Railroad capacity tools and methodologies in the U.S. and Europe

Hamed Pouryousef • Pasi Lautala • Thomas White

Abstract A growing demand for passenger and freight transportation, combined with limited capital to expand the United States (U.S.) rail infrastructure, is creating pressure for a more efficient use of the current line capacity. This is further exacerbated by the fact that most passenger rail services operate on corridors that are shared with freight traffic. A capacity analysis is one alternative to address the situation and there are various approaches, tools, and methodologies available for application. As the U.S. continues to develop higher speed passenger services with similar characteristics to those in European shared-use lines, understanding the common methods and tools used on both continents grows in relevance. There has not as yet been a detailed investigation as to how each continent approaches capacity analysis, and whether any benefits could be gained from cross-pollination. This paper utilizes more than 50 past capacity studies from the U.S. and Europe to describe the different railroad capacity definitions and approaches, and then categorizes them, based on each approach. The capacity methods are commonly divided into analytical and simulation methods, but this paper also introduces a third, “combined simulation–analytical” category. The paper concludes that European rail studies are more unified in terms of capacity, concepts, and techniques, while the U.S. studies represent a greater variation in methods, tools, and objectives. The majority of studies on both continents use either simulation or a combined simulation–analytical approach. However, due to the significant differences between operating philosophy and network characteristics of these two rail systems, European studies tend to use timetable-based simulation tools as opposed to the non-timetable-based tools commonly used in the U.S. rail networks. It was also found that validation of studies against actual operations was not typically completed or was limited to comparisons with a base model.

Keywords Railroad capacity • Simulation • Railroad operation • The U.S. and European railway characteristics

1 Introduction

Typically, the capacity of a rail corridor is defined as the number of trains that can safely pass a given segment within a period of time. The capacity is affected by variations in system configurations, such as track infrastructure, signaling system, operation philosophy, and rolling stock.

The configuration differences between European and the U.S. rail systems may lead to different methodologies, techniques, and tools to measure and evaluate the capacity levels. There are high utilization corridors in Europe where intercity passenger, commuter, freight, and even high-speed passenger services operate on shared tracks, and all train movements follow their predefined schedule in highly structured daily timetables that may be planned for a full year in advance. On the contrary, the prevalent operations...
pattern on current shared corridors in the U.S. follows unstructured (improvised) philosophy, where schedules and routings (especially for freight trains) are often adjusted on a daily or weekly basis. Recently, the U.S. has placed an increasing emphasis on either the development of new higher speed passenger services, or to incrementally increase the speeds of current passenger services on selected shared corridors [1]. At the same time, the slower speed freight rail transportation volumes are also expected to increase [2]. These increases in volumes and operational heterogeneity can be expected to add pressure for higher capacity utilization of the U.S. shared-use corridors. Capacity measurement and analysis approaches (and their methods and tools) will play a crucial part in preparing the U.S. network for these changes. To maximize the efficiency of future improvements, such as new passenger and high-speed rail services, the accuracy and applicability of capacity tools and methods in the U.S. environment need to be carefully evaluated. Whether the analytical and operational approaches utilized in Europe would provide any benefits for the U.S. shared-use corridors should also be reviewed.

This paper starts by identifying the various definitions of capacity and by discussing the similarities and differences between the U.S. and European rail systems that may affect both the methods and outcomes of capacity analysis. It will also identify different approaches to conduct the analysis and concludes with an examination of several past capacity studies from both continents.

2 What is capacity?

2.1 Capacity concept and definitions

The definition used for rail capacity in the literature varies based on the techniques and objectives of the specific study. For instance, Barkan and Lai [3] defined capacity as “a measure of the ability to move a specific amount of traffic over a defined rail corridor in the U.S. rail environment with a given set of resources under a specific service plan, known as level of service (LOS)”. They listed several infrastructure and operational characteristics which affect capacity levels, including length of subdivision, siding length and spacing, intermediate signal spacing, percentage of number of tracks (single, double, and multi-tracks), and heterogeneity in train types (train length, power-to-weight ratios). In another paper, Tolliver [4] introduced freight rail capacity as the number of trains per day for typical track configurations depending on several factors, such as track segment length, train speed, signal aspects and signal block length, directional traffic balance, and peaking characteristics. The American Railroad Engineering and Maintenance-of-Way Association (AREMA) offers a simplified approach for line capacity that estimates practical capacity by multiplying theoretical capacity \( C_t \) and dispatching efficiency \( E \) of the line \( C = C_t \times E \). AREMA’s method for calculating theoretical capacity and dispatching efficiency requires consideration of various factors, such as number of tracks, the operations rules (single or bi-direction operation), stopping distance between trains (or headway), alignment specifications (grade, curves, sidings, etc.) trains specifications (type of train, length, weight, etc.), maintenance activities requirements, and the signaling and train control systems [5]. A capacity modeling guidebook for the U.S. shared-use corridors, released by the Transportation Research Board (TRB), defines capacity as “the capability of a given set of facilities, along with their related management and support systems, to deliver acceptable levels of service for each category of use.” Similar to the other capacity definitions, TRB notes that different parameters and variables should be considered in the capacity analysis, such as train dispatching patterns, train type and consist, signaling system, infrastructure, track maintenance system, etc. [6].

In Europe, the most common method for capacity analysis is provided by the International Union of Railways (UIC) code 406. According to UIC 406, there is no single way to define capacity, and the concerns and expectations vary between different points of view by railroad customers, infrastructure and timetable planners, and railroad operators. UIC also emphasizes that the capacity is affected by interdependencies and the interrelationships between the four major elements of railway capacity including average speed, stability, number of trains, and heterogeneity, as shown in Fig. 1 [7]. According to the figure, a rail

\[1\] The state of keeping the same train schedule by providing time margins/buffers between trains arrival/departures; despite of minor delay which may occur during operation.

\[2\] Diversity level of train types which are in operation along a shared-use corridor.
line with various types of trains on the same track (mixed traffic operations or shared-use corridor) has a higher heterogeneity level compared to the urban metro (subway) system with dedicated right-of-way and homogeneous operations. While the average speed of a mixed traffic corridor might be higher than a dedicated metro line, the various train types reduce the stability of train schedules, as well as the total number of trains that can operate on the corridor, due to increased headway requirements.

