the workshops, we developed a common understanding of the item heterogeneity problem and learned about the practical implications of loading stability and its impact on safe air cargo loading and transportation. Our new understanding allowed us to give special consideration to the practical requirements of our research goal, to systematically compare and benchmark different static stability algorithms. Specifically, items arriving at an air cargo hub are highly heterogeneous—they differ in shape, size, packaging material, weight, weight distribution, load bearing, etc. Boxes and rectangular shapes make up most air cargo items. According to experts, items with a rectangular outer hull make up approximately 95% of all items (Mazur et al., 2020). These items are typically packed in cardboard or wood and are often wrapped in plastic foil. They frequently arrive prepalletized on a wooden subpallet (Brandt and Nickel, 2019). The subpallets enable forklift packing, which is the most important form of packing in air cargo and allows for lifting and loading heavy items. Barrels, such as for transporting liquids, are often employed apart from boxes. Barrels have a circular lower face, which leads to different base support calculations. We frequently observed items with displaced centers of mass indicated with a marker. Such items are loaded in an iterative fashion, typically using a forklift. The large objects on or in which items are loaded are either pallets or containers, which are known as unit-loading devices. Pallets are preferred since they can be loaded more easily and do not require a specific contour (Brandt and Nickel, 2019). Pallets are rigid and made out of metal, for example, aluminum. Figure 2 shows three exemplary air cargo buildups. In the following, we formulate the loading stability problem considering the practical requirements of air cargo.

3.2. From static stability to loading stability

Currently, there are multiple definitions of static stability in the literature, which we depict in Table 2. All definitions focus on cargo loading and unloading situations. Some definitions focus on item position changes during loading; they either restrict any changes at all or allow positional

Table 2
Overview of static stability definitions commonly used in the literature

Study	Definition
Ramos et al. (2017)	Static stability refers to the ability of each box to *maintain its loading position* during loading operations.
Krebs and Ehmke (2021)	Vertical stability *prevents items from falling down* on the ground, on top of other items, or on the operator while (un-)loading.
Olsson et al. (2020)	Vertical stability constraints *prevent boxes from falling down* when the container is still.
de Queiroz and Miyazawa (2014)	Cargo stability refers to packing items such that *no item can rotate or fall down* after being packed considering the gravitational force or other forces.
de Queiroz et al. (2019)	For the vertical stability, due to the action of gravity, the bottom side of items must be supported by the top side of other items to *avoid items rotating and falling down*.
Junqueira and de Queiroz (2022)	Static stability implies that the items are organized in such a way that they *do not move (i.e., rotate and fall)* due to the action of gravity.
Martinez-Franco and Alvarez-Martinez (2018)	Freight is considered to be stable if its *geometric configuration and physical integrity are preserved* during loading and unloading operations.

changes with some tolerance. Some definitions focus on cargo rotation and translation as measurable outcomes of the opposite—namely, instability. When comparing the definitions with current static stability approaches, we observe the following inconsistencies:

- Currently, static stability is widely treated in a static, time-invariant fashion. However, according to current definitions, static stability should refer to a loading process and is therefore time-dependent.
- Currently, static stability approaches do not widely make explicit use of the metric item translation or rotation. Since this is a temporal metric, it requires consideration of a dynamic loading process.

Further, from a practical perspective, cargo loading is a dynamic problem since items may collide with each other and since items may slightly (due to soft spring) or considerably (due to falling down) change their initial loading positions and orientations during loading. Contact between items and imperfect rigidity always lead to slightly moving and rotating items. The slightest positional change would be labeled unstable by current approaches but would not be considered so in practice. Problems only arise when items significantly change their positions and rotations. To include the dynamics of the loading process and to account for the rotational and translational elements of static stability, we define loading stability as follows: Loading stability refers to the process of safely loading items on a pallet (or inside a container). If an item can be safely loaded—that is, if it can maintain its spatial position and orientation across time and leads to no *considerable* spatial position and orientation changes in other loaded items—it is stable. If all items with a given sequence in a layout can be safely loaded, the entire layout is stable. Loading stability is a practical requirement; it can be observed and simulated and enables comparison and validation against a benchmark with simulated or real-world data. Static stability is a necessary but not sufficient condition for loading stability.

3.3. Loading stability problem definition

The input for the focal problem consists of a set $\mathcal{A} = \{A_1, \ldots, A_N\}$ of N pre-calculated, given item layouts, each specifying item-loading positions and an item order. Therefore, we model each layout as an item sequence $A_i = (b_{i,1}, \ldots, b_{i,J})$ for $i = 1, \ldots, N$, where each sequence consists of J items $b_{i,j}$ with $j = 1, \ldots, J$. The item sequence is an order in which the items can be loaded consecutively onto the pallet and one layout consists of exactly one item sequence. Thereby, i runs through all layouts in the dataset and j consecutively runs through all items in the loading sequence of one layout. For better readability, we omit explicit indexing and consider an arbitrary layout $A \in \mathcal{A}$ instead.

Every item b_j in the layout has its own given three-dimensional loading position x_j, y_j, z_j; cuboid shape with width w_j, height h_j, and depth d_j; and weight m_j. Item positions, widths, heights, and depths are discrete values and organized in a grid of points with sets X, Y, and Z that include all possible coordinates along the width, height, and depth of the pallet. Items are positioned such that their bottom left coordinate lies on one of the points in the grid. An item's three-dimensional geometric center $CG_j = (CG_{j,x}, CG_{j,y}, CG_{j,z})'$ is obtained as $CG_{j,x} = x_j + w_j/2$, $CG_{j,y} = y_j + h_j/2$, and $CG_{j,z} = z_j + d_j/2$. In the literature, a common assumption about item mass distributions is that each item has a uniform mass distribution, so its geometrical center coincides with its center of mass. In this study, we include a three-dimensional center of mass $CM_j = (CM_{j,x}, CM_{j,y}, CM_{j,z})$, which can be obtained as $CM_{j,x} = CG_{j,x} + r_{j,x}$, $CM_{j,y} = CG_{j,y} + r_{j,y}$, and $CM_{j,z} = CG_{j,z} + r_{j,z}$, where $r_j = (r_{j,x}, r_{j,y}, r_{j,z})'$ represents the displacement from the geometric center to the center of mass. We refer to $r_j = (0, 0, 0)'$ as a centered center of mass and $r_j \neq (0, 0, 0)'$ as a displaced center of mass (DCOM). Items must be loaded entirely within the pallet's dimensions and are placed using 90° rotations on the X-, Y-, and Z-axes. Regarding the items' materials, we assume most items are made of hard materials, such as wood, cardboard, or metal, and that they are sometimes further wrapped in plastic foil. Consequently, we model all items as rigid bodies whose shape cannot be altered or deformed. The pallet is also assumed to be a rigid body. During loading, there is no lateral side support since nets and straps are attached after cargo loading. We assume the pallet contour to be either cuboidal with dimensions W, H, and D, so the total available space can be obtained as $W * H * D$ or is defined by a vertex list on the X and Y axes with a fixed depth $Z \in \{Z_1, Z_2\}$. The problem is now to develop a loading stability metric that integrates a consecutive loading of items, accounts for item dynamics like positions and rotations, and compares current static stability approaches (*ss*) against the loading stability metric (*ls*). We first develop our loading stability metric, before selecting and benchmarking current approaches.

