

Utilizing Bayesian Optimization for Timetable-Independent Railway Junction Performance Determination

Tamme Emunds ^{a,*}, Paul Brunzema ^b, Sebastian Trimpe ^b, Nils Nießen ^a

^a*Institute of Transport Science, RWTH Aachen University, Aachen, Germany*

^b*Institute for Data Science in Mechanical Engineering, RWTH Aachen University, Aachen, Germany*

Abstract

The efficiency of railway infrastructure is significantly influenced by the mix of trains that utilize it, as different service types have competing operational requirements. While freight services might require extended service times, passenger services demand more predictable schedules. Traditional methods for addressing long-term traffic assignment problems often rely on fixed-value capacity limitations, determined based on specific assumptions about traffic composition. This paper introduces a methodology for determining timetable-independent capacity within the traffic rate assignment problem, enabling the calculation of junction capacities under dynamic traffic distributions. We solve the underlying non-linear constrained optimization problem maximizing the traffic throughput using Bayesian optimization (BO). This setting combines a known objective function with expensive-to-compute capacity constraints, motivating an adaption of standard BO problems, where objective functions are usually unknown. We tailor the acquisition process in BO to this specific setting and increase performance by incorporating prior knowledge about the shape of the constraint functions into the Gaussian process surrogate model. Our derived approaches are benchmarked on a railway junction near Paris, significantly outperforming fixed traffic composition models and highlighting the benefits of dynamic capacity allocation.

Keywords: railway junction, performance determination, Bayesian optimization, continuous time Markov chain

*Corresponding author

Email address: emunds@via.rwth-aachen.de (Tamme Emunds )

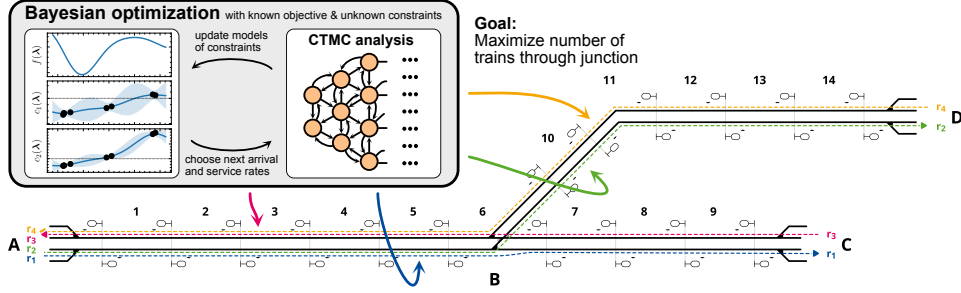


Figure 1: *Overview of the proposed method.* We aim to maximize the number of trains through a junction by using Bayesian optimization to find globally optimal traffic rate assignments respecting computationally expensive to evaluate capacity constraints that are modeled with continuous-time Markov chains.

1. Introduction

The global need for effective transportation modes is accelerating the development and enhancement of railway transportation networks. To effectively allocate available resources, infrastructure managers must assess the performance of all elements within their railway networks. Railway infrastructure is designed to last multiple decades, and infrastructure dimensioning takes place in very early planning stages. Therefore, it is crucial to assess infrastructure performance independent of timetables, which may be fixed only a couple of years in advance. However, the performance capabilities of railway infrastructure depend heavily on traffic distribution patterns, including the allocated rolling stock units and microscopic routes. While recent research [1] introduced a timetable-independent method for evaluating the performance of railway junction infrastructure, its traffic distribution remained fixed for every considered number of trains. Long-term planning problems—such as frequency assignment to traffic types and routes, or optimal line-plan determination—require methods that can accommodate dynamic traffic distributions.

This work introduces an approach for incorporating non-linear timetable capacity constraints into traffic rate assignment problems. With this, we can describe the influence of dynamic traffic rate assignments onto the performance capability of railway infrastructure. We solve the resulting non-linear global optimization problem—characterized by a known objective function with computationally expensive capacity constraints—using Bayesian optimization (BO). We adapt constrained BO algorithms, which assume black-box (i.e., unknown) objective functions and constraints, to our setting where the objective function is known but the capacity constraints require expensive computation and are therefore treated as black-box functions. Furthermore, we embed domain knowledge about the con-

straint function’s shape into the Gaussian process (GP) surrogate models that approximate the capacity constraints, resulting in improved performance. In summary, our main contributions are:

- (i) A novel method for efficiently determining the timetable-independent capacity of railway junctions under dynamic traffic distributions.
- (ii) Adapting classic constrained BO algorithms to the setting of an known objective function and prior knowledge of an exponential trend of black-box constraints.
- (iii) A case study demonstrating the achievable performance improvement of allowing non-static traffic distributions for a given junction infrastructure.

The rest of the paper is organized as follows: In Section 2, we begin with a brief review of related work, followed by background on the methodology in Section 3. Afterwards, Section 4 presents the traffic rate assignment problem formulation. The queueing-based model for the non-linear black-box constraints is described in Section 5, followed by our BO solution approach in Section 6. Section 7 demonstrates the efficacy of our method, and Section 8 concludes the work.

2. Related Work

In order to determine the performance capability of their network, infrastructure managers can choose from a variety of methods. For example, the UIC Code 406 [2] defines the metric of *capacity utilization* for railway lines (and nodes) with an occupancy time rate as the quotient of summed occupancy times in a given timetable and the considered time period. For railway lines and junctions, this can be calculated by compressing the blocking times of all subsequent train runs within the section [2, 3, 4]. By formulating the occupation parameters of a given timetable in Max-Plus-Algebras, capacity utilization may also be computed for railway networks [5, 6]. However, this occupation ratio is strongly dependent on the sequence of train runs and therefore the given timetable. Even though methods to efficiently sample possible timetable requests to achieve independence from the timetable itself exist [7, 8, 9], other capacity metrics have been considered as well. These metrics can be partitioned by their capacity understanding, [7, 1] distinguish between (i) *theoretical capacity*, describing the theoretical maximum number of trains on an infrastructure, satisfying mainly safety and driving dynamic constraints; (ii) *timetable capacity*, corresponding to the maximum number of requests in a timetable for the given infrastructure, additionally considering operating program specific settings and quality thresholds; and (iii) *operational capacity*, the

maximum number of trains with acceptable operational quality, also respecting disturbances and delay propagation.

Theoretical capacity analysis is mainly utilizing mixed integer programming (MIP) formulations, saturating a given timetable with additional train requests [10, 11] or independently allocating occupation slots to requested train journeys either for capacity determination directly [12, 13] or within timetabling problems [14, 15]. However these methods typically consider mainly conflict-freeness and refrain from including quality concerns. In order to assess the operational quality, infrastructure managers might decide to analyze operational data directly [16, 17], utilize simulation approaches [18, 19, 20] or perform analytical analysis of delay propagation [21, 22, 23].

Other analytical approaches focus on determining the timetable capacity of railway lines [24, 25], junctions [26, 27], stations [28, 29], or networks [30, 31], modeling the timetabling process with queueing systems. In particular, we build on the approach from [1] which calculates the timetable capacity of railway junction infrastructure by formulating a continuous-time Markov chain (CTMC) and evaluating it with probabilistic model-checking.

Existing approaches to integrate dynamic traffic distributions into performance estimations for railway infrastructure are either limited to linear approximations of gridwise predetermined capacity results [32], or calculate theoretical capacity, e.g., by directly including timetabling problems into line planning formulations [33], which does not consider timetable quality indicators.

In order to include dynamic traffic distributions, this paper introduces an approach to iteratively enhance the assignment of route frequencies, obtaining queue-length estimations for the selected assignment in each step. This process necessitates numerous calls to the model-checker to evaluate constraints partially described by CTMCs. To limit the amount of required iterations, we leverage BO for an efficient optimization under expensive to evaluate constraints.

BO increased in popularity for optimizing unknown objective function in a data-efficient manner with applications ranging from robotics [34, 35, 36] to drug discovery [37, 38, 39]. In railway systems, BO has been applied to network optimization [40], train delay prediction modelings [41, 42], and interval dynamic analysis [43]. BO has not been applied to optimizing railway junction performance.

A key factor contributing to the popularity of BO is the flexibility of its framework. It has been extended, e.g., to high-dimensional settings [44, 45, 46], time-varying settings [47, 48, 49], as well as optimizing under unknown constraints [50, 51, 52]. For a comprehensive overview, we refer to [53]. Furthermore, the GP surrogate model used in BO can directly incorporate prior knowledge about the underlying problem by, e.g., designing task specific kernels [54] or mean functions. In this paper, we build on BO with unknown constraints and discuss the

specific case where the objective function is known. To our knowledge, this work represents the first explicit treatment of BO scenarios with known objectives but black-box constraints. Furthermore, we demonstrate that performance can be significantly improved by incorporating a problem-specific mean function for modeling the black-box constraints.

3. Background and Preliminaries

This paper combines BO with queuing-based performance estimation methods to obtain solutions for traffic rate assignment problems. In this Section, we recall the performance metric *timetable capacity* (Section 3.1) along with its associated quality thresholds (Section 3.2), which will be used to formulate the constraints on our optimization problem. Additionally, we revisit fundamentals of GPs (Section 3.3) and BO (Section 3.4). All notation is summarized in the [Appendix A](#) in Table A.4 to facilitate parsing and to serve as a reference throughout the paper.

3.1. Timetable Capacity of Railway Systems

Successful railway transportation of goods and passengers depends on critical long-term planning processes that occur several years before operations begin. To ensure adequate design of costly infrastructure, evaluation must be conducted independently of specific timetables, which may change throughout the infrastructure’s multi-decade lifespan.