According to UIC, the absolute maximum capacity, or “Theoretical Capacity”, is almost impossible to achieve in practice, and it is subject to:

- Absolute train-path harmony (the same parameters for majority of trains)
- Minimum headway (shortest possible spacing between all trains)
- Providing best quality of service [7].

In addition to the UIC literature, research conducted as part of European Commission’s “Improve Rail” project produced a definition of ultimate capacity that was similar to the UIC’s theoretical capacity definition, but placed higher emphasis on the train schedules and running time [8].

2.2 Capacity metrics

The literature categorizes the main metrics of capacity level measurements into three groups: throughput (such as number of trains, tons, and train-miles), LOS (terminal/station dwell, punctuality/reliability factor, and delay), and asset utilization (velocity, infrastructure occupation time, or percentage) [9]. In 1975, The Federal Railroad Administration (FRA) introduced a parametric approach developed by “Peat, Marwick, Mitchell and Company” to measure capacity in the U.S. rail network based on delay units (hours per 100 train-miles) [4]. The European rail operators typically use throughput metrics (number of trains per day or hours) to measure the capacity levels, although punctuality and asset utilization metrics are also applied as secondary units [8, 10].

3 Differences between the U.S. and European rail systems

The U.S. and European rail networks have several similarities, such as mixed operations on shared-use corridors, and using modern signaling and traffic control systems (e.g., developing ETCS in Europe and PTC in the U.S.). On the other hand, significant differences also exist and they may change the preferred methodologies, tools, and the outcomes of capacity analysis. Figure 2 and the following discussion uses the literature review to highlight several key differences between infrastructure, signaling, operations, and rolling stock in Europe and the U.S.

3.1 Infrastructure characteristics

- Public versus private ownership of infrastructure The ownership of rail infrastructure is one of the important differences between Europe and the U.S. rail networks.
More than 90% of the infrastructure is owned and managed by private freight railroads in the U.S., while in Europe almost all infrastructure is owned and managed by governments or public agencies. In addition, operations and infrastructure are vertically separated in Europe, while in the U.S. the majority of operations (mainly freight) are controlled by the same corporations who own the infrastructure. The ownership and vertical separation have wide impact in the railroad system. Perhaps the greatest effect is on the prioritization of operations and accessibility for operating companies, but other aspects, such as operations philosophy, maintenance strategy and practices, signaling and train control systems, rolling stock configuration, and capital investment strategies are also affected [4, 11].

- **Single versus double track** More than 46% of rail corridors in Europe are at least double track [12, 13], while approximately 80% of the U.S. rail corridors are single track [2, 4].

- **Directional versus bidirectional** Most of the U.S. double tracks operate in bidirectional fashion and use crossovers along the corridor, while directional operation with intermediate sidings and stations is the common approach in Europe [4].

- **Distance between sidings** The distances between stations and sidings in the European rail network are generally shorter than in the U.S. The average distance between sidings/stations throughout the European network (total route mileage vs. number of freight and passenger stations) is approximately four miles between sidings/stations in both UK and Germany [13, 14]. In the U.S., the distance between sidings varies greatly between corridors. On double track sections, passing sidings are typically further apart than in Europe, often more than twice the average European distance [11, 15].

- **Siding length** Siding/yard tracks in the U.S. are typically longer than the European rail network, but in many cases are still not sufficient for the longest freight trains operating today [11, 16].

- **Track conditions** Typically, railroad structure in the U.S. is designed for higher axle loads, but has tighter horizontal curves (smaller radius) and lower maximum speed operations than the European rail network [11, 16].

- **Grade crossings** There are approximately 227,000 active grade crossings along the main tracks in the U.S. [17, 18], while there are few grade crossings on the main corridors in Europe, partially due to higher train speeds. High frequency of grade crossings and difficulty of their elimination cause operational and safety challenges for increased train speeds in the U.S. [19].

### 3.2 Signaling characteristics

- **Manual blocking versus signaling systems** Manual blocking is absent on main passenger corridors in the U.S. today, but relatively common on lower density branch ones, including some of the lines proposed for passenger corridors. In Europe, most shared-use corridors are equipped with one of the common signaling systems [20].

- **Cab signaling** A more significant difference is the extensive use of cab signaling and enforced signal systems, such as ETMS and ATS in Europe. Implementation of automatic systems is limited in the U.S., despite the current effort to introduce the positive train control (PTC) on a large portion of corridors [11].

### 3.3 Operation characteristics

- **Improvised versus structured operation** While some specific freight trains (mainly intermodal) have tight schedules, the U.S. operations philosophy is based on the improvised pattern with no long-term timetable or dispatching plan. On the passenger side, the daily operation patterns of many Amtrak and commuter trains are also developed without details, anticipating improvised resolution of conflicts among the passenger trains, or between passenger and freight trains. In Europe, almost all freight and passenger trains have a regular schedule developed well in advance, known as structured operations [21].

- **Freight versus passenger traffic** The majority of the U.S. rail traffic is freight, while the majority of European rail traffic is passenger rail [4, 22].

- **Delay versus waiting time** Delay (deviation of train arrival/departure time from what was predicted/planned) and waiting time (scheduled time spent at stations for passing or meeting another train) are two fundamental concepts in the railroad operations. The waiting time concept is typically used in Europe to manage rail operations, due to the structured operations pattern with strict timetables. Delay is more commonly used in the U.S. capacity analysis as the main performance metric, while it is limited in Europe to the events that are not predictable in advance [21].

- **Punctuality** The punctuality criteria of trains are quite different in the U.S. and Europe. Amtrak’s trains are considered on-time if they arrive within 15 min of a scheduled arrival (departure) time from what was predicted/planned and waiting time (scheduled time spent at stations for passing or meeting another train) are two fundamental concepts in the railroad operations. The waiting time concept is typically used in Europe to manage rail operations, due to the structured operations pattern with strict timetables. Delay is more commonly used in the U.S. capacity analysis as the main performance metric, while it is limited in Europe to the events that are not predictable in advance [21].
the freight trains performance or infrastructure failure [23]. The passenger trains in Europe have shorter average delay per train. For instance, Network Rail in the UK reported that approximately 90% of all short-distance passenger trains had less than 5 min deviation from planned timetable, while for long-distance trains, the same was true for deviation less than 10 min [24]. In Switzerland, more than 95% of all passenger trains are punctual with an arrival delay of 5 min or less [25]. The punctuality of European freight trains in 2003 was reported to be approximately 70% [26].