4. Loading stability metric

To represent the consecutive loading state of the layout A at sequence step j, we introduce the loaded subsequence $A_j \subseteq A$, defined as

$$A_j = (b_0, \ldots, b_j) \tag{1}$$

which contains all previously loaded and stable items $b_0, \ldots, b_{j-1}$ plus the new focal item b_j. A_0 represents the empty arrangement before any items are loaded, and A_J equals the fully loaded

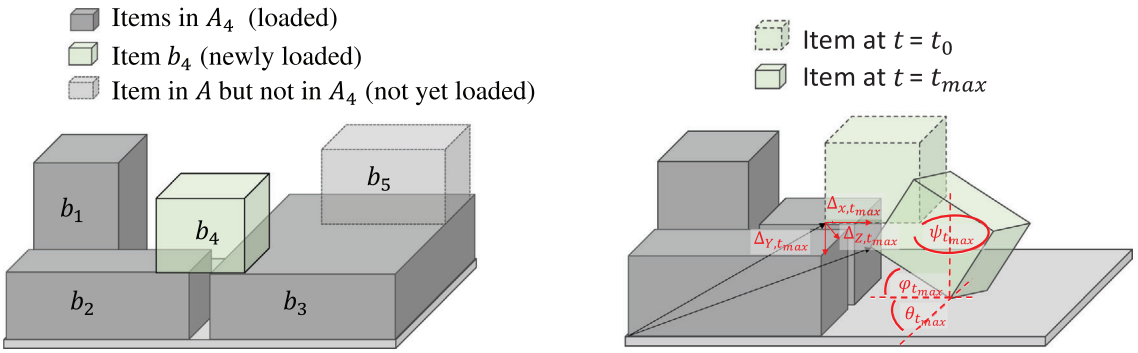

Fig. 3. Left: Simplified loading situation with $j = 4$, $A = (b_1, b_2, b_3, b_4, b_5)$, and $A_4 = (b_1, b_2, b_3, b_4)$. The loading and stability of the new item b_4 might change and depend on the stability of any previously loaded item $b_1, \ldots, b_3$ Right: Loading stability metric illustration. The loading of the green unstable item leads to its translational displacement $\Delta_{x,j,k,t}$, $\Delta_{y,k,t}$, $\Delta_{z,k,t}$, and rotational displacement $\varphi_{j,k,t}$, $\theta_{j,k,t}$, $\psi_{j,k,t}$ after a number of timesteps t_{max}.

complete layout A. The consecutive loading is defined as

$$A_j = A_{j-1} \cup \{b_j\}, \quad \text{for } j \geq 1. \tag{2}$$

The loading is finished if $j = J$. This consecutive loading of items reflects real-world pallet loading. When the new item b_j is loaded, the stability of the entire loaded subsequence A_j needs to be evaluated. This is necessary since loading one additional item b_j might change the stability of any previously loaded and stable item (Ramos et al., 2016b). Any newly placed item might change the force distribution of all existing items and might affect the loading stability of the entire layout (Le Jean et al., 2024). The objective is to determine the highest index j where a given loading sequence remains stable. We exemplify the loading situation in Fig. 3.

Given the dependence of stability on the consecutive loading, we now formulate our loading stability metric. The loading stability $ls(A)$ of an entire item layout is now defined as

$$ls(A) = \frac{j_{max}}{J}. \tag{3}$$

The loading stability is a normalized value between 0 and 1. It is defined as the ratio between the highest stable loading sequence j_{max} and the total number of items J in the layout. The loading stability indicates the number of safely loadable items and represents an item layout's amount of stability. The loading stability depends on the highest stable item-loading sequence j_{max}, which is defined as

$$j_{max} = \max(j | ls(A_j) = 1)). \tag{4}$$

Thereby, j_{max} equals the number of items that can be safely loaded with static stability. The layout is safely loadable up to the sequence j_{max}. When $j_{max} = J$, the entire layout is stable. Since j_{max}

depends on the loading stability of every loading subsequence A_j, we define it as

$$ls(A_j) = \begin{cases} 1 & \text{if } \prod_{k=0}^{j} ls(b_k) = 1 \\ 0 & \text{else} \end{cases}. \tag{5}$$

Thereby, the subsequence A_j is stable only if all items b_k in the subsequence are stable. Finally, the loading stability of a single item is defined as

$$ls(b_k) = \begin{cases} 1 & \text{item } b_k \text{ is stable} \\ 0 & \text{else} \end{cases}. \tag{6}$$

In this definition, each item in the subsequence A_j is stable ($ls(b_k) = 1$) or unstable ($ls(b_k) = 0$). Note that we neglect loading sequences with (hypothetical) subsequent items $j_{max} + 1, \ldots, J$, since the buildup of this arrangement would stop at this stage and subsequent loading of items would not be feasible in reality.

Since our definition of loading stability focuses not only on the consecutive nature of cargo loading but also on the dynamics of the cargo loading process, it includes considerable item positions and rotation changes. To account for the loading dynamics, for every item b_k in every subsequence A_j, there is a set of time steps t, $t = t_0, \ldots, t_{max}$ that models a given time frame, during which the items move, collide, and rotate. These physical interactions lead to changes in spatial positions $p_{j,k,t}$ and orientations quaternion $q_{j,k,t}$, which is given by the following: $p_{j,k,t} = (x_{j,k,t}, y_{j,k,t}, z_{j,k,t})'$ and $q_{j,k,t} = q_{j,k,t,w}, q_{j,k,t,x}, q_{j,k,t,y}, q_{j,k,t,z}$ To account for the translation and rotation part of our loading stability definition, we introduce the amount of item movement Δ on the X, Y, and Z axes and the item rotation angles φ, θ, and ψ. For every time frame t, we calculate the distance of item movement $\Delta_{x,j,k,t} = x_{j,k,t} - x_{j,k,t_0}$, $\Delta_{y,j,k,t} = y_{j,k,t} - y_{j,k,t_0}$, and $\Delta_{z,j,k,t} = z_{j,k,t} - z_{j,k,t_0}$ and item rotation angles $\varphi_{j,k,t}$, $\theta_{j,k,t}$, and $\psi_{j,k,t}$, which represent Euler roll, pitch, and yaw angles around the X-, Y-, and Z-axes and can be efficiently calculated using the formula presented by Bernardes and Viollet (2022). We assume that the rotation angles at t_0 to be $\varphi_{j,k,t_0} = \theta_{j,k,t_0} = \psi_{j,k,t_0} = 0$. Since we assume rigid bodies, deformation is not possible. Finally, the item b_k in subsequence A_j at loading step j is unstable according to our loading stability approach ls if for any k, t, the observed spatial displacement exceeds a translational threshold ϵ_t (e.g., 10 cm) or the observed rotation angles exceed a rotational threshold ϵ_R (e.g., 10°):

$$ls(b_k) = \begin{cases} 1, & \text{if } (|\Delta_{x,j,k,t}| > \epsilon_t) \vee (|\Delta_{y,j,k,t}| > \epsilon_t) \vee (|\Delta_{z,j,k,t}| > \epsilon_t) \\ & \vee (|\varphi_{j,k,t}| > \epsilon_R) \vee (|\theta_{j,k,t}| > \epsilon_R) \vee (|\psi_{j,k,t}| > \epsilon_R) \\ 0, & \text{else} \end{cases}. \tag{7}$$

We assume these tolerance values ϵ_T and ϵ_R will depend on the practical scenario at hand. We illustrate our novel loading stability metric in Fig. 3. Having clarified our loading stability metric, we now benchmark different static stability approaches *ss* against this loading stability.

Table 3
High-level approach classification

	Time		Focus		Scope	
	Static	Dynamic	Base related	Force related	Single item	Layout
FBS	x		x		x	
PBS	x		x		x	
SME	x			x		x
PS		x		x		x

5. Conceptual comparison

5.1. Approach selection and high-level classification

For our comparison of current static stability approaches *ss*, we distinguish between FBS (e.g., Moura and Bortfeldt, 2017), PBS with minimal supported area (e.g., Mack et al., 2004), PBS with minimal number of supported corners (e.g., Paquay et al., 2016), SME with support polygon (Ramos et al., 2016a), hybrid SME and support ratio (Krebs and Ehmke, 2021), and PS (Mazur et al., 2020)

For our high-level classification, we first distinguish between static and dynamic approaches. Static approaches are independent of a time component, capturing a loading situation at a specific point in time. Cargo items appear immediately in their final loading position, which excludes movements and collisions. FBS, PBS, and SME are static approaches. Dynamic approaches include a time component that captures an item's spatial position changes in a given time interval. Dynamic approaches use a spatial displacement score to measure instability (Ramos et al., 2015). PS is a dynamic approach that enables the modeling of loading situations in which items are moved a distance to their final positions and could potentially collide with already-placed items or the pallet. Dynamic approaches can be seen as simulated consequences of static approaches.