This work employs the performance metric of *timetable capacity* (see also [24, 1]). It corresponds to the capacity of the railway infrastructure in the timetabling process, i.e., the assignment of conflict-free occupation slots to requested train journeys. Since multiple train journeys might request the occupation of conflicting infrastructure routes at the same time, the infrastructure manager needs to reschedule some requests to later occupation times. Given a route r , the number of rescheduled requests awaiting their corresponding occupation slot can be represented as a queue for any point in time. Note that the average queue length of occupation requests depends on their stochastic arrival and service processes. In this work, we utilize continuous-time Markov chains to obtain an approximation of the expected queue length L_r for every route r (see Section 5.2). These average queue lengths on all routes can be utilized as performance metrics for the infrastructure of the railway junction by comparing them to thresholds for adequate quality.

3.2. Thresholds on Queue Lengths

To qualitatively evaluate performance capabilities, different thresholds have been established for corresponding capacity metrics. For example, maximum occupation rates have been defined [2] for the use with timetable compression methods. Following the methods of the largest European infrastructure manager, DB

InfraGO [55], we use a threshold for the expected average queue length. Developed in a survey [56] under railway dispatchers to evaluate timetable quality on a railway line, the limit on the average queue length is defined as

$$L_{\text{limit}, r} = 0.479 \cdot \exp(-1.3 \cdot p_{\text{pt}, r}), \quad (1)$$

in dependence of the proportion of passenger traffic $p_{\text{pt}, r}$. Note that this threshold is defined for railway lines and we hence apply it independently for every route within a railway junction.

3.3. Basics on Gaussian Processes

We will use GPs [57] to model the the black-box constraints formulated in the previous subsection. A GP is fully defined by a mean function $m: \Lambda \rightarrow \mathbb{R}$ and a kernel function $k: \Lambda \times \Lambda \rightarrow \mathbb{R}$ and we will denote it as $\mathcal{GP}(m(\lambda), k(\lambda, \lambda'))$. This prior GP can be conditioned on a dataset $\mathcal{D}_t := \{(\lambda_i, y_i)\}_{i=1}^{t-1}$ and evaluated on a test point λ yielding a normal predictive distribution $\mathcal{N}(m_{\mathcal{D}_t}(\lambda), \sigma_{\mathcal{D}_t}^2(\lambda))$ with mean and variance as

$$m_{\mathcal{D}_t}(\lambda) = m(\lambda) - \mathbf{k}_t(\lambda)^\top (\mathbf{K}_t + \sigma_n^2 \mathbf{I})^{-1} (\mathbf{y}_t - m(\lambda)) \quad (2)$$

$$\sigma_{\mathcal{D}_t}^2(\lambda) = k(\lambda, \lambda) - \mathbf{k}_t(\lambda)^\top (\mathbf{K}_t + \sigma_n^2 \mathbf{I})^{-1} \mathbf{k}_t(\lambda). \quad (3)$$

Here, $\mathbf{K}_t = [k(\lambda_i, \lambda_j)]_{i,j=1}^{t-1}$ is the Gram matrix, $\mathbf{k}_t(\lambda) = [k(\lambda_i, \lambda)]_{i=1}^{t-1}$ is a vector, and $\mathbf{y}_t = [y_1, \dots, y_{t-1}]^\top$ are noisy measurements. A key advantage of GPs is that their mean and kernel functions can encode prior knowledge about the unknown function's behavior, which we leverage to enhance constraint expressiveness and overall performance.

3.4. Overview on Bayesian Optimization

Because they can yield well-calibrated uncertainty estimates, GPs are widely adopted as surrogate models within BO [53]. As a sample-efficient global optimization method for black-box objective functions, BO is especially suited for problems where function evaluations are expensive. In its standard form, BO aims to find $\lambda^* = \arg \max_{\lambda \in \Lambda} f(\lambda)$ by sequentially querying the unknown objective $f(\lambda)$ and receiving zeroth-order feedback, i.e., only the function value and not its gradient. Here, $\Lambda \subseteq \mathbb{R}^d$ is the feasible set where d is the dimension of decision variable. As stated above, typically, $f(\lambda)$ is modeled as a GP and the next query point λ_t at each iteration $t \in \mathcal{I}_T := \{1, \dots, T\}$ is selected by optimizing an acquisition function $\alpha: \Lambda \rightarrow \mathbb{R}$ as

$$\lambda_{t+1} \leftarrow \arg \max_{\lambda \in \Lambda} \alpha(\lambda \mid \mathcal{D}_t). \quad (4)$$

This acquisition function usually balances exploration (sampling in uncertain regions) and exploitation (sampling near well-performing regions) to efficiently converge to the optimum. There exist various choices for an acquisition function such as upper confidence bound [58, 59], expected improvement [60], and Thompson sampling [61]. In this paper, we will consider variants of the latter two acquisition functions but we will tailor them to our specific setting of a known objective function under black-box constraints.

4. Problem Definition

One major long-term planning task is to assign frequencies to paths within a railway network, in order to best satisfy a given demand of traffic for origin-destinations combinations. This problem can be formulated for large-scale transportation networks and even include considerations regarding different vehicle-sizes or passenger connections between the various lines. The focus of this work, however, is not to solve such a *line-planning* problem for a complex railway network, but rather on introducing a novel method to model capacity restrictions for railway junctions, connecting the lines between stations within such networks. This allows for assessing infrastructure performance indicators with non-static traffic distributions, enhancing the level of flexibility included into applications for long-term planning problems.

We model a railway junction $J = (\mathcal{R}, C)$ with a set of k routes $\mathcal{R}$ and a *conflict matrix* $C \in \{0, 1\}^{k \times k}$, describing whether the parallel occupation of two routes r, r' is permitted ($C_{r,r'} = 0$). One major influence to the occupation time of railway infrastructure is the rolling stock, we therefore notate an occupation request $o \in \mathcal{O}$ as a tuple $o = (r, u) \in \mathcal{R} \times \mathcal{U}$ of a selected route r and a corresponding rolling stock unit (train type) $u \in \mathcal{U}$. We sometimes refer to $C_{o,o'} := C_{r,r'}$ to model conflicts for a combination of requests $(o, o') = ((r, u), (r', u'))$.

In this work, we measure the timetable capacity of a railway junction within a fixed time horizon t_U . The number (or *frequency*) f_o of trains per request $o \in \mathcal{O}$ can be modeled as an *arrival rate*

$$\lambda_o = \frac{f_o}{t_U}, \quad (5)$$

describing the number of trains of request o , arriving at the modeled junction in the time horizon t_U . The number of different requests is denoted with $d = |\mathcal{O}|$, describing the dimension of the assignment vector $(\lambda_o)_{o \in \mathcal{O}} = \lambda \in \mathbb{R}^d$.

For a given upper bound function $\text{ub} : \mathcal{O} \rightarrow \mathbb{R}, o \mapsto \text{ub}_o$, we can formulate the

constraints

$$\lambda_o \leq \text{ub}_o, \quad \forall o \in \mathcal{O}, \quad (6)$$

$$c_r(\lambda) \leq 0, \quad \forall r \in \mathcal{R}, \quad (7)$$

describing the feasible set of λ . In this work, the *capacity restriction* functions $c_r : \mathbb{R}^d \rightarrow \mathbb{R}$ are utilized to include limitations arising from the timetable-independent capacity metric (see Section 3.1) computed via queuing systems for the railway junction infrastructure.¹

When determining the performance capability of a railway junction, the intuitive objective function $\max_{\lambda} \sum_o \lambda_o$, might give the theoretical maximum of traffic rate assignments. However, the resulting rates might not represent an optimal assignment in practice, since neither the traffic nor the capacity consumption are distributed equally between all combinations. Therefore, it is necessary to extend the objective function with terms representing desired traffic parameters.

We utilize two different approaches to tackle the practicality of obtained solutions in this work. In detail, the deviation from two different distribution targets, i.e., the spreading $\tilde{p}_u$ (or $\tilde{p}_r$) between train types $u \in \mathcal{U}$ or routes $r \in \mathcal{R}$, are discussed. Naturally, a chosen traffic rate assignment λ also defines the corresponding distributions between train types $p_u(\lambda)$ and routes $p_r(\lambda)$. An objective formula of the form

$$\max_{\lambda} \sum_{o \in \mathcal{O}} \lambda_o - w_U \cdot \sum_{u \in \mathcal{U}} (p_u(\lambda) - \tilde{p}_u)^2, \quad (8)$$

can be utilized to enforce adherence to a given train type distribution $(\tilde{p}_u)_{u \in \mathcal{U}}$. The objective (8) incorporates a penalty, corresponding to the summed squared differences $\sum_{u \in \mathcal{U}} (p_u - \tilde{p}_u)^2$, into the objective value by weighting it with a weighting factor $w_U \in \mathbb{R}$. An analogous objective formula can be developed for distributions across other traffic contexts, such as the different train routes, or combinations of these parameters. With this, we can formulate the *Penalized Traffic Rate Assignment Problem (PTRAP)* that we aim to solve in this paper.

¹Note that the core ideas of this paper are not restricted to employing a queuing-based capacity model. However, we will tailor parts of our method in Section 6) to this modeling approach and exploit the exponential behavior of the capacity constraints for improved performance.

$$\lambda^* = \arg \max_{\lambda} \sum_{o \in \mathcal{O}} \lambda_o - w_U \cdot \sum_{u \in \mathcal{U}} (p_u(\lambda) - \tilde{p}_u)^2 \quad (\text{a})$$

$$\text{subject to } \lambda_o \leq \text{ub}_o \quad \forall o \in \mathcal{O} \quad (\text{b})$$

$$c_r(\lambda) \leq 0 \quad \forall r \in \mathcal{R} \quad (\text{c})$$

5. Queuing-Based Capacity Restrictions

After deriving **PTRAP**, the remaining step is to formalize the capacity restriction functions $c_r(\lambda)$ in **(PTRAP-c)**. For this, Section 5.1 introduces a model of queueing systems for railway infrastructure dimensioning along with the technique to obtain queue-length estimations for railway junctions in Section 5.2. In a final step in Section 5.3, we obtain constraint formulations for the traffic rate assignment problem **PTRAP** considered in this work.

