IEEE Open Journal of
Intelligent Transportation Systems

Received 29 November 2021; accepted 2 May 2022. Date of publication 4 May 2022; date of current version 12 May 2022.
Digital Object Identifier 10.1109/OJITS.2022.3172509

A Methodology for Monitoring Rail Punctuality Improvements

C. W. PALMQVIST [1] AND I. KRISTOFFERSSON [2]
[1] Department of Technology and Society, Lund University, 221 00 Lund, Sweden
[2] Traffic Analysis and Logistics Department, Swedish National Road and Transport Research Institute, 114 28 Stockholm, Sweden
CORRESPONDING AUTHOR: C. W. PALMQVIST (e-mail: carl-william.palmqvist@tft.lth.se)
This work was supported in part by the Swedish Transport Administration under Grant TRV2019/70093.

ABSTRACT Punctuality is an important aspect of train operations, highly valued by passengers. Both Swedish and Norwegian railways have introduced frameworks to systematically improve punctuality in their systems, inspired by an extensive literature on Total Quality Management. After about a decade with these frameworks, we can see that punctuality has risen by about 2-3 percentage points. However, this pace of improvements is slower than desired. We propose that there is a gap between what most individual improvement efforts deliver, and what can be detected by directly monitoring punctuality. This gap stifles the desired culture of constant improvements. We instead propose a methodology for how to monitor punctuality improvements, by focusing on the constituents of a train trip. Using 20 years of data from commuter trains in three metropolitan regions (Stockholm, Gothenburg & Malmö), we show the frequency of runtime and dwell time delays is directly related to punctuality. These delay frequencies are also easy to measure and target, and more easily capture the intended effects of specific improvement efforts. Our hope is that this framework and measures such as these will better enable systematic efforts to improve railway punctuality.

INDEX TERMS Public transport, rail transportation, railway monitoring, rail transportation reliability.

I. INTRODUCTION

THE IMPORTANCE of travel time reliability to rail passengers has long been acknowledged in scientific literature [1]–[4]. An extended travel time that is known ahead of time can be planned for, whereas an unreliable service forces travelers to add extra time to their travel plans – the size of which is unclear. A reliable rail travel service depends on high train punctuality, and to achieve this, systematic quality work is needed.

Several methodological papers concerning timetable development have been published, with the goal of creating timetables that are robust towards delays [5]–[7]. In comparison, the literature on methods for systematic punctuality improvement based on an already decided timetable is scarce. To our knowledge, only a few such frameworks have been attempted and published. For instance, [8] developed methods for rail quality improvement in Great Britain, [9] developed the PULS model in Sweden.

In the most recent attempt we are aware of, [10] created the Punctuality Improvement Method System (PIMS) method in Norway. The PIMS method acknowledges that systematic punctuality improvement work is similar to other quality work. It is thus based on basic principles from the Total Quality Management (TQM) research area, such as customer focus, continuous improvement, total participation, and societal networking [11]–[13]. Other approaches to quality work are Lean and Six Sigma, but [14] argue that the basic ideas of Lean and Six Sigma are the same as in TQM, so for brevity we refer to these approaches as TQM. TQM emerged in the manufacturing sector, but has subsequently spread to the service sector in particular [15]. Reference [16] conducted a systematic literature review of TQM work, emphasizing the importance of customer focus for successful implementation.

Even though the PIMS method development work was based on well-known TQM principles and included both

The review of this article was arranged by Associate Editor Margarida Coelho.

researchers and practitioners, punctuality levels in Sweden and Norway have only improved marginally since the introduction of systematic quality work. The suggestion in this paper is that the monitoring part of PIMS (and similar approaches) needs to be amended to achieve a more substantial improvement in rail punctuality levels. We argue that an important reason for the stagnation in punctuality improvement is that the gap between the final goal of high train punctuality, and the traditionally measured factors that influence punctuality, such as factors within categories such as infrastructure, operations, timetable and weather [17]–[24], is too large. The relationships between these traditional influencing factors and punctuality are weak and it is still unclear to what extent an improvement in one of the influencing factors will affect overall punctuality levels.

The aim of this paper is therefore to amend the monitoring part of punctuality improvement frameworks, so that the relation between what is monitored/measured and overall punctuality levels become clearer and can contribute to actual punctuality improvements. The purpose of this paper is not to provide evidence for measures which improve punctuality per se, only how to monitor such efforts. The following section of this paper provides with a background description of existing rail quality work. A framework for monitoring rail punctuality improvements is then developed in Section III. In Section IV, the methodology is demonstrated by application to data. Section V concludes the paper.