3.4 Rolling stock characteristics

- **Train configuration (length and speed)** Typically freight trains in the U.S. are longer and heavier than freight trains in Europe. Based on the Association of American Railroads (AAR), the typical number of cars in a U.S. freight train varies between 63–164 cars in the West and 57–110 cars in the East, while the typical number in Europe is 25–40. From speed perspective, the average speed of intercity passenger trains in Europe is significantly faster than in the U.S. [2, 11, 16]. Freight trains also typically operate at higher speeds and with less variability in Europe.

- **Diversity of freight versus passenger trains** The U.S. rail transportation is more concentrated on the freight trains than Europe, and there is a great diversity between the types, lengths, etc., of freight trains. On the passenger side, Europe has more diverse configurations (such as speed, propulsion, train type, power assignment, HSR services, diesel, and electric multiple unit (EMU) trains) in comparison to the U.S. [2, 20].

While the principles of rail capacity remain the same in all rail networks, the above characteristics have an effect on capacity and its utilization. What remains unclear is how these differences have been considered in various capacity analysis tools and methodologies used and how much they limit the applicability of the U.S. tools in the European environment and vice versa. This paper introduces some of the common tools and methodologies, including examples of their use in past studies, but excludes any direct comparisons between the capabilities of individual tools. A more detailed (case study based) comparative analysis of selected U.S. and European simulation tools and methodologies is provided by the authors in separate papers [27, 28].

4 Capacity measurement, analytical, simulation, and combined approaches

Generally speaking, there are two main approaches to improve the capacity levels: either by applying new capital investment toward upgraded or expanded infrastructure or by improving operational characteristics and parameters of the rail services [29]. In either approach, it is necessary to assess and analyze the benefits, limitations, and challenges of the approach, often done through capacity analysis. The literature classifies capacity analysis approaches and methodologies in several different ways. Although the approaches differ, the input typically includes infrastructure and rolling stock data, operating rules, and signaling features. Abril et al. [30] classified the capacity methodologies as analytical methods, optimization methods, and simulation methods. Pachl [31] divided the capacity methodologies into two major classes: analytic and simulation. Similar categorization was used in research conducted by Murali on delay estimation technique [32]. Khadem Sameni, and Preston et al. [9] categorized capacity methods to timetable-based and non-timetable-based approaches. The capacity guidebook developed by TRB also divides capacity evaluation methods into two approaches: simple analysis, and complex simulation modeling [6]. Finally, in research conducted at the University of Illinois, Sogin, Barkan et al. [3, 33] divided capacity methods to theoretical (analytical), parametric, and simulation methods. Overall, the analytical and simulation methods are the most common methods found in the literature. For our review, we divided methods into three groups: analytical, simulation, and combined. Although the term “combined methodology” was not used commonly in the reviewed literature, it was added as a new class to address the fact that many reviewed studies took advantage of both analytical and simulation methods.

4.1 Analytical approach

The **analytical approach** typically uses several steps of data processing through mathematical equations or algebraic expressions and is often used to determine theoretical capacity of the segment/corridor. The outcomes vary based on the level of complexity of the scenario and may be as simple as the number of trains per day, or a combination of several performance indicators, such as timetable, track occupancy chart, fuel consumption, speed diagrams, etc. Analytical methods can be conducted without software developed for railroad applications, such as Microsoft Excel, but there are also analytical capacity tools specifically developed for rail applications. One example is SLS PLUS in Germany, which is used in the German rail network (DB Netz AG) for capacity estimation through analytical determination of the performance, asynchronous simulation, and manual timetable construction [34]. Figure 3 presents the different levels of analytical approach and how complexity can be added to the process to provide more detailed results. In some cases, analytical models are
called optimization methods or parametric models, taking advantage of different modeling features, such as probabilistic distribution or timetable optimization. The latter method, timetable optimization, is typically achieved using specialized software or simulation tools [30, 31].

Timetable compression method is one of the main analytical approaches in Europe to improve the capacity levels, especially on the corridors with pre-determined timetables (structured operation pattern). A majority of techniques and tools for improving the capacity utilization in Europe, including the UIC method (leaflet 406), are partly developed based on timetable compression [7, 10, 35–37]. The UIC’s method modifies the pre-determined timetable and reschedules the trains as close as possible to each other [30]. Figure 4 provides an example of the methodology where a given timetable along a corridor with quadruple tracks (Scenario a) is first modified by compressing the timetable (Scenario b) and then further improved by optimizing the order of trains (Scenario c). As demonstrated in the figure, the third scenario could provide a higher level of theoretical capacity in comparison to the Scenarios a and b [10]. It should be emphasized that due to the unstructured nature of the U.S. rail operation philosophy, timetable compression technique has not been practically applied yet in the U.S. rail environment.

4.2 Simulation approach

Simulation is an imitation of a system’s operation which should be as close as possible to its real-world equivalent [30]. In this approach, the process of simulation is repeated several times until an acceptable result is achieved by the software. The data needed for the simulation are similar to the analytical methods, but typically at a higher level of detail. The simulation practices in rail industry started in the early 1980s through the development of models and techniques, such as dynamic programming and branch-and-bound, proposed by Petersen, as well as heuristic methods developed by Welch and Gussow [30]. Today, the simulation process utilizes computer tools to handle sophisticated computations and stochastic models in a faster and more efficient way. The simulation approaches use either general simulation tools, such as AweSim, Minitab, and Arena [32, 38]; or commercial railroad simulation software specifically designed for rail transportation, such as RTC, MultiRail, RAILSIM, OpenTrack, RailSys, and CMS [9, 30]. The use of general simulation tools requires the user to develop all models, equations, and constraints step by step (often manually). This requires more expertise, creativity, and effort, but it can also offer more flexible and customization when it comes to results and outputs. The commercial railroad simulation tools offer an easier path toward development of different scenarios, in addition to providing a variety of outputs in a user-friendly way, but the core decision models and processes are not easily

![Fig. 3 Levels of analytical approaches for capacity analysis](image)

![Fig. 4 Actual timetable for a quadruple-track corridor (a) compressed timetable with train order maintained (b) compressed timetable with optimized train order (c) (Note chart layout follows typical European presentation and solid and dot lines represent different types of trains) [10]](image)
customizable or reviewable, which may reduce the flexibility of applying these tools.