5.1. Queueing Systems for Railway Performance Measures

The timetable capacity of a railway junction $J = (\mathcal{R}, C)$ not only depends on the layout of the infrastructure $\mathcal{R}, C$ and the utilized rolling stock units $\mathcal{U}$, but additionally on the distribution of traffic to different routes. A railway junction might be traversed with heterogeneous traffic types, such as passenger and freight traffic, all of which utilizing different rolling stock units. Therefore, the occupation time $b : \mathcal{O} \rightarrow \mathbb{R}, o \mapsto b_o$ of a request $o = (r, u)$ depends on its route r and rolling stock unit u .

Furthermore, the occupation time of a request within a railway junction is not only influenced by the train itself, but additionally by the following train. In railway operations research, the occupation time is therefore usually modeled as a minimum headway time $h : \mathcal{O} \times \mathcal{O} \rightarrow \mathbb{R}, (o, o') \mapsto h_{o,o'}$ for a sequence of trains (o, o') , describing the minimum time-span between the start of the occupation of the first train o , until the start of the occupation of the following train o' . This minimum headway time is calculated once for every possible request combination on microscopic infrastructure models, more details regarding the blocking-time model and the computation of minimum headway times can, e.g., be found in [62].

For timetable-independent planning problems, the sequence of trains is not predetermined. Hence, the occupation time b_o needs to be formulated without knowledge of following requests. We therefore assume a uniform distribution of request

sequences and describe the average occupation time

$$b_o = \frac{\sum_{\substack{o' \in O \\ C_{o,o'}=1}} \lambda_{o'} \cdot h_{o,o'}}{\sum_{\substack{o' \in O \\ C_{o,o'}=1}} \lambda_{o'}} \quad (9)$$

of a single request o by weighting the minimum headway times $h_{o,o'}$ of conflicting request combinations o, o' with their arrival rates.

In this work, we model the timetabling process on the railway junction as a queuing system (see Section 5.2). Within this queueing system, traffic on the junction $J = (\mathcal{R}, C)$ is decomposed along the routes $r \in \mathcal{R}$, and performance indicators are calculated for every route. Hence, we need to abstract the occupation times from the request based formula in (9) to the average occupation time

$$b_r = \frac{\sum_{o=(r,u) \in O} \lambda_o \cdot b_o}{\lambda_r} \quad (10)$$

for a route r by weighting the request occupations times b_o for all requests $o = (r, u) \in O$ on the route r with their respective arrival rates. In (10), $\lambda_r = \sum_{(r,u) \in O} \lambda_{(r,u)}$ defines the total arrival rate to a route r . The with (10) obtained route-based occupation times b_r allow us to define the *service rate* $\mu_r = 1/b_r$ for a route r , giving the average rate with which the requests are serviced within the queuing system. These service rates μ_r can be compared to the arrival rates λ_r , in the *occupation ratio* $\rho_r = \lambda_r/\mu_r$, serving as an indicator of the utilization of a route r .

In the modeled queuing system, arrived but not yet started requests are gathered in an individual queue for every route. The expected number of request in the queue of route r , denoted as the *expected queue-length* L_r , gives another performance indicator of the queuing system. The next Section will describe the utilized approach to calculating these L_r from obtained arrival and service rates.

5.2. Obtaining Queue-Length Estimations

The queueing system for the railway junction with arrival and service rates derived in Section 5.1 can be formulated as a CTMC with state space S , see [1] for full details.

To compute the expected queue-length L_r for each route r in the modeled railway junction, the stationary distribution of the CTMC can be analyzed. The stationary distribution is a probability distribution over the state set S , i.e., it maps a probability $p(u)$ to each state $u \in S$, indicating the likelihood that the system will be in the corresponding state in the long run.

Such a steady state analysis has been used by a number of analytical performance determination approaches, see f.e. [24, 26, 25]. In this work, we build on an

approach introduced in [1] which builds the CTMC in the formal PRISM language [63] and obtains the queue-length estimations L_r with probabilistic model-checking [64]. This approach relies on an approximation of the queueing system by limiting the number of waiting positions in the queue of each route r with a finite number $B \in \mathbb{N}$. We use values of $B = 3$ and $B = 5$ for the computations in Section 7, resulting in the loss of every request arriving to a route with three (five) waiting requests, therefore under-approximating the theoretical average queue-length. Further influences of this parameter B are discussed in [1].

Note that due to the Markov property of the CTMC, resulting estimations are obtained for exponentially distributed inter-arrival and service times only (M/M systems [65]). To adapt this system to other probability distributions (GI/GI systems [65]), we can utilize the formula of Hertel [66], where $L_r(M/M) \cdot \frac{1}{\gamma} \approx L_r(GI/GI)$ with parameters $\gamma = \frac{2}{c \cdot v_S^2 + v_A^2}$ and $c = \left(\frac{\rho_r}{s}\right)^{1-v_A^2} \cdot (1 + v_A^2) - v_A^2$, and scale the queue-lengths L_r accordingly, using coefficients of variation for the service and arrival process. In this work, we scale the queue-length with a variation coefficient of the arrival of $v_A = 0.8$ and of the service process of $v_S = 0.3$, according to standard values. The next section highlights a formulation of the capacity constraints by using the obtained queue-length estimates.

5.3. Capacity Constraint Formulation

After obtaining the estimated queue-lengths L_r for every route, we can formulate the constraints for the traffic rate assignment problem from Section 4. For this, constraint (PTRAP-c) can be updated with the comparison of the in Section 5.2 estimated queue-length $L_r = L_r(\lambda)$ for a given rate assignment λ with the threshold value $L_{\text{limit}, r}$ from Section 3.2 to the constraint

$$\text{(PTRAP-c): } c_r(\lambda) = L_r(\lambda) - L_{\text{limit}, r} \leq 0, \forall r \in \mathcal{R}, \quad (11)$$

for every route r .

In practice, estimations of $L_r(\lambda)$ can be expensive to compute and are nonlinear. To still efficiently solve PTRAP, we approximate the constraints (PTRAP-c) with GPs and leverage BO for an informed search.

6. Utilizing Bayesian Optimization

We next tailor standard BO approaches to our optimization problem at hand, i.e., a known objective function and black-box constraints. This differs from the standard BO formulation where the objective is usually unknown. In this section, we will include the additional information about the objective and its unconstrained optimum into the design of our method. We begin by describing how to model the black-box constraints.

6.1. Modeling Unknown Constraints on the Capacity

Since the constraints $c_r(\lambda)$ are expensive to evaluate, we treat them as unknown and model them with a GP.

6.1.1. Kernel function

In this work, we will choose as the kernel function a Matérn kernel [57] with $\nu = 5/2$ and individual lengthscales ℓ_i , scaling the correlation between inputs for all dimensions, as

$$k(\lambda, \lambda') = \sigma_f^2 \left(1 + \sqrt{5}\Delta + \frac{5}{3}\Delta^2 \right) \exp(-\sqrt{5}\Delta), \quad (12)$$

where σ_f^2 is the output scale, and $\Delta := \sqrt{\sum_{o=1}^d [(\lambda_o - \lambda'_o)^2 / \ell_o^2]}$ is the normalized Euclidean distance.

6.1.2. Mean function

For the mean function, we include intuition on the shape of the constraints. For this, we visualize the constraint for a one-dimensional problem on a dense grid as shown in Fig. 2.

Observing an exponential trend in the constraint c_r , we leverage the fact that embedding prior knowledge into the GP model is possible and formulate the following mean function for our GP prior of c_r as

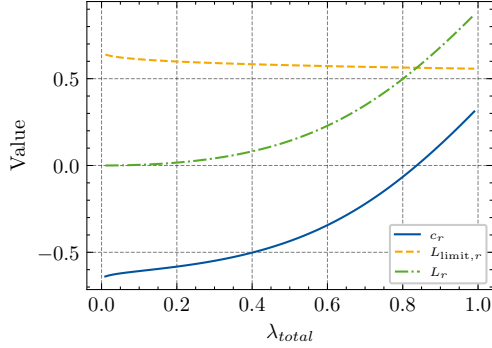


Figure 2: Exemplary plot of a constraint c_r and its components $L_r, L_{r,\text{limit}}$ for different traffic rates λ_{total} .

$$m_{\text{exp}}(\lambda) = \beta \exp \left(\sum_{i=1}^d w_i \lambda_i \right) - \gamma. \quad (13)$$

Here, β , w_i , and γ are learnable parameters. To ensure the observed positive trend, we enforce positivity on β and w_i by re-parameterizing them as

$$\beta = \exp(\beta_r) \text{ and } w_i = \exp(w_{i,r}) \quad \forall i \in \mathcal{I}_d \quad (14)$$

where $\beta_r \in \mathbb{R}$ and $w_{i,r} \in \mathbb{R}$ are the actual learnable parameters in an unconstrained space. In our experiments, we will also compare this to the standard approach of using a constant mean function. Importantly, the formulation in (13) defines

an abstraction that naturally encompasses the constant mean function as a special case. Specifically, when $\beta \rightarrow 0$ or $w_i \rightarrow 0$ for all i , the mean function simplifies to a constant, making it a subset of our proposed mean’s solution space. This flexibility enables the model to recover a simple constant offset when exponential patterns are absent, while capturing exponential trends when they are present in the data. To summarize, the set of learnable parameters for each constraint is $\Theta_r = \{\sigma_n, \ell_0, \dots, \ell_d, \sigma, \beta_r, w_0, \dots, w_d, \gamma_r\} \in \mathbb{R}^{2d+4}$. We will learn these parameters at each iteration given all data set by maximizing the marginal log-likelihood [57].

6.2. Choosing the Next Query with Known Objectives under Unknown Constraints

We next discuss how to design an acquisition function for our specific problem formulation. Here, we will evaluate two approaches: *i*) reformulating constrained expected improvement to a known objective and *ii*) using Thompson sampling.