We observe a conceptual difference between item information usage when distinguishing between base-related and force-related approaches. In a base-related approach, an item's stability is evaluated based on the geometry of its lower face. Base-related approaches perform collision tests between an item's bottom face or its vertices and already-placed items' top faces. FBS and PBS are base-related approaches. Force-related approaches can directly model and approximate the forces acting on items and indirectly use information about an item's geometry. SME and PS are force-related approaches.

We also distinguish between approaches that natively use information about the entire layout and approaches that operate at the single-item level. The base-related approaches determine the supported base of a single item, whereas both force-related approaches (SME and PS) include information about items placed later in the sequence through collision detection and normal force calculations. Table 3 shows our high-level approach classification.

5.2. Systematic errors

To explore approach errors systematically, we first introduce an approach accuracy classification scheme. Note that we assume that layouts are not generated using the using the approach *ss* as a

Table 4
Systematic error classification model

	Correct evaluation	Underestimation	Overestimation
Condition	$ss = ls$	$ss < ls$	$ss > ls$
Consequence	Correct modeling	Overly restrictive; rejects stable layouts as unstable	Potentially dangerous; accepts unstable layouts as stable

Table 5
Systematic error classification

	Approach			
Influence	FBS	PBS	SME	PS
Top influence	UE	UE + OE	UE + OE	UE + OE
Bottom influence	UE	UE	−	UE + OE
Side influence	UE	UE	UE	UE + OE
Weight distribution	UE	UE + OE	UE + OE	UE + OE

constraint during layout generation. We assume *ss* works independently from layout generation. We distinguish between three combinations of approach outcome *ss* and our loading stability benchmark $ls(\epsilon_T, \epsilon_T)$: correct evaluation, underestimation (UE), and overestimation (OE). *ss* refers to the *ss*'s prediction of a layout's static stability. *ls* represents the loading stability benchmark. It can be empirically observed through item loading and evaluating if a cargo item significantly moves or falls down. Its definition depends on the thresholds ϵ_T and ϵ_R, which depend on the practical use case—for example, an item might be permitted to move by a few millimeters or centimeters but not more. *ls* may also depend on the item's material. UE results in a stable layout sequence being evaluated as unstable. This error causes overly restrictive behavior (Ramos et al., 2016a) and is undesirable because it rejects a potentially sound layout. In the long run, applying this approach represents an unnecessary restriction of the solution space and leads to lower overall solution quality (e.g., in terms of volume utilization) since it inhibits the search for good solutions. OE occurs when an approach incorrectly evaluates an unstable layout as stable. This false assumption of safety has negative consequences; for example, it could lead to falling cargo items that could cause injuries and/or damage or break those items (Bortfeldt and Wäscher, 2013; Zhao et al., 2016). We illustrate this classification scheme in Table 4.

Having provided the accuracy classification scheme, we now present problem characteristics that influence an item's static stability. We first distinguish between an item's surroundings that influence it from the (1) top, (2) bottom, and (3) side. We then include (4) weight distribution characteristics. We provide an overview of systematic errors in Table 5.

Top influence. The top influence represents loading situations in which an item's stability changes because of other items placed above. Weight forces from above can shift an item's resultant force point. FBS and PBS ignore top influences, focusing exclusively on the bottom. If FBS continues to hold for a layout evaluated as stable, placements from above cannot make the layout unstable according to FBS. Thus, FBS's evaluation of a layout as unstable could be incorrect, as is the case when an unstable item (according to FBS) is supported from above (e.g., by a heavy item), leading

to UE. PBS similarly ignores top influences. An item placement evaluated by PBS as stable based on its α level could become unstable if another item is placed on top of it and shifts the center of mass such that it is no longer supported (OE). Further, when a layout has been evaluated as unstable by PBS, loaded items could support other items from the top such that they become stable (UE). Similarly, SME does not consider top forces. Although Ramos et al. (2016a) mentioned top force calculation, they did not publish an algorithm. Future research might calculate and include top forces in the SME model, but there are two limitations: there is only a limited number of forces that can be included in the model, and SME assumes that forces act on a single point. Therefore, SME is prone to OE and UE when action forces are either omitted or not correctly calculated.

Bottom influence. The bottom influence is the major driver for support effects that cause reaction forces that stabilize items. FBS captures the bottom influence through the required supported bottom area. If we assume that only the gravitational force acts and there are only rigid bodies, an item is statically stable if it is fully supported from below. Hence, FBS already considers the intersection of the bottom area of an item with all items below, and no further potential for improvement exists. FBS underestimates a layout if a stable item's bottom area is only partly supported (UE). For PBS, the α level determines its accuracy. If set correctly, an α value that is large enough (e.g., $\alpha > 40\%$) is expected to be stable when considering only bottom influences and neglecting displaced centers of mass. If set wrongly (e.g., smaller α), unstable layouts might be evaluated as stable (OE). Further, a high α level, even as large as 50%, does not guarantee stability since the resulting force point might still not be supported (Zhu et al., 2024). In turn, when a stable item is only marginally supported from below (e.g., if all corners are supported), PBS evaluates a stable layout as unstable (UE) (Zhu et al., 2024). For SME, we observe that forces from below (i.e., reaction forces) are not considered. They are approximated through the assumption of a supported resulting force point. The resulting force point must be supported either by the floor, an underlying item, or the support polygon (De Castro Silva et al., 2003; Ramos et al., 2016a), which leads to correct evaluations.

Side influence. Lateral support is present either through other items or rigid (container) walls. Since we focus on pallets, we assume only lateral support from other items exists. The side influence plays a major role in dynamic stability (Ramos et al., 2015). In the case of static stability, side support means items that have been placed in an unstable way are nevertheless supported from the side(s) such that they remain stable. FBS does not consider side influences since it generates only stable layouts. However, it would evaluate potentially stable layouts as unstable (UE), as would PBS (UE). With respect to SME, algorithms have thus far considered only top-down forces since a layout does not move and thus no external side forces exist. However, tilting items do exceed side forces on other items without any movement. Therefore, side forces could result in UE.

Weight distribution. For FBS, a DCOM has no impact since fully supported items are also supported at their center of mass. However, layouts that would be stable if a DCOM existed are rejected (UE). PBS considers only area, and a DCOM might be located on parts of the bottom area that are not supported (OE). Further, PBS might underestimate a layout's static stability since a DCOM might stabilize an item such that less base support area is necessary for loading (UE). With respect to SME, DCOMs can be considered when calculating the resulting force points.

Systematic errors of the PS approach. All present sources of systematic errors can be omitted by using the PS approach. However, to a great extent, its worth depends on the parameter configuration of the simulation timespan, its resolution, and the displacement threshold ϵ_T. If ϵ_T is set incorrectly, collisions between bodies might be missed, which would in turn lead to false stability

Table 6
Overview of datasets

Dataset No.	Number of layouts	Mean (SD) items per layout	Generation stability	Generation algorithm
1	1276	15.88 (5.46)	Base support	SeqACLPP-WAC
2	3923	3.30 (1.36)	—	160 different algorithms

predictions (OE). If the ϵ_T value is not set well, PS will be overly conservative and result in UE. Therefore, PS is prone to all types of errors depending on the parameter configuration.

6. Benchmarking

6.1. Data generation and preparation

To strengthen the generalizability of our results, we benchmark on two highly heterogeneous datasets: (1) air cargo load planning problem instances published by Brandt and Nickel (2019) and (2) randomly assigned instances based on Ali et al. (2024). Both datasets are different in terms of items, number of items per layout, whether or not stability was enforced during the generation of layouts, and how the layouts were generated. Table 6 provides an overview of the benchmark datasets. The generation algorithm refers to the generation heuristics originally used to create the layouts. Generation stability refers to whether or not static stability was enforced during layout generation.