The commercial railroad simulation software typically revolves around two key simulation components: (1) train movement and (2) train dispatching. The first component uses railroad system component data provided as an input, such as track and infrastructure characteristics (curvature and grades), station and yard layout, signaling system, and rolling stock characteristics, to calculate the train speed along the track. Train dynamics is typically determined based on train resistance formulas, such as Davis equation and train power/traction. The dispatching simulation component typically emulates (or attempts to emulate) the action of the dispatcher in traffic management, but in some cases, it can be also used as part of a traffic management software to help traffic dispatchers to manage and organize the daily train schedules (Fig. 5) [21].

According to Pachl, the simulation method can also be divided into asynchronous and synchronous methods. Asynchronous simulation software is able to consider stochastically generated train paths within a timetable, following the scheduling rules and the train priorities. In synchronous simulation, the process of rail operations is followed in real-time sequences, and the results are expected to be closely aligned with real operations. In contrast to the asynchronous method, synchronous methods cannot directly simulate the scheduling, or develop a timetable, without use of additional computer tools and programs to create a timetable [31]. The outputs of simulation software typically include several parameters such as delay, dwell time, waiting time, elapsed time (all travel time), transit time (time between scheduled stops), trains speed, and fuel consumption of trains [21, 30].

4.2.1 Simulation methods: timetable based versus non-timetable based

The commercial railroad simulation software can be classified in two groups: non-timetable based and timetable based. The non-timetable-based simulations are typically utilized by railroads that use the improvised (unstructured) operation pattern without an initial timetable, such as the majority of the U.S. rail networks. In this type of simulation, after loading the input data in the software, the train dispatching simulation process uses the departure times from the initial station that are provided as part of the input data. The software may encounter a problem to assign all trains and request assistance from the user to resolve the issue by manually adjusting the train data, or by modifying the schedule constraints [9, 21].

The simulation procedure in timetable-based software (typically used in Europe) is based on the initial timetable of trains and the objective is to improve the timetable as much as possible. The UIC’s capacity approach is often one of the main theories behind the timetable-based simulation approach. The simulation process in this methodology begins with creating a timetable for each train. In the case of schedule conflict between the trains, the user must adjust the timetable until a feasible schedule is achieved. However, the user actions are more structured compared to the improvised method, and is implemented as part of the simulation process [21]. There are several common software tools in this category, such as MultiRail (U.S.), RAILSIM (U.S.), OpenTrack (Switzerland), SIMONE (The Netherlands), RailSys (Germany), DEMIURGE (France), RAILCAP (Belgium), and CMS (UK) [9, 30]. A comprehensive capability review of various simulation tools is outside the scope of this paper, but three simulation packages (RTC, RailSys, and OpenTrack) are briefly introduced to demonstrate some key differences between non-timetable-based and timetable-based software.

The rail traffic controller (RTC), developed by Berkeley Simulation Software, is the most common software in the non-timetable-based category, used extensively by the U.S. rail industry [9]. RTC was launched in the U.S. (and North American) rail market in 1995 and has since been
continuously developed and upgraded. Since majority of the U.S. train services (particularly freight trains) have frequent adjustments in their daily schedules, RTC has several features and tools for simulating the rail operations in non-scheduled environment, including train movement animation, automated train conflict resolution, and randomization of train schedule. The dispatching simulation component of RTC is based on a decision support core, called “meet-pass N-train logic”. For any dispatching simulation practice, “meet-pass N-train logic” will decide when the given trains should exactly arrive and depart from different sidings, based on the defined train priorities and preferred times of departure. The simulation outcomes may include variation between the simulated departure times and preferred times [39]. Besides its decision core fitting the U.S. operational philosophy, RTC has other system characteristics, such as attention to grade crossing, that make it well suited to the U.S. market.

RailSys, developed by Rail Management Consultants GmbH (RMCon) in Germany, is an operation management software package that includes features, such as timetable construction/slot management, track possession planning, and simulation. It has been in the market since 2000 and it is one of the commonly used timetable-based simulation software in Europe. The capacity feature of RailSys uses the UIC code 406 which is based on the timetable compression technique [40, 41]. OpenTrack is another common simulation package in Europe. It was initially developed by Swiss Federal Institute of Technology-Zurich (ETH-Zurich) and has since 2006 been supplied by OpenTrack Railway Technology Ltd. OpenTrack is also a timetable-based simulation tool with several features, such as automatic conflict resolution based on train priority, routing options and delay probabilistic functions, as well as several outputs and reporting options, such as train diagram, timetable and delay statistics, station statistics, and speed/time diagram [42, 43].

4.3 Combined analytical–simulation approach

In addition to the analytical and simulation approaches, a combined analytical–simulation method can also be used to investigate the rail capacity. Parametric and heuristic modelings (in analytical approach) are more flexible when creating new aspects and rules for the analysis. On the other hand, updating the railroad component input data and criteria tends to be easier in the simulation approach, and the process of running the new scenarios is generally faster, although simulation may place some limitations when adjusting the characteristics of signaling or operation rules. A combined simulation–analytical methodology takes advantage of both methodologies’ techniques and benefits, and the process can be repeated until an acceptable set of outputs and alternatives is found (Fig. 6). There are several ways to combine analytical and simulation tools. For instance, finding a basic and reasonable schedule of trains through simulation, followed by analytical schedule can be considered as one example of combined analytical–simulation approach. Another example would be application of a simplistic analytical model to provide the basic inputs, such as determining the type of signaling system, or developing train schedule, followed by more extensive and detailed analysis in commercial rail simulation tools.

5 Review of capacity studies in the U.S. and Europe

The approaches, methodologies, and tools highlighted in previous section have been applied in numerous U.S. and European capacity studies. The team reviewed 51 total studies using all three approaches (17 analytical studies, 22 simulation studies, and 12 combined simulation–analytical approaches). Then, 25 of them that had sufficient details of the study approach and respective results were used to conduct a detailed assessment of studies conducted in Europe versus in the U.S.