6.2.1. Adaptive Trust Regions

For both approaches, we will constrain the feasible set to a trust region as in the algorithm TuRBO proposed in [45]. TuRBO maintains a trust region $\mathcal{TR}_t$ as the hyperbox around parameters of the best observed value $\hat{\lambda}_t^*$ based on the lengthscales of the GP surrogate of the objective. Crucially, in our setting, the objective is known. Therefore, we also omit the scaling and define the trust region

$$\mathcal{TR}_t = \left\{ \lambda \in \Lambda \mid |\lambda_j - \hat{\lambda}_{t,j}^*| \leq \frac{L_t}{2}, \forall i \in \mathcal{I}_d \right\}. \quad (15)$$

Here, L_t is the hyperbox length at the current iteration. We adapt this parameter over time in the same fashion as in [45], i.e., we shrink it if we have not improved in the past few iterations and expand it if improved in successive iterations. We use the same hyperparameter as [45] to which we refer to for further details.

6.2.2. Adapting Constrained Expected Improvement to Known Objective

With knowledge of the objective function, we no-longer have to consider the expected improvement (EI) of a next query compared to the previous one as in the standard formulation, as the expectation is taken with respect to a fixed objective. Additionally, the improvement over the best query is merely a shift in the function values and does not impact its optimizer. Therefore, we can directly reformulate the classic EI acquisition function as

$$\alpha_{\text{CI}}(\lambda) := f(\lambda) \cdot \mathbb{P}(c_r(\lambda) \leq 0) \quad (16)$$

where the feasibility probability is given by

$$\mathbb{P}(c_r(\lambda) \leq 0) = \prod_{r \in \mathcal{R}} \Phi\left(\frac{-m_r(\lambda)}{\sigma_r(\lambda)}\right), \quad (17)$$

and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.

To ensure numerical stability, we follow advise from [67] and instead of optimizing (16), we optimize a logarithmic version for numerical stability as

$$\lambda_{t+1} \leftarrow \arg \max_{\lambda \in \mathcal{TR}_t} \log f(\lambda) + \sum_{r \in \mathcal{R}} \log \mathbb{P}(c_r(\lambda) \leq 0). \quad (18)$$

Interestingly, this formulation closely resembles a log-barrier approach commonly used in classical numerical optimization [68]. In traditional log-barrier methods, constraints are enforced by introducing penalty terms of the form $-\sum_j \log(-c_r(\lambda))$. As the optimization progresses, these barrier terms encourage feasible solutions while smoothly guiding the optimizer towards the constraint boundary. Similarly, as our model of the constraint improves, the acquisition function will increasingly favor points closer to the active constraints.

In Fig 3, we show a small example for the applicability of our revised mean function in combination with an adaptive trust region and our EI formulation. We can observe that over time, the algorithm will converge to a feasible solution respecting our unknown constraints.

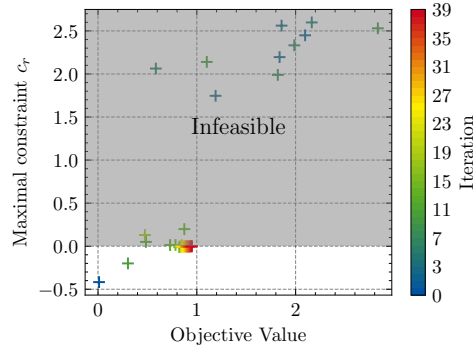


Figure 3: Iterations of EI-Exp-TR (see Section 7.1) and their corresponding maximal constraint $\max c_r$ and objective value.

6.2.3. Thompson Sampling in Constrained Bayesian Optimization

Besides an EI-like acquisition function, we also apply constrained Thompson sampling (TS) to solve the **PTRAP**. The core idea of TS is to choose the next query location as the global optimizer of a posterior sample from the surrogate model. Here, we directly follow a constrained version of TS proposed in [52] but adapt it to the setting of a known objective function. For this, we select M candidates $\mathcal{M} = \{\lambda^{(0)}, \dots, \lambda^{(M)}\} \subseteq \mathcal{TR}_t$ from a pseudo random Sobol sequence from the current trust region. We next sample realizations from the posteriors distributions of all the constraints and select the query location for the next iteration as

$$\lambda_{t+1} \leftarrow \begin{cases} \arg \min_{\lambda \in \mathcal{F}} f(\lambda), & \text{if } \mathcal{F} \neq \emptyset, \\ \arg \min_{1 \leq j \leq M} \sum_{r \in \mathcal{R}} \max(0, c_r(\lambda^{(j)})), & \text{otherwise,} \end{cases} \quad (19)$$

where $\mathcal{F} = \{\lambda \in \mathcal{M} \mid c_r(\lambda) \leq 0, \forall r \in \mathcal{R}\}$ denotes the set of feasible points among the M candidates.

6.3. Recenter Trust Region on Probable Optimum

In addition to the conventional TuRBO framework, which iteratively establishes a trust region centered around the highest previously known objective value to bound the optimization area for the subsequent optimization, we propose an alternative method for determining the trust region. With this approach, we aim to leverage the availability of an analytical description of the objective function.

In detail, the analytical description of the objective function (**Ptrap-a**) as a composition of differentiable functions allows us to determine the gradient analytically. Under the assumption that constraints (**Ptrap-b**) and (**Ptrap-c**) are satisfied within a region $\mathcal{F}_\Lambda \subseteq \mathbb{R}^d$, we can determine a local maximum $\lambda_t^* = \arg \max_{\lambda \in \mathcal{F}_\Lambda} (\text{Ptrap-a})$ using non-linear optimization methods such as sequential least-squares programming [69].

However, with the probabilistic representation of the constraints, defining a feasibility region $\mathcal{F}_\Lambda$ involves managing uncertainties. Specifically, we aim to solve the non-linear chance constrained optimization problem

$$\begin{aligned} \lambda_t^* &= \arg \max_{\lambda \in \Lambda} (\text{Ptrap-a}) \\ \text{subject to } &(\text{Ptrap-b}) \\ &\mathbb{P}\{c_r(\lambda) \leq 0\} \leq \alpha_r, \quad \forall r \in \mathcal{R}. \end{aligned} \tag{20}$$

Given that $c_r(\lambda_o)$ is Gaussian distributed, we can state that

$$\mathbb{P}\{c_r(\lambda) \leq 0\} \leq \alpha_r \Leftrightarrow m_{\mathcal{D}_t, r}(\lambda) + \Phi^{-1}(\alpha_r) \sigma_{\mathcal{D}_t, r}(\lambda) \leq 0 \tag{21}$$

with $\Phi^{-1}(\cdot)$ as the inverse CDF which again yields a tractable optimization problem. We utilize its optimal solution λ_t^* as the center of a new optimization region $\mathcal{TR}'_t \subseteq \mathbb{R}^d$ to be used with the acquisition functions discussed in Section 6.2. By formulating (20) as a chance constrained problem, we ensure feasibility of the center λ_t^* of $\mathcal{TR}'_t$ with high probability.²

In this work, we compute the gradient with SymPy [70], and use SLSQP [69] as implemented in SciPy [71] to find the maximizer λ_t^* . Subsequently, the next optimization region $\mathcal{TR}'_t \subseteq \mathbb{R}^d$ is defined as a hypercube centered around λ_t^* , with an iteratively adjusted edge length, similar to the approach used in the TuRBO framework (cf. Section 6.2.1).

²In all following example, we directly choose $\Phi^{-1}(\alpha_r) = 2$ which roughly corresponds to $\alpha_r \approx 0.975$, i.e., 97.5% satisfaction probability.

7. Results

We next benchmark our approaches. First, we evaluate different combinations of our three main proposed adaptations on a small exemplary railway network (Section 7.1). These adaptations comprise: (i) the inclusion of prior knowledge in the GP model, (ii) different acquisition functions, and (iii) the determination of the trust region center. Afterwards (Section 7.2), selected methods are applied in a case study to maximize the traffic frequency on a realistic railway junction with eight possible routes. All the following baselines are implemented in BoTorch [72] and to evaluate the CTMCs for all constraints we use [64].

7.1. Ablation and Model Selection

In order to compare the different methods, we consider a small exemplary network, consisting of three stations, connected via double-track lines (Fig. 1). The routes from **A** to **C** (r_1, r_3) and from **A** to **D** (r_2, r_4) separate into two double-track railway lines in the railway junction **B**.

Formulating the queuing process of this junction necessitates the knowledge of minimum headway times in order to estimate the service rates for every route. Usually, infrastructure managers would obtain them by applying some microscopic calculation tool based on infrastructure layout and driving dynamics (see e.g., [62]). For this artificial example, however, we assume minimum headway times $h_{o,o'}$ as in Table 1, with first trains o in rows and following trains o' in columns for the infrastructure in Fig. 1. Trains $(r_i, u) \in \mathcal{O}$ are indicated as ' r_i-u ', using the abbreviations 'fr' for freight, 'ld' for long-distance and 'lo' for local trains to represent the different rolling stock units $u \in \mathcal{U}$.

In detail, we compare eight variants that differ across three key components: (i) the acquisition function used, (ii) the underlying mean function of the GP for capacity constraints, and (iii) the trust region definition approach—either the conventional TuRBO framework (TR) or our proposed adaptation with re-centering to the probable optimum (CP0). Table 2 gives an overview regarding the different methods, their characteristics and their respective names.

In this example, we utilize the objective function from (8) for a given target parameter

$$\tilde{p}_u = \begin{cases} 0.5 & , u = \text{local train} \\ 0.3 & , u = \text{freight train} \\ 0.2 & , u = \text{long distance train} \end{cases} \quad (22)$$

of the traffic type distribution across the junction.