The first dataset is initially based on real bookings of more than 1,300 generated layouts from 400 flights in 2014, is the only published air cargo datasets, and is thus well suited as the basis for benchmarking. This dataset is the most realistic, but it also introduces a bias toward base support. In the final dataset, there are 41 layouts with an A contour (cuboidal with $\{(0, 0, Z), (244, 0, Z), (244, 244, Z), (0, 244, Z)\}$ with $Z \in \{0, 606\}$), 643 with a D contour (pentagonal with $\{(0, 0, Z), (0, 300, Z), (166, 300, Z), (244, 244, Z), (244, 0, Z)\}$, with $Z \in \{0, 318\}$), 21 with an AKE-ULD container contour (pentagonal, we assume no walls, with {(52, 0, 0), (0,45,0), (0,158,0), (196,158,0), (196, 0, 0)} with $Z \in \{0, 141\}$), and 51 with an F contour (hexagonal with $\{(0, 50, Z), (0, 163, Z), (406, 163, Z), (406, 50, Z), (362, 0, Z), (44, 0, Z)\}$, with $Z \in \{0, 244\}$).

The second dataset comprises three subsets, each of which consists of 60 items. The item dimensions of those sets are randomly generated with a uniform distribution of intervals [1, 10], [1, 35], and [1, 100], respectively. The size of the large objects is 30 * 30 * 30 for the first subset and 100 * 100 * 100 for the second and third subsets. The instances are solved using 160 packing heuristics developed by combining four bin selection strategies, eight space selection strategies, and five placement rules (Ali et al., 2024). The dataset contains item dimensions, placement coordinates, and loading sequences, and weights are proportional to item volumes. This dataset is rather unrealistic with respect to the item characteristics; however, it does not include a static stability generation bias. Since no mass distribution information was present in the datasets, we assign values according to three scenarios of varying complexity:

- The first scenario (S_1) represents a simplified situation and matches the original dataset. All items have uniform mass distribution, which means the geometric center coincides with the center of mass.

- The second scenario (S_{2a}) includes DCOMs. We assign each item a center of mass in the items' dimensional bounds according to a Gaussian distribution around the items' geometric centers, with μ at the geometric center and $\sigma = (dimension * 0.8)/6$, so approximately 99.7% of the center of mass positions are within 80% of the items' bounds. If this value is exceeded, we clip the center of mass positions at 80%. Thus, the majority of items have centers of mass close to their geometric centers.
- For the last scenario (S_{2b}), we use the same procedure as in S_{2a} but change from Gaussian distribution to uniform distribution.

We selected only layouts with at least two items. Due to extensive simulation runtime for layouts with large amounts of items, we capped the item amount at 20. If a layout contained more than 20 items, we selected only the first 20 items of the respective layout. Since the simulation became unstable for very small, thin, or flat items, we selected only layouts in which all items have minimal dimensions (width, height, depth) of at least 10 cm. We rounded item weights to the nearest multiple of ten. We ended up with the number of layouts depicted in Table 6.

6.2. Benchmark simulation

Since there is a general lack of realistic datasets for pallet loading (Daios et al., 2024) and no pallet loading dataset that incorporates information about item (in)stability, we decided to simulate loading operations using an independent simulation with the multibody simulator MSC ADAMS[2] to obtain the benchmark loading stability labels ls. ADAMS is market-leading software for multibody dynamics and motion analysis for physics simulation. Therefore, ADAMS is a better fit for this task compared to other simulation software that has been used in related literature, such as Autodesk Inventor, which is a three-dimensional computer-aided design program with add-ons for standard simulation tasks. We model each item as a rigid body and use a RAPID geometry engine and HHT integrator (error rate $= 1.0 \times 10^{-7}$), which is well suited for a large number of bodies and collisions in a simulation. With respect to forces, we include the gravitational force and contact forces between each pair of items. Contact forces are discontinuous forces that act between two rigid bodies. ADAMS uses its own contact force based on the IMPACT contact force model, which is defined as follows: $N = kg^e + STEP(g, 0, 0, d_{max}, c_{max})\frac{dg}{dt}$. The IMPACT model is based on Hertzian contact theory and includes penetration mechanics—that is, a controlled overlapping g of rigid bodies up to a maximum penetration distance d_{max}. The normal force N is based on two components: a stiffness component and a step function. The stiffness component is driven by k, the stiffness factor; g, the penetration distance; and e, the exponent. The stiffness factor depends not only on the material property of the item packaging but also on the geometry of the item. From the penetration distance on, the maximum damping c_{max} acts and adds up to the normal force. Determining good estimates for these parameters is difficult, so they should be obtained using an iterative process. We base our initial contact parameter values on ADAMS recommendations and iteratively adjust them until we observe realistic force development relationships. To further confirm our results, we check the

[2]https://hexagon.com/de/products/product-groups/computer-aided-engineering-software/adams

Table 7
Base simulation specification

Parameter type	Parameter	Value	Unit
Integrator settings	Error rate	$1.0e-7$	cm
Contact	k	$6e7$	N/cm
forces	e	1.2	–
	d_{max}	0.001	cm
	c_{max}	800	N × s/cm
Coulomb	μ_s	0.5	–
friction	μ_d	0.375	–
	V_s	0.01	cm/s
	V_d	1	cm/s
Pallet	E	$7,31e7$	N/cm^2
material	v	0.35	–
	ρ	$2.79e-3$	kg/cm^3
Item	E	$2.8e5$	N/cm^2
material	v	0.29	–
	G_{xz}	$1.4e4$	N/cm^2
	G_{xy}	$3.7e4$	N/cm^2
Simulation	ϵ_T	10	cm
settings	ϵ_R	10	degree
	t_{max}	0.3	s
	F_G	980	cm/s^2

normal forces for contacts and compare stiffness and damping values. Stiffness should contribute most to the contact force at its peak. We end up with the values shown in Table 7.

Together with the contact forces, we include the Coulomb friction model. It includes the coefficient of static friction (μ_s), which acts up to the threshold velocity; the stiction transition velocity (V_s); and the coefficient of dynamic friction (μ_d), which acts from a threshold friction transition velocity on (V_d). We assign a coefficient of static friction value of $\mu_s = 0.5$, matching empirical dry cardboard–cardboard and dry cardboard–wooden pallet friction estimates. For dynamic friction, we assign a value of $0.75 * \mu_s = 0.375$ (IMO/ILO/UNECE, 2014). For the stiction and friction transition velocities, we use recommendations from ADAMS.

Further, a material configuration exists for each rigid body. The material configuration is based on the following parameters: Young's modulus (E), shear modulus (G), Poisson's ratio (v), and density (ρ). Young's modulus, as defined in Hooke's law, is a mechanical property that influences the relative strain of a solid body and defines a material's stiffness (Hibbeler, 2010). For the pallet material, we use common values for aluminum for Young's modulus, Poisson's ratio, and density (Hibbeler, 2010). In air cargo, items are often packed in cardboard and wrapped in plastic foil (Brandt and Nickel, 2019). Therefore, we assign the elasticity modules E and G to match values for corrugated cardboard. Since cardboard has different mechanical properties depending on its direction, we assume orthotropicity and distinguish between machine direction (MD), cross-machine direction (CD), and thickness direction (ZD). For our case, we assume $ZD = G_{xz}$, $CD = G_{xy}$, and $MD = E$, so the weakest direction of the cardboard is along the cardboard thickness direction. Further, we set the parameter values to 300 g/m^2 paperboard under standard conditions (Fadiji et al., 2017). The density ρ of each item is determined by the input data.

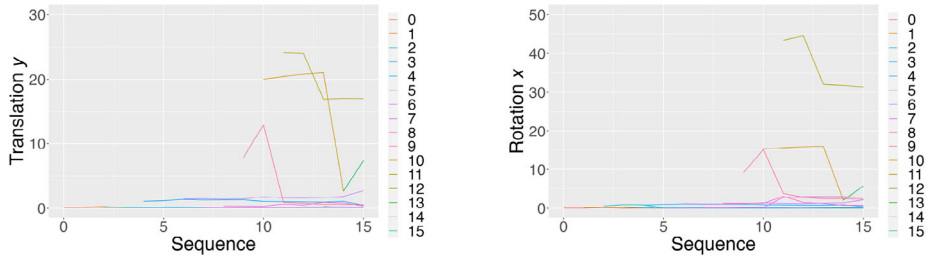

Fig. 4. Demonstration of the maximal y-translations and x-rotations of all items in the demonstration layout depending on the item sequence.