5.1 Studies with analytical approach

One of the first analytical models was developed by Frank [44] in 1966 by studying the delay levels along a single track corridor considering both directional and bidirectional scenarios. He used one train running between two consecutive sidings (using manual blocking system) and a single average speed for each train to calculate the number of possible trains (theoretical capacity) on the given segment. Petersen [45] expanded Frank’s idea in 1974 by considering two different speeds, independent departure

Fig. 6 Basic diagram of combined analytical–simulation approach for capacity analysis
times, equal spacing between sidings, and constant delays between two trains. Higgins et al. [46] developed a model in 1998 for urban rail networks to evaluate the delays of trains by considering different factors such as trains’ schedule, track links, sidings, crossings, and the directional/bidirectional operation patterns throughout the network.

De Kort et al. [47] analyzed the capacity of new corridors in 2003 by applying an optimization method and considering uncertainty of demand levels on the planned route. Ghoseiri et al. [48] introduced a multi-objective train scheduling model of passenger trains along single and multiple tracks of rail network, based on minimizing the fuel consumption cost as well as minimizing the total passenger time of trains. Burdett and Kozan [49] developed analytical techniques and models in 2006 to estimate the theoretical capacity of a corridor based on several criteria, such as mixed traffic, directional operation pattern, crossings and intermediate signals along the track, length of the trains, and dwell time of trains at sidings or stations. Wendler [50] used queuing theory and the semi-Markov chains in 2007 to provide a technique of predicting the waiting times of trains based on the arrival times, minimum headway of trains, and the theory of blockings. Lai and Barkan [3] introduced an enhanced technique of capacity evaluation tools in 2009 based on the parametric modeling of capacity evaluation, which was initially developed by CN Railroad. The railroad capacity evaluation tool (RCET), developed by Lai and Barkan, can evaluate the expansion scenarios of network by estimating the line capacity and investment costs, based on the future demand and available budget.

Lindner [51], recently, reviewed the applicability of timetable compression technique, UIC code 406, to evaluate the corridor and station capacity. He used several case studies and examples to conclude that UIC code 406 is a good methodology for evaluating the main corridor capacity of network systems. Corman et al. [52] conducted another study in 2011 to analyze an innovative approach of optimization of multi-class rescheduling problem. The problem focused on train scheduling with multiple priority classes in different steps, using the branch-and-bound algorithm.

In addition to specific studies on railroad capacity, a book edited by Hansen and Pachl [53], containing several articles and sections conducted by different railroad studies mostly by European universities and academic centers, was released in 2008 as one of the latest resources of timetable optimization and train rescheduling problem. The book covers articles on various topics, such as cyclic timetabling, robust timetabling, use of simulation for timetable construction, statistical analysis of train delays, rescheduling, and performance evaluation.

5.2 Studies with simulation and combined approach

Studies using analytical approach preceded simulation and combined approaches. One of the first general simulation studies was conducted by Petersen et al. in 1982 by dividing a given corridor into different track segments where each segment represented the distance between two siding/switches [54]. Kaas [55] developed another general simulation model in 1991, called “Strategic Capacity Analysis for Network” (SCAN), by defining different factors of simulation which could determine the rail network capacity. In another study, Dessouky et al. (1995) [56] used a general simulation model for analyzing the track capacity and train delay throughout a rail network. Their model included both single and double track corridors, as well as other network parameters, such as trains length, speed limits, and train headways. Sogin et al. [57] recently used RTC to simulate several case studies at University of Illinois, Urbana-Champaign. One of their studies evaluated the impact of passenger trains along U.S. shared-use double track corridors, considering different speed scenarios. They concluded that increasing speed gap between the trains can result in higher delays.

The Missouri DOT used the combined analytical–simulation approach in 2007 to analyze the rail capacity on the Union Pacific (UP) corridor between St. Louis to Kansas City to improve the passenger train service reliability and to reduce the freight train delay. Six different alternatives were generated based on a theory of constraints (TOC) analysis 3 and then compared with each other using the Arena simulation method. A set of recommendations and capital investment for each proposed alternative were proposed with respect to delay reduction [38].

In another project, Washington DOT (WSDOT) conducted a master plan in 2006 to provide a detailed operation and capital plan for the intercity passenger rail program along Amtrak Cascades route. The capacity of the corridor was also evaluated using the combined simulation–analytical approach. First, analytical methods were used to determine the proposed infrastructure. Then the proposed traffic and infrastructure were simulated with RTC software to test the proposed infrastructure and operational results. After running simulation on RTC software, a heuristic (analytical) method, called root cause analysis (RCA), was applied to evaluate the simulation output. The objective of RCA method was to identify the real reason of a delay along the rail corridor by comparing

--

3 TOC is a management technique that focuses on each system constraints based on five-step approach to identify the constraints and restructuring the rest of the system around it. These steps are: 1) identify the constraints, 2) decision on how to exploit the constraints, 3) subordinate everything around the above decision, 4) elevate the system’s constraints, and 5) feedback, back to step 1.
the output reports of each delayed train with other train services and to re-adjust the simulation outputs to be more accurate, in addition to locating infrastructure bottlenecks which caused the capacity issues and delays [58].

The Swedish National Rail Administration (Banverket) carried out a research project in 2005 to evaluate the application of the UIC capacity methodology (timetable compression) for the Swedish rail network. RailSys software was used for the simulations and the research team analytically evaluated the capacity consumption, its relationship with time supplements (or buffer times), and the service punctuality. The research concluded that the buffer times are absolutely necessary for the service recovery, in case of operation interruption. When there is no buffer time, the service punctuality can be significantly degraded due to increased capacity consumption. Banverket also confirmed the validity of the framework and the results of the UIC’s approach and asked their experts and consultants to implement this analytical approach in their network [36].

In research conducted through combined analytical–simulation approach, Medeossi et al. [59] applied stochastic approach on blocking times of trains to improve the timetable planning using OpenTrack simulation software. They redefined timetable conflicts by considering a probability for each train conflict as a function of process time variability. The method repeatedly simulated individual train runs on a given infrastructure model to show the occupation staircase of trains in different color spectrums, while each color represents the probability of trains’ conflict which should be resolved.