We conducted 20 experiments for all eight variants, using a weight factor of $w_U = 5$. All computations were done on eight cores of an Intel Xeon Platinum

Table 1: Minimum Headway Times in Minutes for the Junction Infrastructure in Fig. 1

		r_1			r_2			r_3			r_4		
		-fr	-ld	-lo	-fr	-ld	-lo	-fr	-ld	-lo	-fr	-ld	-lo
r_1	-fr	5.0	5.0	5.0	0.0	0.0	0.0	5.0	5.0	5.0	0.0	0.0	0.0
	-ld	2.0	2.0	2.0	0.0	0.0	0.0	2.0	2.0	2.0	0.0	0.0	0.0
	-lo	3.0	4.0	3.0	0.0	0.0	0.0	3.0	4.0	3.0	0.0	0.0	0.0
r_2	-fr	0.0	0.0	0.0	5.0	5.0	5.0	1.7	1.7	1.7	5.0	5.0	5.0
	-ld	0.0	0.0	0.0	2.0	2.0	2.0	1.0	1.0	1.0	2.0	2.0	2.0
	-lo	0.0	0.0	0.0	3.0	4.0	3.0	1.0	1.3	1.0	3.0	4.0	3.0
r_4	-fr	5.0	5.0	5.0	1.7	1.7	1.7	5.0	5.0	5.0	0.0	0.0	0.0
	-ld	2.0	2.0	2.0	1.0	1.0	1.0	2.0	2.0	2.0	0.0	0.0	0.0
	-lo	3.0	4.0	3.0	1.0	1.3	1.0	3.0	4.0	3.0	0.0	0.0	0.0
r_4	-fr	0.0	0.0	0.0	5.0	5.0	5.0	0.0	0.0	0.0	5.0	5.0	5.0
	-ld	0.0	0.0	0.0	2.0	2.0	2.0	0.0	0.0	0.0	2.0	2.0	2.0
	-lo	0.0	0.0	0.0	3.0	4.0	3.0	0.0	0.0	0.0	3.0	4.0	3.0

8468 Sapphire Processor (2.1 GHz), utilizing 19.84 GB of working memory. The termination criteria included a time limit of 60 min or a maximum of 200 iterations.

Fig. 4 exemplarily illustrate the resulting objective function values for all variants. The results in Fig. 4a show the mean value and standard deviation (mean $\pm \sigma$) of the best objective value at each iteration, averaged across 20 experiments per variant. Initially, variants EI-C-TR and EI-C-CP0 exhibit good performance during the first 20 iterations, achieving higher mean objective values. However, as the iterations progress, EI-Exp-TR and EI-Exp-CP0 demonstrate improved efficacy and the best performance. After approximately 50 iterations, methods TS-Exp-TR, TS-C-CP0, and TS-C-TR converge to similar mean objective values as those achieved by EI-Exp-TR and EI-Exp-CP0. Upon termination of all methods (Fig. 4c), EI-Exp-CP0, TS-Exp-TR, and TS-C-TR achieve the highest mean objective values. It is noteworthy that three TS methods—TS-Exp-TR, TS-C-CP0, and TS-C-TR—achieve their results more rapidly concerning actual computing time (Fig. 4b) compared to the other methods. This efficiency can be attributed to the reduced method overhead, resulting in faster iteration times. For this exemplary junction, the model-checking required in each iteration takes approximately 0.2 s to determine the capacity constraints (Section 5.3) for the small railway junction analyzed. For more complex infrastructures, model-checking time may increase significantly, potentially diminishing the impact of method overhead on the overall computation time. The following case study introduces a more complex railway

Table 2: Proposed Variants and Their Defining Properties

Name	Acquisition Function	GP Mean Function	Trust Region Method
EI-C-TR	EI	Const.	Conv. TuRBO
EI-C-CP0	EI	Const.	Centered prob. Opti.
EI-Exp-TR	EI	Exp	Conv. TuRBO
EI-Exp-CP0	EI	Exp	Centered prob. Opti.
TS-C-TR	TS	Const.	Conv. TuRBO
TS-C-CP0	TS	Const.	Centered prob. Opti.
TS-Exp-TR	TS	Exp	Conv. TuRBO
TS-Exp-CP0	TS	Exp	Centered prob. Opti.

junction infrastructure requiring more time for the formulation of capacity constraints in every iteration.

7.2. Case Study on a Realistic Junction

In this Section, we study the performance of a railway junction, referred to the *Triangle of Gagny*, located in the proximity of Paris, France.

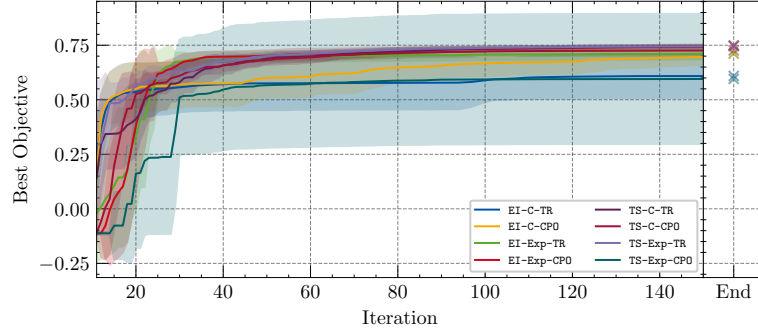
This junction (Fig. 5) features eight routes between two double-track railway lines, **A-B** and **C-D**. Similarly to the previous example, we consider straight routes as main routes and therefore have $\mathcal{R}_{\text{main}} = \{r_3, r_4, r_5, r_8\}$.

This infrastructure has already been analyzed regarding rescheduling algorithms [73] and static timetable capacity [1]. Despite this, microscopic data is not openly available, we therefore assume the minimum headway times in Table 3.

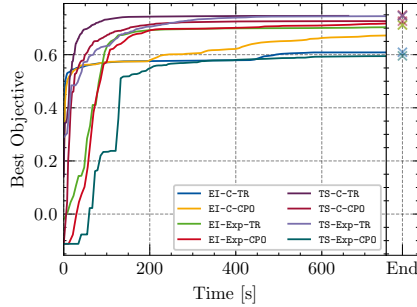
Table 3: Minimum headway times in minutes for the Triangle of Gagny Junction (Fig. 5).

	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8
r_1	3.0	3.0	2.3	0.0	2.2	0.0	0.0	0.0
r_2	3.0	3.0	2.3	0.0	1.5	0.0	1.5	2.2
r_3	1.5	1.5	1.5	0.0	0.0	0.0	0.0	0.0
r_4	0.0	0.0	0.0	1.5	0.0	1.5	1.5	0.0
r_5	1.8	1.5	0.0	0.0	1.8	1.8	1.5	0.0
r_6	0.0	0.0	0.0	2.7	2.7	2.7	2.7	0.0
r_7	0.0	1.5	0.0	3.0	1.5	2.7	3.0	3.0
r_8	0.0	1.8	0.0	0.0	0.0	0.0	1.8	1.8

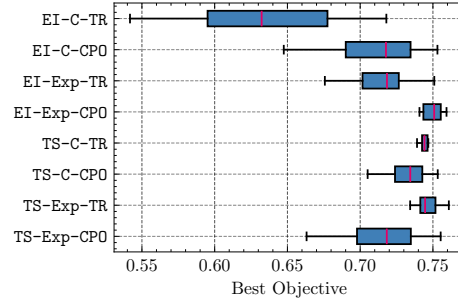
Note that we restrict the case study to one type of trains, i.e., local passenger trains. Due to the route decomposition approach, this does not significantly reduce the required effort for the queue-length determination method from Section 5.2.



(a) Objective per Iteration



(b) Objective per Time



(c) Objective at Termination

Figure 4: Objective distribution between all methods and experiments for $w_U = 5$.

Since this junction features eight routes, the CTMC for queueing-length estimations is more complex and requires substantially more computation time (cf. [1] for a more detailed scaling analysis). We therefore restrict the number of waiting positions to $B = 3$ and the comparison to three good performing methods from Section 7.1 only: EI-Exp-TR, EI-Exp-CP0 and TS-C-TR. Furthermore, a time limit of 60 minutes has been set, in order to test the real-world applicability of the methods. For each of the three methods, 20 experiments have been computed for every selected weighting parameter w_R on six cores of an Intel Xeon Platinum 8468 Sapphire Processor (2.1 GHz), utilizing 30.35 GB of working memory.

Regarding the objective function, we assume that an infrastructure manager may be particularly interested in introducing a target parameter for the route distribution. We assume this target distribution to be

$$\tilde{p}_r = \begin{cases} 1/6 & , r \in \mathcal{R}_{\text{main}} \\ 1/12 & , \text{otherwise} \end{cases} \quad (23)$$

which doubles the traffic on main routes. The objective function has been formu-

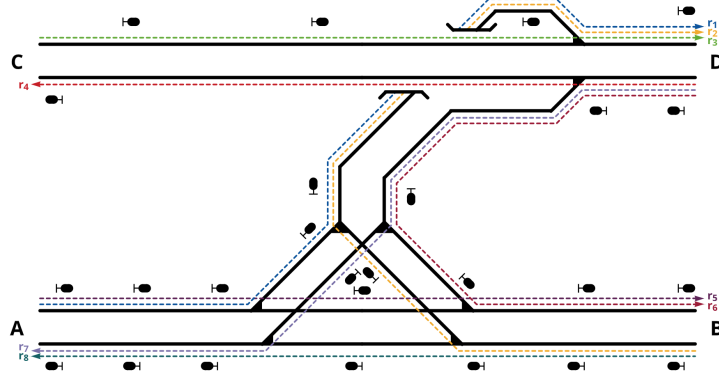


Figure 5: Infrastructure layout of the Triangle of Gagny. Adapted from [73, 1]

lated as

$$(\text{PTRAP-a}): \max_{\lambda \in \Lambda} \sum_{o \in O} \lambda_o - w_R \sum_{r \in R} (p_r - \tilde{p}_r)^2. \quad (24)$$

for different values of the weighting parameter $w_R \in \{0, 1, 2, 5, 10, 20, 50, 100\}$.

Within the one hour time limit, the methods were able to perform between 30 and 70 iterations. On average, the algorithms were able to perform approximately 45 iterations, while the first 10 iterations were used as an initial training budget and their respective query points λ were randomly selected from a pseudo-random Sobol sequence.