Since there is no common threshold value in the literature, we set the loading stability thresholds to $\epsilon_T = 10$ and $\epsilon_R = 10$, respectively. This means an item is unstable if it moves 10 cm in the X, Y, or Z direction or rotates 10° in either direction. In practice, these values might change depending on item dimensions, time spans, and materials. Since our benchmark simulation results heavily depend on the choice of ϵ_T and ϵ_R, we include a sensitivity analysis for these values. We simulated 0.3s per loading step, in which a newly loaded item is placed in the layout. The earth gravity's is set to 980 cm/s^2 and points downward in $-y$-direction.

6.3. Benchmark results

We benchmark the approaches on a workstation with an AMD Ryzen ThreadRipper 3960X 3.8 GHz CPU and 256 RAM. Figure 4 presents the results from our computational experiments, and Table 8 details the values. We present the percentage of correct predictions ($ss = ls$) (which we label as accuracy), UE ($ss < ls$), and OE ($ss > ls$) to the total number of layouts N in the datasets. Runtime refers to the average computing time in milliseconds to assess a single layout given the approach ss. PBS $_\alpha$ and PS $_{\epsilon_T}$ represent the PBS and PS approaches with a fixed α or ϵ_T value. We further include FBS, PBS $_3$ (Paquay et al., 2016), SME (Ramos et al., 2016a), and SME $_K$ (Krebs and Ehmke, 2021). We first exemplify the loading stability metrics using a demonstration layout. Figure 5 illustrates the maximal translation (in centimeters) on the y-axis and the maximal rotation (in degrees) on the x-axis for each loading sequence. From the data, we can observe that items continuously move and rotate at a low level. However, we note that specific item sequences—for example, from $j = 9$ onward—produce significant translations and rotations. Items that significantly move or rotate during loading may indicate instability and might not be safely loaded.

Our benchmark reveals that for both datasets, PS achieves the highest accuracy, while FBS performs the worst. SME achieves close accuracy compared to PS, while PBS strongly depends on the α value. As can be seen in Table 8, the closer α tends toward 1, the closer PBS comes to FBS and the more restrictive it gets, which is associated with more UE. The optimal α-value varies across both datasets. For Dataset 1, PBS performs considerably better, which may be due to the dataset's bias toward base support. In case of minimal base support during layout generation, PBS computation for α values below the generation stability α value is of limited informative value, since PBS with smaller α-values will result in positive stability assessments. PBS is an assessment of the layout generation, not of the stability. Therefore, PBS is strongly related to the base support distribution in the datasets. For Dataset 1, one can observe a substantial drop in accuracy and an increase in

Table 8
Benchmark simulation results for the approaches on Datasets 1 (left, $N = 1276$) and 2 (right, $N = 3923$) for scenarios S_1, S_{2a}, and S_{2b}. The values correspond to the percentage of correct predictions (labeled as accuracy), UE, and OE to the total number of layouts N in the respective dataset. Runtime refers to the average runtime per approach to evaluate the entire layout

Dataset- Scenario	Approach	Accuracy	UE	OE	Runtime
1-S $_1$	FBS	34.17	65.83	0.00	0.184
	PBS $_{(0.90)}$	44.44	55.56	0.00	0.234
	PBS $_{(0.80)}$	67.87	32.13	0.00	0.292
	PBS $_{(0.70)}$	94.59	5.41	0.00	0.172
	PBS $_{(0.60)}$	94.91	5.09	0.00	0.281
	PBS $_{(0.50)}$	95.61	4.39	0.00	0.219
	PBS $_3$	36.29	63.71	0.00	1.084
	SME	93.10	6.90	0.00	1.016
	SME $_K$	92.71	7.29	0.00	0.324
	PS $_{(1.0)}$	96.32	3.68	0.00	306.565
	PS $_{(2.0)}$	98.12	1.88	0.00	308.015
	PS $_{(3.0)}$	99.06	0.94	0.00	307.386
	PS $_{(4.0)}$	99.84	0.16	0.00	309.202
	PS $_{(5.0)}$	99.92	0.08	0.00	314.751
1-S $_{2a}$	FBS	34.40	65.60	0.00	0.064
	PBS $_{(0.90)}$	44.83	55.09	0.08	0.054
	PBS $_{(0.80)}$	67.63	31.82	0.55	0.065
	PBS $_{(0.70)}$	93.10	5.33	1.57	0.074
	PBS $_{(0.60)}$	93.42	5.02	1.57	0.056
	PBS $_{(0.50)}$	94.04	4.39	1.57	0.067
	PBS $_3$	36.60	63.40	0.00	0.853
	SME	90.91	8.70	0.39	0.727
	SME $_K$	91.30	7.21	1.49	0.079
	PS $_{(1.0)}$	94.91	4.23	0.86	289.176
	PS $_{(2.0)}$	96.87	2.27	0.86	294.003
	PS $_{(3.0)}$	97.65	1.49	0.86	298.117
	PS $_{(4.0)}$	98.28	0.71	1.02	299.695
	PS $_{(5.0)}$	98.35	0.47	1.18	296.654
1-S $_{2b}$	FBS	36.36	63.64	0.00	0.042
	PBS $_{(0.90)}$	46.71	53.06	0.24	0.052
	PBS $_{(0.80)}$	69.20	29.00	1.80	0.06
	PBS $_{(0.70)}$	90.13	4.55	5.33	0.054
	PBS $_{(0.60)}$	90.20	4.39	5.41	0.074
	PBS $_{(0.50)}$	90.91	3.68	5.41	0.061
	PBS $_3$	38.24	61.60	0.16	0.795
	SME	86.83	12.30	0.86	0.579
	SME $_K$	88.87	6.03	5.09	0.08
	PS $_{(1.0)}$	91.54	6.90	1.57	277.567
	PS $_{(2.0)}$	94.28	4.00	1.72	283.852
	PS $_{(3.0)}$	95.06	2.98	1.96	287.56
	PS $_{(4.0)}$	95.30	2.12	2.59	289.148
	PS $_{(5.0)}$	95.14	1.65	3.21	287.177

Continued

Table 8
(Continued)

Dataset- Scenario	Approach	Accuracy	UE	OE	Runtime
2-S $_1$	FBS	65.64	34.36	0.00	0.022
	PBS $_{(0.90)}$	71.71	28.29	0.00	0.019
	PBS $_{(0.80)}$	76.32	23.68	0.00	0.023
	PBS $_{(0.70)}$	83.81	16.19	0.00	0.022
	PBS $_{(0.60)}$	90.93	9.07	0.00	0.021
	PBS $_{(0.50)}$	94.26	5.56	0.18	0.017
	PBS $_3$	65.64	34.36	0.00	0.375
	SME	96.25	3.42	0.33	0.044
	SME $_K$	80.22	19.78	0.00	0.024
	PS $_{(1.0)}$	97.32	1.73	0.94	13.391
	PS $_{(2.0)}$	97.76	1.30	0.94	13.398
	PS $_{(3.0)}$	98.01	0.94	1.05	13.404
	PS $_{(4.0)}$	98.37	0.38	1.25	13.471
	PS $_{(5.0)}$	97.94	0.18	1.89	13.409
2-S $_{2a}$	FBS	67.14	32.86	0.00	0.023
	PBS $_{(0.90)}$	73.39	26.56	0.05	0.019
	PBS $_{(0.80)}$	77.98	21.92	0.10	0.022
	PBS $_{(0.70)}$	84.63	14.84	0.54	0.02
	PBS $_{(0.60)}$	90.24	8.49	1.27	0.019
	PBS $_{(0.50)}$	92.00	5.79	2.22	0.019
	PBS $_3$	67.14	32.86	0.00	0.373
	SME	96.23	3.24	0.54	0.051
	SME $_K$	81.54	18.17	0.28	0.018
	PS $_{(1.0)}$	96.79	1.91	1.30	13.297
	PS $_{(2.0)}$	97.35	1.35	1.30	13.264
	PS $_{(3.0)}$	97.43	1.10	1.48	13.274
	PS $_{(4.0)}$	97.40	0.79	1.81	13.306
	PS $_{(5.0)}$	97.32	0.41	2.27	13.286
2-S $_{2b}$	FBS	69.08	30.92	0.00	0.018
	PBS $_{(0.90)}$	75.48	24.52	0.00	0.019
	PBS $_{(0.80)}$	79.23	20.24	0.54	0.028
	PBS $_{(0.70)}$	83.89	13.97	2.14	0.019
	PBS $_{(0.60)}$	87.41	8.67	3.93	0.02
	PBS $_{(0.50)}$	88.02	6.55	5.43	0.021
	PBS $_3$	69.08	30.92	0.00	0.375
	SME	96.02	3.26	0.71	0.07
	SME $_K$	82.11	16.72	1.17	0.022
	PS $_{(1.0)}$	96.23	1.86	1.91	13.103
	PS $_{(2.0)}$	96.64	1.33	2.04	13.041
	PS $_{(3.0)}$	96.79	1.15	2.06	13.149
	PS $_{(4.0)}$	96.66	0.84	2.50	13.106
	PS $_{(5.0)}$	96.43	0.61	2.96	13.171