Recently, a new “Web-based Screening Tool for Shared-Use Rail Corridors” was developed in the U.S. by Brod and Metcalf [60] to perform a preliminary feasibility screening of proposed shared-use passenger and freight rail corridor projects. The outcomes can be used to either reject projects or move them to more detailed analytical/simulation investigations. The concept behind the tool is based on a simplified simulation technique which does not provide optimization features or complex simulation algorithms. The tool requires development of basic levels of infrastructure, rolling stock, and operation rules (trains schedule) of the given corridor; and a conflict identifier assists the user in identifying locations for a siding or yard extension needed to resolve the conflict between existing and future train services.

5.3 Detailed assessment of selected studies

Only a subsection of reviewed studies offered sufficiently detailed explanation of the study approach and respective results. These studies were broken into several categories and subcategories for a comparison between the studies conducted in Europe versus in the U.S. Table 1 and the following discussion summarize the approach, tools used,

| Table 1 | Category/subcategory breakdown of 25 selected studies in the U.S. and Europe [2, 3, 9, 10, 21, 25, 29, 32, 35, 36, 38, 57–70] |
| Category/subcategory | The U.S. (14 studies) | Europe (11 studies) |
|-----------------------|-----------------------|---------------------|
| Capacity approach      | Analytical            | 4 studies [2, 3, 29, 32] |
|                       | Simulation            | 5 studies [57, 61, 62, 64, 65] |
|                       | Combined analytical–simulation | 5 studies [9, 21, 38, 58, 60] |
| Tools/software         | Only mathematical/parametric modeling | 3 studies [2, 3, 29] |
|                       | General simulation software | 3 studies [32, 38, 60] |
|                       | Timetable-based simulation software | – |
|                       | Non-timetable-based simulation software | 8 studies [9, 21, 57, 58, 61, 62, 64, 65] |
| Purpose of research   | New methodology development/methodology approval | 5 studies [3, 9, 21, 29, 60] |
|                       | Master plan/capacity analysis | 3 studies [2, 38, 58] |
|                       | Academic research/project | 6 studies [32, 57, 61, 62, 64, 65] |
| Type of outcomes/solutions | Delay analysis/Improvement | 3 studies [32, 57, 61] |
|                       | Infrastructure development, Rescheduling/operation changes | 1 study [2] |
|                       | Combination of above solutions | 2 studies [21, 62] |
|                       | New tools/methodology approval | 4 studies [38, 58, 64, 65] |
| Validation of simulation results | No comparison | 4 studies [3, 9, 29, 60] |
|                       | Base model | 6 studies [2, 9, 29, 57, 61, 62] |
|                       | Base and alternative results | 7 studies [3, 21, 38, 58, 60, 64, 65] |

J. Mod. Transport. (2015) 23(1):30–42
study purpose, types of outcomes, and validation methods of the 25 studies selected for more detailed comparison.

Approach Most studies used either simulation or combined analytical–simulation approaches. However, research conducted by AAR [2], University of Illinois at Urbana-Champaign (UIUC) [3, 29], and University of Southern California (USC) [32] applied analytical-only methodologies.

Tools and software All European studies used timetable-based simulation software (e.g., RailSys, OpenTrack, ROMA), while the U.S. studies relied on other tools like optimization/parametric modeling (UIUC and USC) [2, 3, 29], general simulation software (e.g., Arena) [38], web-based screening tools [60], and non-timetable-based rail simulation software (RTC).

Purpose of Research Three main purposes were identified for studies: (1) introducing new methodology for capacity evaluation, (2) evaluating the capacity status of a given corridor as part of a corridor master plan development, and (3) academic research on various capacity issues. The majority of European studies (Denmark, Austria, Germany, the Netherlands, and Sweden) were conducted by industry or academic research teams to justify and evaluate the UIC’s approach (UIC code 406) for capacity evaluation [10, 35, 36, 67, 70], while the objectives of the U.S. studies included all three subcategories.

Type of outcomes or solutions The outcomes and solutions obtained from the U.S. studies included variety of different types such as delay analysis (UIUC by using RTC and USC by using Awesim/Minitab), rescheduling and recommendations related to current operations (UIUC and White) [21, 62], infrastructure development, and combination of all outcomes mentioned above (typically as part of a master plan). In addition, new tools and parametric models were also developed as the final outcome of three U.S. studies (mainly by UIUC). The outcomes of European studies were not so diverse, as they either approved the application of UIC’s capacity methodology to be used on their network [10, 36], or suggested network rescheduling and operational changes (the timetable compression concept) [25, 35, 63, 67, 70]. One of the common conclusions of various studies was the identification of operational heterogeneity as a major reason of delay, especially in the U.S. rail network with unstructured operation pattern.

Validation of simulation results None of the studies using analytical method compared the results to a real-life scenario, but some of the simulation-based studies validated the results with one of the following three types of comparisons:

- No comparison No specific information or comparison was provided between simulated results and actual practices. As presented in Table 1, approximately one-third of the studies (9 out of 25) did not validate the simulation results, either because the study was not based on actual operational data, or comparison was not conducted as part of the research.
- Base model Only the results of a base model were compared with the real data. More than half of the studies (14 out of 25) compared the simulation results only with the base model.
- Base and alternative results In addition to base model comparison, the alternative outcomes were compared with the real data. Only two studies belonged to this category.

6 Summary and conclusions

This paper has provided an overview of capacity definitions, alternative analysis approaches, and tools available to evaluate capacity. It has also highlighted the key similarities and differences between the U.S. and European rail systems and how they affect related capacity analysis. Finally, the paper has reviewed over 50 past capacity studies and selected 25 of them for more detailed investigation.

The review revealed no single definition of railroad capacity. Rather, the definition varies based on the techniques and objectives of the specific study. The capacity analysis approaches and methodologies can also be classified in several ways, but are most commonly divided into analytical and simulation methods. This paper also introduced a third “combined” approach that uses both analytical and simulation approaches.

While the objective of capacity analysis is common, there are several differences between the U.S. and European rail systems that affect the approaches, tools, and outcomes of analysis. Europe tends to use a structured operations philosophy and thus uses often timetable-based simulation approaches for analysis, while the improvised U.S. operations warrant non-timetable-based analysis. Other factors, such as differences in ownership, type and extent of double track network, distance between and length of sidings, punctuality of service, dominating type of traffic (passenger vs. freight), and train configuration also affect the analysis methods and tools.