A comparison of the performance over iterations of the three selected variants is shown in Fig. 6a for a weighting factor of $w_R = 5$. In comparison with the results from Section 7.1, the variant EI-Exp-TR utilizing the base TuRBO framework outperforms the other two methods clearly in early stages. However, EI-Exp-CP0 is able to retrieve similar results at the end of the computation period—only TS-C-TR shows significant worse performance demonstrating that the additional prior knowledge in the GP model also here helps to significantly improve performance. Note that these results are exemplary for all other experiments with different w_R values.

In order to compare the total traffic sum across methods, Fig. 6b displays the median and confidence intervals of the traffic volume for all selected weights. Furthermore, the timetable capacity of $n_{\max} \approx 41.92$ trains per hour for a static traffic distribution (23), calculated by solving the one-dimensional problem in [1], determining only the total number of trains, has been denoted. While all three methods exhibit large deviations for $w_R < 5$, relatively stable results have been obtained for weights $w_R > 5$. In particular, variant TS-C-TR, using TS instead of EI, shows relatively constant standard deviations of 5 to 7 trains per hour. In comparison, EI-Exp-CP0 exhibits standard deviations of up to 20 trains per hour for small w_R ,

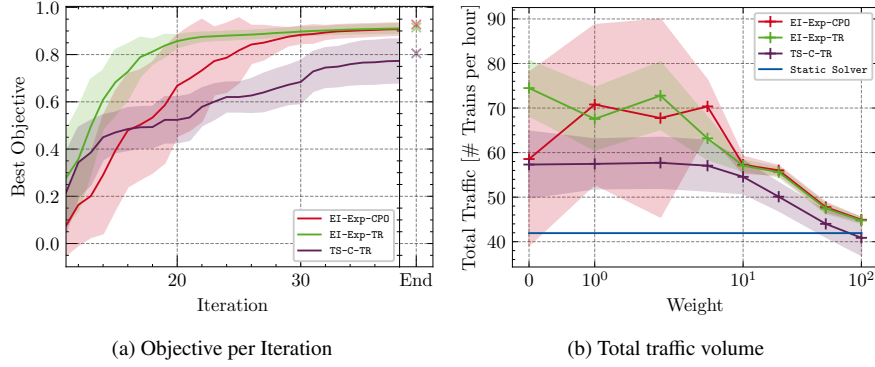


Figure 6: Results for the modeled junction from Fig. 5. Objective distribution per iteration for $w_R = 5$ (6a) and total traffic volume over different weights (6b). The total traffic volume with our variants is compared to a solution from a static solver.

but under one train per per hour for $w_R = 100$. Note that the lowest standard deviations are obtained with EI-Exp-TR, between 7.2 and 0.6 trains per hour.

Comparing the optimal traffic volume (and objective), EI-Exp-TR and EI-Exp-CPO show similar results for $w_R \geq 10$, finding solutions with 2 to 8 trains per hour more than TS-C-TR. Their medians are, depending on w_R , between 30 and 2.5 trains per hour higher, than the timetable capacity of $n_{\max} \approx 41.92$ trains per hour for a static traffic distribution.

Higher weight factors w_R enforce smaller deviations from the in (23) specified route distribution $\tilde{p} = (\tilde{p}_r)_{r \in \mathcal{R}}$. In the following, we analyze the *penalty* or distance $\text{dist}(p, \tilde{p})$ between the route distribution $p = (p_r)_{r \in \mathcal{R}}$ for a fixed solution λ to the specified $\tilde{p}$. We utilize the Euclidean metric $\text{dist}(p, \tilde{p}) = \sqrt{\sum_{r \in \mathcal{R}} (p_r - \tilde{p}_r)^2}$ as the distance here.

Based on all obtained solutions with method EI-Exp-TR, Fig. 7 compares this penalty to the traffic volume difference $\Delta_{\text{Traffic}} = n(\lambda^*) - n_{\max}$ between the traffic volume $n(\lambda^*)$ for the obtained solution λ^* and the static timetable capacity $n_{\max}$.

In accordance with the previously discussed results, deviations between solutions are particularly high for lower weights $w_R < 10$, and more distinct clusters can be recognized for higher weights.

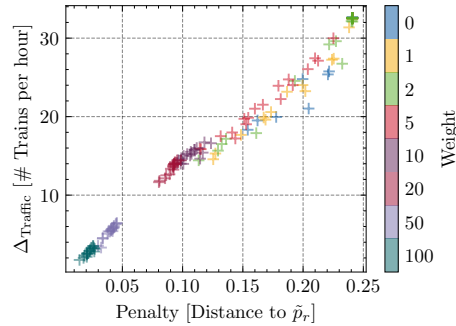


Figure 7: Traffic volume difference and penalty of all obtained solutions for EI-Exp-TR.

However, the results indicate a significant correlation between Δ_{Traffic} and the penalty of the obtained solutions for EI-Exp-TR. For solutions close to the given distribution with small penalties, less traffic volume can be gained than for solutions with higher distances to $\bar{p}$. Furthermore, instances with small weights w_R exhibit far more variation in their balance between adherence to the given route distribution and achieved traffic volume.

These observed effects also follow intuition on the influence of the weight. As the weight increases, the optimization problem **Ptrap** becomes more regularized, smoothing the landscape of the objective. Conversely, smaller weights yield more flexibility at the cost of multiple local minima.

Therefore, infrastructure managers, applying the introduced methods for long-term planning problems, should carefully select the highest weight factor suitable for their exact needs, in order to take advantage of the enhanced convergence and stability of solutions for increased penalty weights.

8. Discussion and Conclusion

In this work, a novel method to determine the performance capability of a railway junction under consideration of dynamic traffic distributions has been introduced. Extensions to an existing queueing-based method have been developed (Section 5) by implementing uncertainties within capacity constraints of a traffic rate assignment problem. The introduced method utilizes BO in order to efficiently solve the presented penalized traffic rate assignment problem **Ptrap** from Section 4. For this, in Section 6, we tailored standard BO algorithms to the setting of known objective under black-box constraints and included domain knowledge in the GP surrogates to increase sample efficiency. On an exemplary infrastructure of a small railway junction, we performed an ablation of our approaches in Section 7.1 of which we selected the most promising candidates for further investigation. These selected variants have been applied to a complex junction infrastructure adapted from a real-world example in Section 7.2, highlighting the method’s ability to assess infrastructure capacity and further quantifying the impact of allowing deviations from traditionally static traffic distributions.

Within the case studies, the real-world applicability of the proposed methods becomes evident, especially when considering scenarios with a high emphasis on small deviations from enforced distributions. However, in the case of small regularization weights, results show high variability for methods utilizing EI on complex infrastructure examples. This indicates that small weights can lead to optimization problems with multiple local optima, which pose a challenge to the introduced methods, iteratively improving their solution quality in the neighborhood of the last best solution. Even for high regularization weights, obtained results are not

guaranteed to form a global optimum to the posed traffic rate assignment problem. Still, all variants outperformed the static solution, indicating a clear potential of adapting the proposed methods in practice. Future work could leverage parametric model-checking alongside classical non-linear solvers with multiple random initializations to efficiently locate promising local optima when solving **PTRAP**. Further extensions may include the adaption of capacity constraints to reflect other capacity metrics or the performance analysis of entire railway networks, utilizing element-wise capacity constraints.

To conclude, the BO variants developed in this work efficiently obtain feasible, locally optimal solutions for the capacity-constrained traffic rate assignment problem under dynamic traffic distributions. Our framework of optimizing known objectives subject to unknown constraints represents a novel contribution to the field of BO with potential applications across multiple domains beyond railway applications where similar problem structures arise. For our specific problem, it enables infrastructure managers to specify acceptable variations in traffic composition when estimating the timetable-independent capacity of railway junctions, making it particularly well-suited for long-term infrastructure planning scenarios where future traffic distributions across train types or lines remain uncertain.

Acknowledgements

This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 2236/2. Computational Experiments were performed with computing resources granted by RWTH Aachen University under project `rwth1635`.

CRedit authorship contribution statement

Tamme Emunds: Conceptualization, Methodology, Software, Formal Analysis, Writing – original draft. **Paul Brunzema:** Conceptualization, Methodology, Software, Formal Analysis, Writing – original draft. **Sebastian Trimpe:** Funding Acquisition, Supervision, Writing - review & editing. **Nils Nießen:** Funding Acquisition, Supervision, Writing - review & editing.

References

- [1] T. Emunds, N. Nießen, Evaluating railway junction infrastructure: A queueing-based, timetable-independent analysis, *Transportation Research Part C: Emerging Technologies* 165 (2024) 104704. [doi:10.1016/j.trc.2024.104704](https://doi.org/10.1016/j.trc.2024.104704).