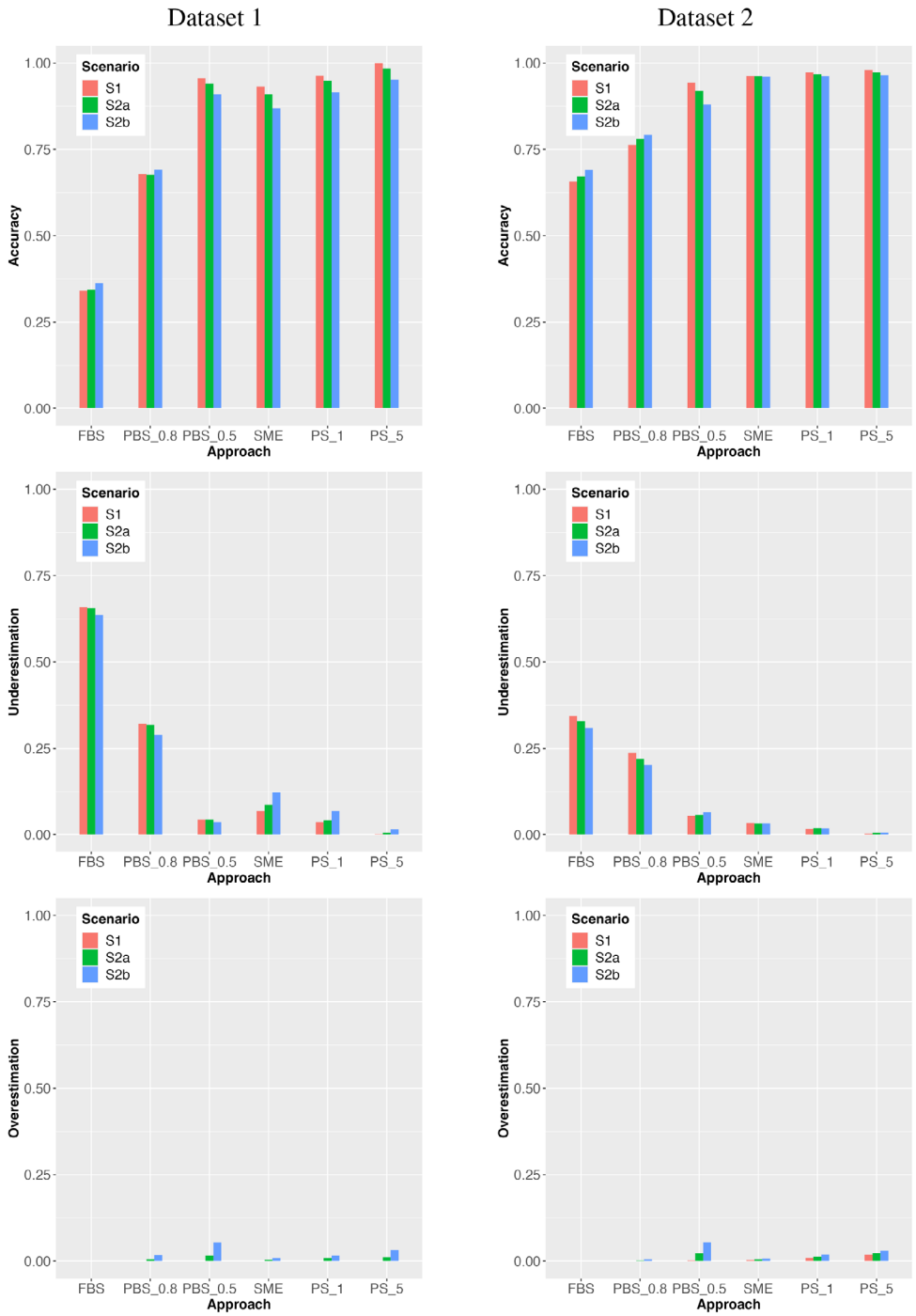

Fig. 5. Benchmark comparison for the FBS, PBS $_{0.8}$, PBS $_{0.5}$, SME, PS $_1$, and PS $_5$ approaches on Datasets 1 ($N = 1276$) and 2 ($N = 3923$): accuracy (top), UE (middle), and OE (bottom). The values correspond to the percentage of correct predictions, UE, and OE to the total number of layouts N in the dataset based on results presented in Table 8.

UE above $\alpha > 0.7$. SME and SME $_K$ both perform well with Dataset 1, while SME $_K$ becomes considerably more imprecise with Dataset 2. This finding could also be attributable to Dataset 1's inherent bias toward stability. Further, FBS is prone to UE, which further reflects and underlines its restrictiveness. For Dataset 2, a lower α-value is clearly attached to a higher accuracy. In Dataset 2, even an α-value of 0.9 achieves an accuracy of about 65–70% compared to Dataset 1, in which PBS $_{0.9}$ achieves an accuracy of about 35%. We find no clear indication that more complex scenarios result in lower accuracy. Given the trade-off between runtime and accuracy, SME and PBS with a low α-value are a good choice: both consume little runtime, considerably less than PS. Achieving higher accuracy is associated with greater runtime. PS achieves the highest accuracy across all datasets and scenarios but also consumes the greatest runtime. Depending on its ϵ_T value, it is prone to both UE and OE. Adjusting PS parameters to find an optimal trade-off depending on the consequences associated with each systematic error can be a goal for future research. In general, we conclude that PS is best equipped to capture practical complexity: it remains the most accurate approach to approximate loading stability for both the simpler and more complex scenarios. As indicated in our structural error analysis, weight distribution seems to lead to more OE in the case of PBS. Further, PS is prone to OE, especially in the S_{2b} scenario. To investigate the dependence of our results on the loading stability value ls and its thresholds ϵ_T and ϵ_R, we conduct a sensitivity analysis. We illustrate the impact of varying ϵ_T and ϵ_R values on the accuracy, UE, and OE of the approaches. We visually depict the results in Fig. 6. As can be observed, the results are sensitive to different threshold value levels. A similar pattern compared to the benchmark applies; for Dataset 1 and scenario S_1, the variation of the thresholds seems to have nearly no effect on the accuracy. A stronger variation is observable in Dataset 2, where the thresholds seem to have a considerable effect on accuracy. The smaller the thresholds, the more OE are observable, the larger, the more UE. This finding is as expected since tighter threshold bounds lead to more strict loading stability evaluations. To summarize, for the investigated threshold values, the results seem quite stable.