The review of over 50 past studies revealed that a majority of analyses (approximately 65 % of studies) utilized either simulation or combined simulation–analytical methods, while the remainder relied on analytical methods. Although the general simulation tools and modeling approaches have been used, most studies use commercial simulation software either in the U.S. (non-timetable based) or in Europe (timetable based). Based on the more detailed review of 25 of the studies, European capacity
analysis tends to be linked to the UIC 406 method, while the U.S. does not seem to have as extensive principles as the European case studies, but the methodologies vary more from one study to another. The outcomes of European studies were also less diverse than in the U.S., and commonly suggested rescheduling and operation changes as the solutions for capacity improvement. Also the studies showed limited effort in comparing the simulation results to the actual conditions (the validation step), especially after recommended improvements were implemented. Only two studies did the full validation, 14 out of 25 only compared the results with the base model, and the remaining one-third of the studies had no validation process. Overall, it was found that there was no major divergence between approaches or criteria used for capacity evaluation in the U.S. and Europe. However, there are differences in the tools used in these two regions, as the tool designs follow the main operational philosophy of each region (timetable vs. non-timetable) and include features that concentrate on other rail network characteristics for the particular region.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

1. FRA (2009) Vision for high-speed rail in America. Federal Railroad Administration (FRA), Washington, DC
2. Cambridge Systematics I (2007) National rail freight infrastructure capacity and investment study. Association of American Railroads (AAR), Cambridge
3. Lai Y-C, Barkan C (2009) Enhanced parametric railway capacity evaluation tool. J Transp Res Board 2117:33–40. doi:10.3141/2117-05 (National Academies, Washington, DC)
4. Tolliver D (2010) Railroad planning & design, CE 456/656 Lecture Notes, North Dakota State University
5. AREMA (2006) Manual for railway engineering. Volume 4-systems management. American Railway Engineering and Maintenance-of-Way Association, Lanham
6. Fox J, Hirsch P, Kanike O, Simpson DP, Cebula AJ, Bing AJ, Horowitz ES, Barkan C (2008) Capacity modeling guidebook for shared-use passenger and freight rail operations, NCHR Report 773. NCHR, Washington, DC
7. UIC (2004) UIC CODE 406 R-capacity, 1st edn. International Union of Railways (UIC), Paris
8. IMPROVERAIL (2003) IMPROVED tools for RAILway capacity and access management-deliverable 6 methods for capacity and resource management. Competitive and Sustainable Growth (GROWTH) Programme—European Commission, Brussels-Belgium
9. Sameni MK, Dingler M, Preston JM, Barkan CPL (2011) Profit-generating capacity for a freight railroad. In: TRB 90th Annual Meeting, TRB, Washington, DC
10. Alex Landex EA (2006) Evaluation of railway capacity. In: Annual Transport Conference, Aalborg University, Denmark
11. White T (2012) Comparison of U.S. and European railroads. Lecture notes for CE-4490. Michigan Tech. University-CEE Department, Houghton
12. Railway-Technical (2008) Railway statistics for Britain. http://www.railway-technical.com/statistics.shtml. Accessed May 2012
13. ERRAC (2004) A comparison of member state public research programmes with the ERRAC SRRA 2020. European Rail Research Advisory Council (ERRAC), Brussels
14. DB-AG (2011) Station categories 2012. DB-AG, Berlin-Germany (in German)
15. Amtrak (2011) Annual report fiscal year 2011. Amtrak, Washington, DC
16. Teixeira P (2009) Railway engineering principles, Technical University of Lisbon, CTIS Master’s course in MIT-Portugal Program, Lisbon
17. FRA (2008) Highway-rail grade crossing safety fact sheet. FRA Public Affairs, Washington, DC
18. DOT (2011) National transportation statistics. U.S. Department of Transportation, Washington, DC
19. NetworkRail (2010) Curtail crazy driving at level crossings or risk more lives, says rail chief. http://www.networkrailmedia centre.co.uk/Press-Released/CURTAIN-CRAZY-DRIVING-AT-LEVEL-CROSSINGS-OR-RISK-MORE-LIVES-SAYS-RAIL-CHIEF-13c1/SearchCategoryID-2.aspx. Accessed on 2012
20. Verkehrswissenschaftlich (2008) Influence of ETCS on the line capacity. International Union of Railways (UIC), Paris
21. White T (2005) Alternatives for railroad traffic simulation analysis. J Transp Res Board 1916:34–41 (TRB-Washington, DC)
22. Systra (2007) Technical classification and main technologies for the HSR systems. Complementary studies of Tehran-Qom-Isfahan HSR, Paris
23. Amtrak (2012) Monthly performance report for December 2011. AMTRAK, Washington, DC
24. Wearden G (2007) UK train times back on track. The Guardian, Friday 23 November 2007
25. Marco Luethi AN, Weidmann U (2007) Increasing railway capacity and reliability through integrated real-time rescheduling. Swiss Federal Railways, SBB AG, Zurich
26. CER (2004) Rail freight quality: meeting the challenge. A Report on the first year of the CER-UIC-CIT Freight Quality Charter. Community of European Railway and Infrastructure Companies (CER), Bruxelles, Belgium
27. Pouryousef H, Lautala P (2013) Evaluating the results and features of two capacity simulation tools on the SHARED-use corridors. In: ASME/ASCE/IEEE 2013 Joint Rail Conference (JRC2013), Knoxville, TN, USA
28. Pouryousef H, Lautala P (2014) Evaluating two capacity simulation tools on shared-use U.S. rail corridor. In: 2014 Annual Meeting of the Transportation Research Board, TRB, Washington, DC
29. Lai Y-C, Barkan C (2011) A comprehensive decision support framework for strategic railway capacity planning. J Transp Eng 137:738–749. doi:10.1061/(ASCE)TE.1943-5436.0000248 (ASCE 137 (October 2011))