- [2] International Union of Railways (UIC), Capacity, 2nd Edition, no. 406 in UIC Code, UIC, Paris, France, 2013.
- [3] A. Landex, L. W. Jensen, Measures for track complexity and robustness of operation at stations, *Journal of Rail Transport Planning & Management* 3 (1-2) (2013) 22–35. doi:[10.1016/j.jrtpm.2013.10.003](https://doi.org/10.1016/j.jrtpm.2013.10.003).
- [4] R. M. Goverde, F. Corman, A. D’Ariano, Railway line capacity consumption of different railway signalling systems under scheduled and disturbed conditions, *Journal of Rail Transport Planning & Management* 3 (3) (2013) 78–94. doi:[10.1016/j.jrtpm.2013.12.001](https://doi.org/10.1016/j.jrtpm.2013.12.001).
- [5] R. M. Goverde, Railway timetable stability analysis using max-plus system theory, *Transportation Research Part B: Methodological* 41 (2) (2007) 179–201. doi:[10.1016/j.trb.2006.02.003](https://doi.org/10.1016/j.trb.2006.02.003).
- [6] N. Bešinović, R. M. P. Goverde, Capacity Assessment in Railway Networks, in: R. Borndörfer, T. Klug, L. Lamorgese, C. Mannino, M. Reuther, T. Schlechte (Eds.), *Handbook of Optimization in the Railway Industry*, Springer International Publishing, Cham, 2018, pp. 25–45. doi:[10.1007/978-3-319-72153-8_2](https://doi.org/10.1007/978-3-319-72153-8_2).
- [7] L. W. Jensen, A. Landex, O. A. Nielsen, L. G. Kroon, M. Schmidt, Strategic assessment of capacity consumption in railway networks: Framework and model, *Transportation Research Part C: Emerging Technologies* 74 (2017) 126–149. doi:[10.1016/j.trc.2016.10.013](https://doi.org/10.1016/j.trc.2016.10.013).
- [8] L. W. Jensen, M. Schmidt, O. A. Nielsen, Determination of infrastructure capacity in railway networks without the need for a fixed timetable, *Transportation Research Part C: Emerging Technologies* 119 (2020) 102751. doi:[10.1016/j.trc.2020.102751](https://doi.org/10.1016/j.trc.2020.102751).
- [9] N. Weik, J. Warg, I. Johansson, M. Bohlin, N. Nießen, Extending UIC 406-based capacity analysis – New approaches for railway nodes and network effects, *Journal of Rail Transport Planning & Management* 15 (2020) 100199. doi:[10.1016/j.jrtpm.2020.100199](https://doi.org/10.1016/j.jrtpm.2020.100199).
- [10] R. L. Burdett, E. Kozan, Techniques for absolute capacity determination in railways, *Transportation Research Part B: Methodological* 40 (8) (2006) 616–632. doi:[10.1016/j.trb.2005.09.004](https://doi.org/10.1016/j.trb.2005.09.004).
- [11] Z. Liao, H. Li, J. Miao, F. Corman, Railway capacity estimation considering vehicle circulation: Integrated timetable and vehicles scheduling on hybrid

- time-space networks, *Transportation Research Part C: Emerging Technologies* 124 (2021) 102961. doi:[10.1016/j.trc.2020.102961](https://doi.org/10.1016/j.trc.2020.102961).
- [12] M. Yaghini, N. Nikoo, H. R. Ahadi, An integer programming model for analysing impacts of different train types on railway line capacity, *Transport* 29 (1) (2014) 28–35. doi:[10.3846/16484142.2014.894938](https://doi.org/10.3846/16484142.2014.894938).
 - [13] R. L. Burdett, Multi-objective models and techniques for analysing the absolute capacity of railway networks, *European Journal of Operational Research* 245 (2) (2015) 489–505. doi:[10.1016/j.ejor.2015.03.020](https://doi.org/10.1016/j.ejor.2015.03.020).
 - [14] G. Caimi, F. Chudak, M. Fuchsberger, M. Laumanns, R. Zenklusen, A New Resource-Constrained Multicommodity Flow Model for Conflict-Free Train Routing and Scheduling, *Transportation Science* 45 (2) (2011) 212–227. doi:[10.1287/trsc.1100.0349](https://doi.org/10.1287/trsc.1100.0349).
 - [15] F. Leutwiler, F. Corman, A logic-based Benders decomposition for microscopic railway timetable planning, *European Journal of Operational Research* 303 (2) (2022) 525–540. doi:[10.1016/j.ejor.2022.02.043](https://doi.org/10.1016/j.ejor.2022.02.043).
 - [16] T. Spanning, B. Büchel, F. Corman, Train Delay Predictions Using Markov Chains Based on Process Time Deviations and Elastic State Boundaries, *Mathematics* 11 (4) (2023) 839. doi:[10.3390/math11040839](https://doi.org/10.3390/math11040839).
 - [17] F. Corman, J. Henken, Estimating aggregate railway performance from realized empirical data: Literature review, a test case and a research roadmap, *Journal of Rail Transport Planning & Management* 22 (2022) 100316. doi:[10.1016/j.jrtpm.2022.100316](https://doi.org/10.1016/j.jrtpm.2022.100316).
 - [18] M. Abril, F. Barber, L. Ingolotti, M. Salido, P. Tormos, A. Lova, An assessment of railway capacity, *Transportation Research Part E: Logistics and Transportation Review* 44 (5) (2008) 774–806. doi:[10.1016/j.tre.2007.04.001](https://doi.org/10.1016/j.tre.2007.04.001).
 - [19] C. Schmidt, U. Martin, Erhöhung der Effektivität und Transparenz bei Leistungsuntersuchungen mit Simulationsverfahren, *Eisenbahntech. Rundsch. (ETR)* 07-08/2010 (2010) 463–468.
 - [20] J. Liang, U. Martin, Y. Cui, Increasing performance of railway systems by exploitation of the relationship between capacity and operation quality, *Journal of Rail Transport Planning & Management* 7 (3) (2017) 127–140. doi:[10.1016/j.jrtpm.2017.08.002](https://doi.org/10.1016/j.jrtpm.2017.08.002).

- [21] W. Schwanhäuß, Die Bemessung der Pufferzeiten im Fahrplangefüge der Eisenbahn, Ph.D. thesis, RWTH Aachen University (1974).
- [22] T. Büker, B. Seybold, Stochastic modelling of delay propagation in large networks, *Journal of Rail Transport Planning & Management* 2 (2012) 34–50. [doi:10.1016/j.jrtpm.2012.10.001](https://doi.org/10.1016/j.jrtpm.2012.10.001).
- [23] İ. Şahin, Markov chain model for delay distribution in train schedules: Assessing the effectiveness of time allowances, *Journal of Rail Transport Planning & Management* 7 (3) (2017) 101–113. [doi:10.1016/j.jrtpm.2017.08.006](https://doi.org/10.1016/j.jrtpm.2017.08.006).
- [24] E. Wendler, The scheduled waiting time on railway lines, *Transportation Research Part B: Methodological* 41 (2) (2007) 148–158. [doi:10.1016/j.trb.2006.02.009](https://doi.org/10.1016/j.trb.2006.02.009).
- [25] N. Weik, Long-term capacity planning of railway infrastructure: A stochastic approach capturing infrastructure unavailability, Ph.D. thesis, RWTH Aachen University (2020).
- [26] N. Nießen, Waiting and loss probabilities for route nodes, in: *International Conference on Railway Operations Modelling and Analysis (RailCopenhagen)*, Copenhagen, 2013.
- [27] C. Schmitz, N. Weik, S. Zieger, N. Nießen, A. Schmeink, Markov models for the performance analysis of railway networks, in: *International Conference on Railway Operations Modelling and Analysis (RailLille)*, Lille, 2017.
- [28] I. Bychkov, A. Kazakov, A. Lempert, M. Zharkov, Modeling of Railway Stations Based on Queuing Networks, *Applied Sciences* 11 (5) (2021) 2425. [doi:10.3390/app11052425](https://doi.org/10.3390/app11052425).
- [29] T. Emunds, N. Nießen, A Queueing-Based Approach for Timetable-Independent Railway Station Performance Analysis, in: *International Conference on Railway Operations Modelling and Analysis (RailDresden)*, Dresden, 2025.
- [30] T. Huisman, R. J. Boucherie, N. M. Van Dijk, A solvable queueing network model for railway networks and its validation and applications for the Netherlands, *European Journal of Operational Research* 142 (1) (2002) 30–51. [doi:10.1016/S0377-2217\(01\)00269-7](https://doi.org/10.1016/S0377-2217(01)00269-7).

- [31] A. Kazakov, A. Lempert, M. Zharkov, An approach to railway network sections modeling based on queuing networks, *Journal of Rail Transport Planning & Management* 27 (2023) 100404. doi:[10.1016/j.jrtpm.2023.100404](https://doi.org/10.1016/j.jrtpm.2023.100404).
- [32] T. Emunds, P. Wang, N. Nießen, Modelling Traffic-Distribution based Capacity in Frequency Assignment Problems for a Railway Junction, in: *International Conference on Railway Operations Modelling and Analysis (RailDresden)*, Dresden, 2025.
- [33] F. Fuchs, A. Trivella, F. Corman, Enhancing the interaction of railway timetabling and line planning with infrastructure awareness, *Transportation Research Part C: Emerging Technologies* 142 (2022) 103805. doi:[10.1016/j.trc.2022.103805](https://doi.org/10.1016/j.trc.2022.103805).
- [34] R. Calandra, A. Seyfarth, J. Peters, M. P. Deisenroth, Bayesian optimization for learning gaits under uncertainty: An experimental comparison on a dynamic bipedal walker, *Annals of Mathematics and Artificial Intelligence* (2016).
- [35] F. Berkenkamp, A. P. Schoellig, A. Krause, Safe controller optimization for quadrotors with Gaussian processes, in: *IEEE International Conference on Robotics and Automation (ICRA)*, 2016.
- [36] M. Neumann-Brosig, A. Marco, D. Schwarzmann, S. Trimpe, Data-efficient autotuning with Bayesian optimization: An industrial control study, *IEEE Transactions on Control Systems Technology* (2019).
- [37] R.-R. Griffiths, J. M. Hernández-Lobato, Constrained Bayesian optimization for automatic chemical design using variational autoencoders, *Chemical Science* (2020).
- [38] L. Colliandre, C. Muller, Bayesian optimization in drug discovery, *High Performance Computing for Drug Discovery and Biomedicine* (2023).
- [39] N. Maus, H. Jones, J. Moore, M. J. Kusner, J. Bradshaw, J. Gardner, Local latent space Bayesian optimization over structured inputs, in: *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.
- [40] B. Hickish, D. I. Fletcher, R. F. Harrison, Investigating Bayesian optimization for rail network optimization, *International Journal of Rail Transportation* 8 (4) (2020) 307–323.