7. Discussion and conclusion

Our study makes a methodological contribution by showing how to benchmark and compare different static stability approaches. This comparison against a common independent loading stability value ensures that the differences between the approaches are adequately measured and do not depend on any optimization algorithm. Our results indicate that FBS is the most restrictive approach and that it underestimates loading stability in a systematic manner. This finding is in line with previous research (Ramos et al., 2016a). PBS depends highly on the base support parameter α and tends to be more accurate with lower α values. Specifically, the higher the α value, the more UE occurs and the closer PBS comes to the restrictive results of FBS. Therefore, it is useful to utilize PBS with smaller values, as other studies have also indicated (Junqueira and de Queiroz, 2022). SME is the most robust approach across the different datasets and scenarios and maintains a good accuracy-to-runtime trade-off. PS provides the greatest accuracy but consumes the most runtime by far.

From a managerial perspective, the choice of loading stability algorithm plays an important role in preventing damaged items, broken loading devices, and injured personnel, all of which are associated with increased costs, unsatisfied customers, and delays. Our study also helps clarify and

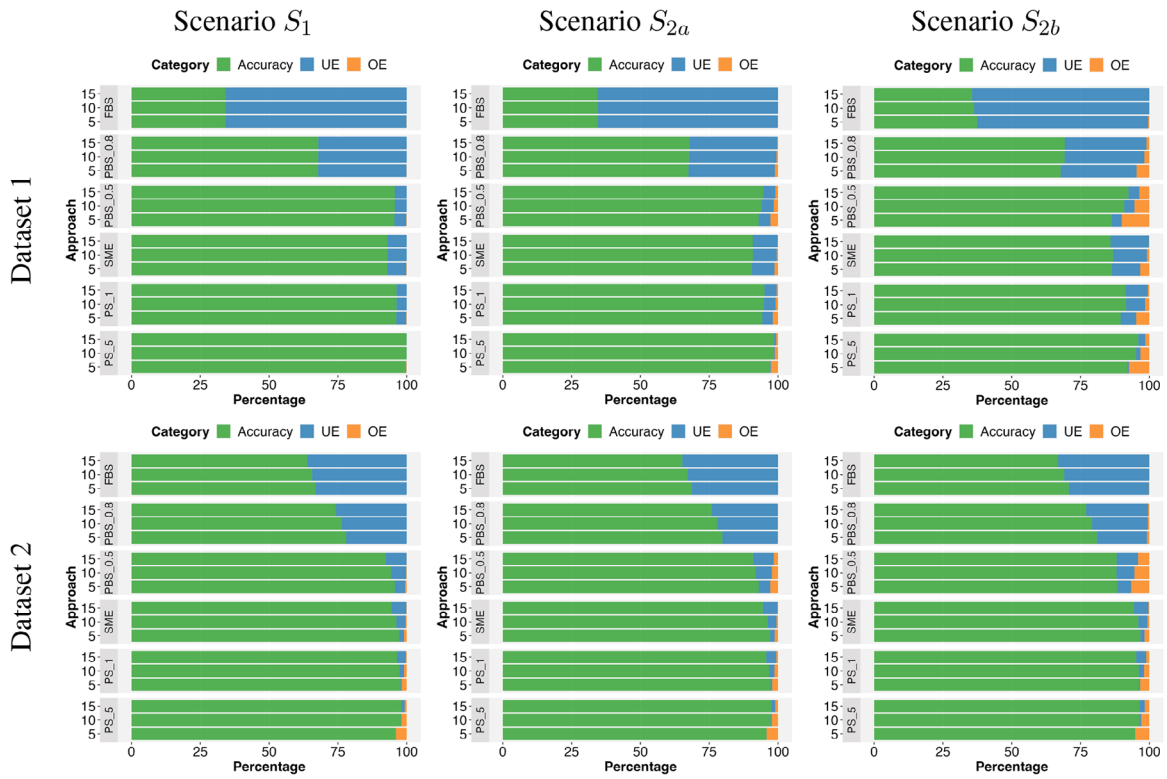

Fig. 6. Sensitivity analysis with varying levels of the loading stability value $ls(\epsilon_T, \epsilon_R)$ for $\epsilon_T = \epsilon_R = 5$, $\epsilon_T = \epsilon_R = 10$, and $\epsilon_T = \epsilon_R = 15$ for approaches FBS, PBS $_{0.8}$, PBS $_{0.5}$, SME, PS $_1$, and PS $_5$ for scenarios S_1 (left), S_{2a} (middle), and S_{2b} (right) on Dataset 1 (top) and Dataset 2 (bottom).

quantify the negative impact of a conservative, overly restrictive stability constraint. As we showed, the most restrictive approach (FBS) produces the most UE. This restrictiveness may prevent PLP systems from finding the best layouts as many layouts that achieve a high load factor are disregarded because they do not fulfill the constraints. Therefore, our results may also have a positive economic impact. Given the different scenarios and approaches, managers can estimate the impact of the individual static stability algorithms. Depending on their focus, they can then decide which algorithm to use based on the costs associated with UE and OE.

Our study also has limitations that need to be addressed in future research. First, our results are based on a conceptual discussion and a benchmark simulation. Thus, the validity of these results depends on the validity of the simulation model. Although we anchored our simulation configuration using the recommended settings from ADAMS and the literature, there still might be validity issues. Therefore, future research needs to confirm the simulation results in a field test. Since such a field test at a logistics company is associated with high effort and costs due to potential damages, an experimental setting might be sufficient. The obtained loading stability benchmark values ls depend on multiple factors, such as the simulation configuration, length, and the value for the thresholds ϵ_T and ϵ_R. Neither value indicates, if there was a damage to the cargo item or an item that has fallen to the ground. However, an item may be unstable even if it does not fall to the ground. The

correct specification of these thresholds, depending on the use case, industry, and item packaging material, is therefore an important next step. Third, we focused on pallets and thus assumed no lateral support for items from walls. Generalizing the results to the CLP is difficult as containers provide much more side support through the walls. Future research might evaluate the results and perform a benchmark study for the CLP in which rigid container walls are simulated. Fourth, we tested the approaches on two datasets—one from the problem domain of air cargo and one that was randomly generated—which limits the generalizability of our results to other industries and transportation modes. To explore stability in other transportation modes, different datasets from these industries must be taken into account. Fifth, the results for PBS for Dataset 1 are of limited informative value since they highly depend on the distribution of support in the dataset. Finally, we selected a set of static stability algorithms that represent the most important ones in our view and have been widely adopted in related studies. Nevertheless, there are more static stability approaches and other configurations, such as PBS with a smaller α level, which we did not test. Future research can include these approaches for a more holistic evaluation. We provided the benchmark instances and simulation results online,[3] so future research can build upon them.

Acknowledgments

This work is co-financed by Component 5 - Capitalization and Business Innovation, integrated into the Resilience Dimension of the Recovery and Resilience Plan within the scope of the Recovery and Resilience Mechanism (MRR) of the European Union (EU), framed in the Next Generation EU, for the period 2021 - 2026, within project Produtech R3, with reference 60.

Open access funding enabled and organized by Projekt DEAL.

References

Ali, S., Ramos, A.G., Carravilla, M.A., Oliveira, J.F., 2024. Heuristics for online three-dimensional packing problems and algorithm selection framework for semi-online with full look-ahead. *Applied Soft Computing* 151, 111168.

Araya, I., Moyano, M., Sanchez, C., 2020. A beam search algorithm for the biobjective container loading problem. *European Journal of Operational Research* 286, 2, 417–431.

Bernardes, E., Viollet, S., 2022. Quaternion to Euler angles conversion: a direct, general and computationally efficient method. *PLOS ONE* 17, 11, e0276302.

Bischoff, E.E., 1991. Stability aspects of pallet loading. *OR Spektrum* 13, 4, 189–197.

Bischoff, E.E., Ratcliff, M.S.W., 1995. Issues in the development of approaches to container loading. *Omega* 23, 4, 377–390.

Bortfeldt, A., Wäscher, G., 2013. Constraints in container loading – a state-of-the-art review. *European Journal of Operational Research* 229, 1, 1–20.