30. Abril M, Barber F, Ingolotti L, Salido MA, Tormos P, Lova A (2007) An assessment of railway capacity. Technical University of Valencia, Valencia, Spain
31. Pachl J (2002) Railway operation and control. VTD rail publishing, Mountlake Terrace
32. Pavankumar Murali MMD, Ordonez F, Palmer K (2009) A delay estimation technique for single and double-track railroads. Department of Industrial & Systems Engineering, University of Southern California, Los Angeles, CA
33. Sogin SL, Barkan CPL (2012) Railroad capacity analysis. In: Railroad engineering education symposium (REES) by AREMA, Kansas City
34. Schulze GI (2012) SLS PLUS (Software Manual). Schultze + Gast Ingenieure, Berlin
35. Robert Prinz JH (2005) Implementation of UIC 406 capacity calculation method at Austrian Railway (OBB). Austrian Railway (OBB), Vienna
36. Banverket (2005) Application of the UIC capacity leaflet at Banverket. Banverket, Swedish National Rail Administration
37. Melody Khadem Sameni MD, Preston JM (2010) Revising the UIC 406 method: revenue generating capacity—JRC2010-36281. In: Joint Rail Rail Conference (JRC) 2010, Urbana, IL, USA, 2010
38. Noble JS, Charles N (2007) Missouri freight and passenger rail capacity analysis. Missouri Department of Transportation (MoDOT), Jefferson City
39. RTC (2011) RTC online help (Manual). Berkeley Simulation Software, CA, USA
40. RMCon (2012) RailSys 8 classic, RailSys 8 Enterprise, network-wide timetable and infrastructure management. Rail management consultants GmbH (RMCon), Hannover
41. RMCon (2010) RailSys user manual. Rail management consultants GmbH (RMCon), Hannover
42. Huerlimann D, Nash AB (2012) Manual of opentrack—simulation of railway networks. Version 1.6. OpenTrack Railway Technology Ltd. and ETH Zurich, Zurich, Switzerland
43. Huerlimann D (2012) OpenTrack—simulation of railway networks. OpenTrack Railway Technology Ltd., Zurich
44. Frank O (1966) Two-way traffic in a single line of railway. Oper Res 14:801–811
45. Petersen ER (1974) Over the road transit time for a single track railway. Transp Sci 8:65–74
46. Higgins A, Kozen E (1998) Modeling train delays in urban networks. Transp Sci 32(4):251–356
47. de Kort AF, Heidergott B, Ayhan H (2003) A probabilistic approach for determining railway infrastructure capacity. Eur J Oper Res 148:644–661
48. Ghoseiril K, Szidaroszvky F, Asgharpour MJ (2004) A multi-objective train scheduling model and solution. Transp Res Part B 38:927–952
49. Burdett RL, Kozen E (2006) Techniques for absolute capacity determination in railways. Transp Res Part B 40:616–632
50. Wendler E (2007) The scheduled waiting time on railway lines. Transp Res Part B 41:148–158
51. Lindner T (2011) Applicability of the analytical UIC Code 406 compression method for evaluating line and station capacity. J Rail Transp Plan Manag 1:14–24
52. Corman F, D’Ariano A, Hansen IA, Pacciarelli D (2011) Optimal multi-class rescheduling of railway traffic. J Rail Transp Plan Manag 1:49–57
53. Hansen IA, Pachl J (2008) Railway timetable & traffic, analysis, modeling, simulation, 1st edn. Eurailpress, Hamburg
54. Petersen ER, Taylor AJ (1982) A structured model for rail line simulation and optimization. Transp Sci 16:192–206
55. Kaas AH (1991) Strategic capacity analysis of networks: developing and practical use of capacity model for railway networks. ScanRail consult, Technical University of Denmark
56. Dessouky MM, Leachman RC (1995) A simulation modeling methodology for analyzing large complex rail networks. Simulation 65(2):131–142
57. Sogin SL, Barken CPL, Lai Y-C, Saat MR (2012) Impact of passenger trains in double track networks. In: 2012 Joint rail conference, Philadelphia, Pennsylvania
58. Transit Safety Management Inc.; HDR Engineering (2006) Amtrak cascades operating and infrastructure plan—technical report. Vol 1. Washington State Department of Transportation-Freight Systems Division
59. Medeossi G, Longo G, Fabris Sd (2011) A method for using stochastic blocking times to improve timetable planning. J Rail Transp Plan Manag 1:1–13
60. Brod D, Metcalf AE (2014) Web-based screening tool for shared-use rail corridors (NCFRP REPORT 27). National cooperative freight research program—Transportation research board (TRB), Washington, DC
61. Sogin SL, Barken CPL, Lai Y-C, Saat MR (2012) Measuring the impact of additional rail traffic using highway & railroad metrics. In: 2012 Joint rail conference (JRC), Philadelphia, Pennsylvania
62. Dingler M, Lai Y-C, Barkan CPL (2009) Impact of operational practices on rail line capacity: a simulation analysis. In: 2009 Annual AREMA conference, Chicago, IL
63. Siggel S (2006) Analysis of capacity and optimization of infrastructure of the line minot (North Dakota)—Havre (Montana). Braunschweig Technical University, Braunschweig-Germany
64. Atanassov I, Dick CT, Barkan CPL (2014) Siding spacing and the incremental capacity of the transition from single to double track. In: 2014 Joint rail conference (JRC), Colorado Springs, CO, USA
65. Shih M-C, Dick CT, Sogin SL, Barkan CPL (2014) Comparison of capacity expansion strategies for single-track railway lines with sparse sidings. In: TRB 2014 Annual meeting, Washington, DC, USA
66. Schlechte T, Bornsdörfer R, Erol B, Graffagnino T, Swaret E (2011) Micro–macro transformation of railway networks. J Rail Transp Plan Manag 1:38–48
67. Gille A, Siefer T (2013) Sophisticated capacity determination using simulation. In: Transportation research board 2013 Annual meeting, Washington, DC, USA
68. Medeossi G, Longo G (2013) An approach for calibrating and validating the simulation of complex rail networks. In: TRB 2013 Annual meeting, Washington, DC, USA
69. Sipila H (2014) Evaluation of single track timetables using simulation. In: Transportation research board 2013 Annual meeting, Washington, DC, USA
70. Goverde RMP, Corman F, D’Ariano (2014) A investigation on the capacity consumption of Dutch railways for various signalling technologies and traffic conditions. In: TRB 2014 Annual Meeting, Washington, DC, USA