- [41] R. Shi, X. Xu, A train arrival delay prediction model using xgboost and Bayesian optimization, in: IEEE International Conference on Intelligent Transportation Systems (ITSC), IEEE, 2020, pp. 1–6.
- [42] J. Luo, P. Huang, Q. Peng, A multi-output deep learning model based on Bayesian optimization for sequential train delays prediction, *International Journal of Rail Transportation* 11 (5) (2023) 705–731.
- [43] H.-P. Wan, Y.-Q. Ni, A new approach for interval dynamic analysis of train-bridge system based on Bayesian optimization, *Journal of Engineering Mechanics* 146 (5) (2020) 04020029.
- [44] J. Kirschner, M. Mutny, N. Hiller, R. Ischebeck, A. Krause, Adaptive and safe Bayesian optimization in high dimensions via one-dimensional subspaces, in: International Conference on Machine Learning, PMLR, 2019, pp. 3429–3438.
- [45] D. Eriksson, M. Pearce, J. R. Gardner, R. Turner, M. Poloczek, Scalable Global Optimization via Local Bayesian Optimization (2020). [arXiv:1910.01739](https://arxiv.org/abs/1910.01739), [doi:10.48550/arXiv.1910.01739](https://doi.org/10.48550/arXiv.1910.01739).
- [46] S. Müller, A. von Rohr, S. Trimpe, Local policy search with Bayesian optimization, *Advances in Neural Information Processing Systems* 34 (2021) 20708–20720.
- [47] I. Bogunovic, J. Scarlett, V. Cevher, Time-varying Gaussian process bandit optimization, in: Artificial Intelligence and Statistics, PMLR, 2016, pp. 314–323.
- [48] A. Bardou, P. Thiran, G. Ranieri, This too shall pass: Removing stale observations in dynamic Bayesian optimization, in: Advances in Neural Information Processing Systems, 2024.
- [49] P. Brunzema, A. von Rohr, F. Solowjow, S. Trimpe, Event-triggered time-varying Bayesian optimization, *Transactions on Machine Learning Research* (2025).
- [50] J. Gardner, M. Kusner, K. Weinberger, J. Cunningham, et al., Bayesian optimization with inequality constraints, in: International Conference on Machine Learning, PMLR, 2014, pp. 937–945.
- [51] M. A. Gelbart, J. Snoek, R. P. Adams, Bayesian optimization with unknown constraints, in: Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence, 2014, pp. 250–259.

- [52] D. Eriksson, M. Poloczek, Scalable constrained bayesian optimization, in: International conference on artificial intelligence and statistics, PMLR, 2021, pp. 730–738.
- [53] R. Garnett, Bayesian optimization, Cambridge University Press, 2023.
- [54] A. Marco, P. Hennig, S. Schaal, S. Trimpe, On the design of LQR kernels for efficient controller learning, in: IEEE Conference on Decision and Control (CDC), IEEE, 2017, pp. 5193–5200.
- [55] DB InfraGO, Richtlinie 405 Fahrwegkapazität, (german only) (2022).
- [56] W. Schwanhäuß, K. Schultze, Ermittlung von Qualitätsmaßstäben für die Berechnung der Leistungsfähigkeit eines Streckenabschnittes und Entwicklung eines Rechenverfahrens zur Ermittlung von Endverspätungen: Forschungsarbeit für die Deutsche Bundesbahn, (german only) (1982).
- [57] C. Rasmussen, C. Williams, Gaussian Processes for Machine Learning, MIT Press, 2006.
- [58] P. Auer, Using confidence bounds for exploitation-exploration trade-offs, Journal of Machine Learning Research (2002).
- [59] N. Srinivas, A. Krause, S. Kakade, M. Seeger, Gaussian process optimization in the bandit setting: No regret and experimental design, in: International Conference on Machine Learning, 2010.
- [60] D. R. Jones, M. Schonlau, W. J. Welch, Efficient global optimization of expensive black-box functions, Journal of Global optimization (1998).
- [61] W. R. Thompson, On the likelihood that one unknown probability exceeds another in view of the evidence of two samples, Biometrika 25 (3-4) (1933) 285–294.
- [62] I. Hansen, J. Pachl, Railway Timetabling and Operations: Analysis, Modelling, Optimisation, Simulation, Performance, Evaluation, Eurail press, 2014.
- [63] D. Parker, G. Norman, M. Kwiatkowska, Prism (2000).
URL <https://www.prismmodelchecker.org/>
- [64] C. Hensel, S. Junges, J.-P. Katoen, T. Quatmann, M. Volk, The probabilistic model checker storm, International Journal on Software Tools for Technology Transfer 24 (4) (2022) 589–610. doi:10.1007/s10009-021-00633-z.

- [65] D. G. Kendall, Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain, *The Annals of Mathematical Statistics* (1953) 338354.
- [66] G. Hertel, Exakte lösung zur berechnung der wartegleiszahl vor im einrichtungsbetrieb befahrenen streckengleisen bei nicht-poisson-ankünften (g/m/1-wartesystem), *Wissenschaftliche Zeitschrift der Hochschule für Verkehrswesen* 31 (1984) 195–205.
- [67] S. Ament, S. Daulton, D. Eriksson, M. Balandat, E. Bakshy, Unexpected improvements to expected improvement for bayesian optimization, *Advances in Neural Information Processing Systems* 36 (2023) 20577–20612.
- [68] J. Nocedal, S. J. Wright, *Numerical optimization*, Springer, 1999.
- [69] D. Kraft, A software package for sequential quadratic programming, *Forschungsbericht- Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt* (1988).
- [70] A. Meurer, C. P. Smith, M. Paprocki, O. Čertík, S. B. Kirpichev, M. Rocklin, A. Kumar, S. Ivanov, J. K. Moore, S. Singh, T. Rathnayake, S. Vig, B. E. Granger, R. P. Muller, F. Bonazzi, H. Gupta, S. Vats, F. Johansson, F. Pedregosa, M. J. Curry, A. R. Terrel, v. Roučka, A. Saboo, I. Fernando, S. Kulkarni, R. Cimrman, A. Scopatz, *Sympy: symbolic computing in python*, *PeerJ Computer Science* 3 (2017) e103. doi:10.7717/peerj-cs.103. URL <https://doi.org/10.7717/peerj-cs.103>
- [71] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy 1.0 Contributors, *SciPy 1.0: Fundamental algorithms for scientific computing in python*, *Nature Methods* 17 (2020) 261–272. doi:10.1038/s41592-019-0686-2.
- [72] M. Balandat, B. Karrer, D. R. Jiang, S. Daulton, B. Letham, A. G. Wilson, E. Bakshy, *BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization* (2020). arXiv:1910.06403, doi:10.48550/arXiv.1910.06403.

- [73] P. Pellegrini, G. Marlière, J. Rodriguez, Optimal train routing and scheduling for managing traffic perturbations in complex junctions, *Transportation Research Part B: Methodological* 59 (2014). [doi:10.1016/j.trb.2013.10.013](https://doi.org/10.1016/j.trb.2013.10.013).

Appendix A. Notation

Table A.4: Overview of Variables and Notation

Symbol	Type	Description
<i>Infrastructure and Network</i>		
J	Set	Railway junction $J = (\mathcal{R}, C)$
$\mathcal{R}$	Set	Set of k routes in junction
C	Matrix	Conflict matrix $C \in \{0, 1\}^{k \times k}$
$\mathcal{U}$	Set	Set of rolling stock units
$\mathcal{O}$	Set	Set of occupation requests $o = (r, u) \in \mathcal{R} \times \mathcal{U}$
<i>Traffic and Capacity</i>		
λ_o	Scalar	Arrival rate for request o
λ_r	Scalar	arrival rate for route r , $\lambda_r = \sum_{(r,u) \in \mathcal{O}} \lambda_{(r,u)}$
λ	Vector	Assignment vector $(\lambda_o)_{o \in \mathcal{O}} \in \mathbb{R}^d$
f_o	Scalar	Frequency of trains per request o
t_U	Scalar	Fixed time horizon
d	Scalar	Dimension of assignment vector, $d = \mathcal{O} $
<i>Service Times and Operations</i>		
$h_{o,o'}$	Function	Minimum headway time between requests o and o'
b_o	Scalar	Average occupation time for request o
b_r	Scalar	Average occupation time for route r
μ_r	Scalar	Service rate for route r , $\mu_r = 1/b_r$
ρ_r	Scalar	Occupation ratio for route r , $\rho_r = \lambda_r/\mu_r$
v_A, v_S	Scalar	Coefficients of variation of the arrival and service process
<i>Optimization Problem</i>		
$f(\lambda)$	Function	Objective function
Λ	Set	Feasible set
λ^*	Vector	Optimal solution
ub_o	Function	Upper bound function for request o
$p_u(\lambda)$	Function	Traffic distribution for train type u
$p_r(\lambda)$	Function	Traffic distribution for route r
$\tilde{p}_u, \tilde{p}_r$	Scalar	Target distributions for types/routes
w_U, w_R	Scalar	Weighting factors for distribution penalties
<i>Constraint Formulation</i>		
L_r	Scalar	Expected queue length for route r
$L_{\text{limit},r}$	Scalar	Threshold for expected queue length on route r

Table A.4: Overview of Variables and Notation

Symbol	Type	Description
$c_r(\lambda)$	Function	Capacity constraint function for route r
$p_{pt,r}$	Scalar	Proportion of passenger traffic on route r
<i>Bayesian Optimization</i>		
$m(\lambda)$	Function	GP mean function
$k(\lambda, \lambda')$	Function	GP kernel function
$\alpha(\lambda)$	Function	Acquisition function
$\mathcal{D}_t$	Set	Dataset at iteration t
$\mathcal{TR}_t$	Set	Trust region at iteration t
L_t	Scalar	Length of the TuRBO hyperbox at iteration t
$\mathcal{F}$	Set	Feasible points in candidates set
$\mathcal{M}$	Set	Candidates for selection
<i>GP Model Parameters</i>		
β, w_i, γ	Scalar	Exponential mean function parameters
σ_f^2	Scalar	GP output scale
ν	Scalar	Parameter for the Matérn kernel
ℓ_i	Scalar	Length scale for dimension i
σ_n^2	Scalar	Noise variance
Θ_r	Set	Set of learnable hyperparameters
<i>General notation</i>		
$\mathcal{I}_T, \mathcal{I}_d$	Set	Set of first T/d positive integers
Δ	Function	Euclidean distance
σ	Scalar	Standard Deviation