Bracht, E.C., de Queiroz, T.A., Schouery, R.C.S., Miyazawa, F.K., 2016. Dynamic cargo stability in loading and transportation of containers. In *2016 IEEE International Conference on Automation Science and Engineering (CASE)*, IEEE, Piscataway, NJ, pp. 227–232.

Brandt, F., Nickel, S., 2019. The air cargo load planning problem - a consolidated problem definition and literature review on related problems. *European Journal of Operational Research* 275, 2, 399–410.

[3]https://zenodo.org/records/13925610

Christensen, S.G., Rousøe, D.M., 2009. Container loading with multi-drop constraints. *International Transactions in Operational Research* 16, 6, 727–743. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1475-3995.2009.00714.x

Daios, A., Kladovasilakis, N., Kostavelis, I., 2024. Mixed palletizing for smart warehouse environments: sustainability review of existing methods. *Sustainability* 16, 3, 1278.

De Castro Silva, J.L., Soma, N.Y., Maculan, N., 2003. A greedy search for the three-dimensional bin packing problem: the packing static stability case. *International Transactions in Operational Research* 10, 2, 141–153.

de Queiroz, T.A., Bracht, E.C., Miyazawa, F.K., Bittencourt, M.L., 2019. An extension of Queiroz and Miyazawa's method for vertical stability in two-dimensional packing problems to deal with horizontal stability. *Engineering Optimization* 51, 6, 1049–1070.

de Queiroz, T.A., Miyazawa, F.K., 2014. Order and static stability into the strip packing problem. *Annals of Operations Research* 223, 1, 137–154.

Fadiji, T., Berry, T., Coetzee, C., Opara, L., 2017. Investigating the mechanical properties of paperboard packaging material for handling fresh produce under different environmental conditions: experimental analysis and finite element modelling. *Journal of Applied Packaging Research* 9, 2, 3.

Gehring, H., Bortfeldt, A., 1997. A genetic algorithm for solving the container loading problem. *International Transactions in Operational Research* 4, 5-6, 401–418.

Hibbeler, R.C., 2010. *Mechanics of Materials* (8. ed.). Prentice Hall, Upper Saddle River, NJ.

Hodgson, T.J., 1982. A combined approach to the pallet loading problem. *A I I E Transactions* 14, 3, 175–182.

IATA, 2019. *Cargo Handling Manual* (3rd ed.). International Air Transport Association, Montreal-Geneva.

IMO/ILO/UNECE,2014. IMO/ILO/UNECE Code of Practice for Packing of Cargo Transport Units (CTU-Code). *International Maritime Organization (IMO), the International Labour Organization (ILO) and the United Nations Economic Commission for Europe (UNECE)* 4, 7, 129–129.

Junqueira, L., Morabito, R., Sato Yamashita, D., 2012. Three-dimensional container loading models with cargo stability and load bearing constraints. *Computers and Operations Research* 39, 1, 74–85.

Junqueira, L., de Queiroz, T.A., 2022. The static stability of support factor–based rectangular packings: an assessment by regression analysis. *International Transactions in Operational Research* 27, 2, 311.

Krebs, C., Ehmke, J.F., 2021. Vertical stability constraints in combined vehicle routing and 3D container loading problems. In Mes, M., Lalla-Ruiz, E. and Voß, S. (eds), *Computational Logistics*, Springer International Publishing, Cham, pp. 442–455.

Le Jean, A., Brauner, N., Briant, O., Cocan, M., David, B., 2024. Stability constraints in a 3D knapsack problem with non parallelepipedic items. *Computers & Industrial Engineering* 189, 109943.

Lee, N.S., Mazur, P.G., Bittner, M., Schoder, D., 2021. An intelligent decision support system for air cargo palletizing. In *Proceedings of the 54th Hawaii International Conference on System Sciences*, January 5-8, 2021, Hawaii, USA.

Mack, D., Bortfeldt, A., Gehring, H., 2004. A parallel hybrid local search algorithm for the container loading problem. *International Transactions in Operational Research* 11, 5, 511–533.

Martinez-Franco, J.C., Alvarez-Martinez, D., 2018. Physx as a middleware for dynamic simulations in the container loading problem. In *2018 Winter Simulation Conference (WSC)*, IEEE, Piscataway, NJ, pp. 2933–2940.

Mazur, P.G., Lee, N.S., Schoder, D., 2020. Integration of physical simulations in static stability assessments for pallet loading in air cargo. In Bae, K.H., Feng, B., Kim, S., Lazarova-Molnar, S., Zheng, Z., Roeder, T. and Thiesing, R. (eds), *Proceedings of the 2020 Winter Simulation Conference, December 13-16, Orlando, Florida, USA*, IEEE, Piscataway, NJ, pp. 1312–1323.

Mazur, P.G., Lee, N.S., Schoder, D., 2022. A GPU-Accelerated Approach of Static Stability Assessments for Pallet Loading in Air Cargo. In *Proceedings of the 55th Hawaii International Conference on System Sciences (2022)*, pp. 1177–1185.

Moura, A., Bortfeldt, A., 2017. A two-stage packing problem procedure. *International Transactions in Operational Research* 24, 1-2, 43–58.

Oliveira, L.d.A., de Lima, V.L., de Queiroz, T.A., Miyazawa, F.K., 2021. The container loading problem with cargo stability: a study on support factors, mechanical equilibrium and grids. *Engineering Optimization* 53, 7, 1192–1211. https://doi.org/10.1080/0305215X.2020.1779250.

Olsson, J., Larsson, T., Quttineh, N.H., 2020. Automating the planning of container loading for Atlas Copco: coping with real-life stacking and stability constraints. *European Journal of Operational Research* 280, 3, 1018–1034.

Paquay, C., Schyns, M., Limbourg, S., 2016. A mixed integer programming formulation for the three-dimensional bin packing problem deriving from an air cargo application. *International Transactions in Operational Research* 23, 1-2, 187–213.

Rajaei, M., Moslehi, G., Reisi-Nafchi, M., 2024. The multiple container loading problem with loading docks. *International Transactions in Operational Research* 31, 3, 1671–1698. https://onlinelibrary.wiley.com/doi/pdf/10.1111/itor.13393

Ramos, A.G., Jacob, J., Justo, J.F., Oliveira, J.F., Rodrigues, R., Gomes, A.M., 2017. Cargo dynamic stability in the container loading problem - a physics simulation tool approach. *International Journal of Simulation and Process Modelling* 12, 1, 29–41.

Ramos, A.G., Oliveira, J.F., Gonçalves, J.F., Lopes, M.P., 2015. Dynamic stability metrics for the container loading problem. *Transportation Research Part C: Emerging Technologies* 60, 480–497.

Ramos, A.G., Oliveira, J.F., Gonçalves, J.F., Lopes, M.P., 2016a. A container loading algorithm with static mechanical equilibrium stability constraints. *Transportation Research Part B: Methodological* 91, 565–581.

Ramos, A.G., Oliveira, J.F., Lopes, M.P., 2016b. A physical packing sequence algorithm for the container loading problem with static mechanical equilibrium conditions. *International Transactions in Operational Research* 23, 1-2, 215–238.

Ranck Júnior, R., Yanasse, H.H., Morabito, R., Junqueira, L., 2019. A hybrid approach for a multi-compartment container loading problem. *Expert Systems with Applications* 137, 471–492.

Witkin, A., Baraff, D., 1997. Differential equation basics. *SIGGRAPH Course Notes 1997*, 24.

Yang, H., Hu, Y., Zhou, W., Gao, Y., Wang, J., 2024. A three-dimensional multiple containers loading approach considering weak heterogeneous containers for multiple customers. *International Transactions in Operational Research* Advance Online Publication. https://doi.org/10.1111/itor.13493

Zhao, X., Bennell, J.A., Bektaş, T., Dowsland, K., 2016. A comparative review of 3D container loading algorithms. *International Transactions in Operational Research* 23, 1-2, 287–320.

Zhu, W., Fu, Y., Zhou, Y., 2024. 3D dynamic heterogeneous robotic palletization problem. *European Journal of Operational Research* 316, 2, 584–596.