II. BACKGROUND

A. PUNCTUALITY IMPROVEMENT SYSTEM (PIMS)

The method development in this paper starts with the PIMS method as a baseline [10], which is based on Total Quality Management (TQM) discussed in the introduction. According to [10], the PIMS method encompasses: i) continuous improvement, ii) systemic combination of abstract thought and empirical data, iii) interaction between a base organization and an improvement project, and iv) user friendliness. Fig. 1 shows the steps in the PIMS method. The PIMS method has been applied to train punctuality quality work in Norway [10]. The red circle in Fig. 1 indicates the Improvement work part that is amended in this paper.

B. TOGETHER FOR TRAINS ON TIME (TTT)

The Swedish analogy to the PIMS work in Norway has been the TTT – Together for Trains on Time. TTT was established in 2013 and constitutes a systematic quality work regarding train punctuality through collaboration between organizations from the railway infrastructure manager, railway operators, the railway real estate owner and public transport authorities. All organizations within TTT share the goal that 95% of trains in Sweden arrive to their destination with a delay of at most 5 minutes after the scheduled arrival time (with the seconds being truncated; whether canceled trains are included or not has changed over the years since TTT started in 2013). TTT releases monthly punctuality reports online [25] and a yearly summary of the quality work done and results

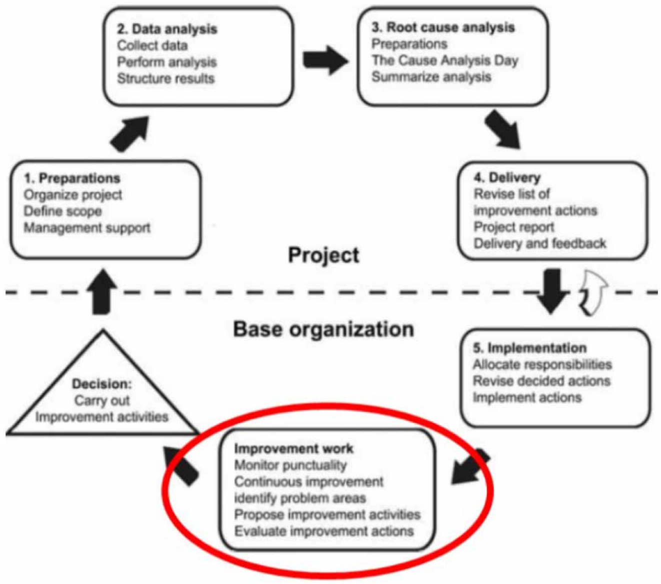

FIGURE 1. An illustration of the PIMS framework [10] with the focus of this paper circled in red.

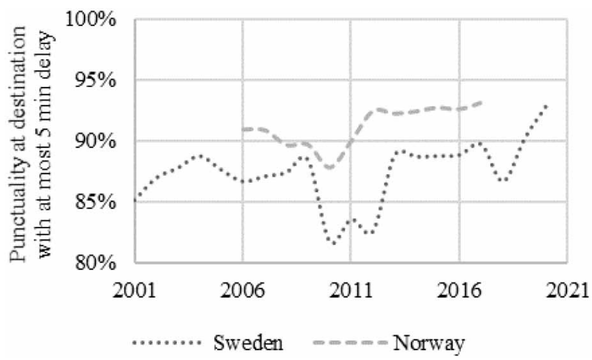

FIGURE 2. Recent development of punctuality in Sweden and Norway. Own analysis of nation-wide operational data for passenger and freight trains. The big dips in punctuality are largely associated with particularly harsh winters, and an insufficient ability to deal with these.

achieved regarding train punctuality in Sweden [26]. The structure of TTT's punctuality monitoring work has changed slightly over the years, but the focus has been on collecting information about delay hours, categorized by the attributed cause of the delay, e.g., infrastructure failure or trespassing. The work within TTT was analyzed and a number of new indicators for monitoring was suggested in [27]. According to [28], the collaboration element of TTT appears to have been successful.

C. DEVELOPMENT OF PUNCTUALITY IN SWEDEN AND NORWAY

Fig. 2 illustrates the development of punctuality in both Sweden and Norway over the last several years. We see very similar trends for the periods where we have overlapping data, though the level in Norway has consistently been higher than in Sweden. Following the work of [10], we can see that the punctuality has reached a higher level than

TABLE 1. Overview of possible scenarios during train operations.

	Departure time deviation at origin		
	<0, early	=0, on time	>0, late
a) First departure	Early departure	Timely departure	Late departure
		Arrival time deviation	
b) Runtime deviation	<0, early	=0, on time	>0, late
<0, faster than scheduled	Too fast	Fully recovers delay	Partially recovers delay
=0, as scheduled	Stays ahead	Stays on time	Unable to recover delay
>0, slower than scheduled	Less early	No longer early	Delay occurs
		Departure time deviation	
c) Dwell time deviation	<0, early	=0, on time	>0, late
<0, faster than scheduled	Too fast	Fully recovers delay	Partially recovers delay
=0, as scheduled	Stays ahead	Stays on time	Unable to recover delay
>0, slower than scheduled	Less early	No longer early	Delay occurs

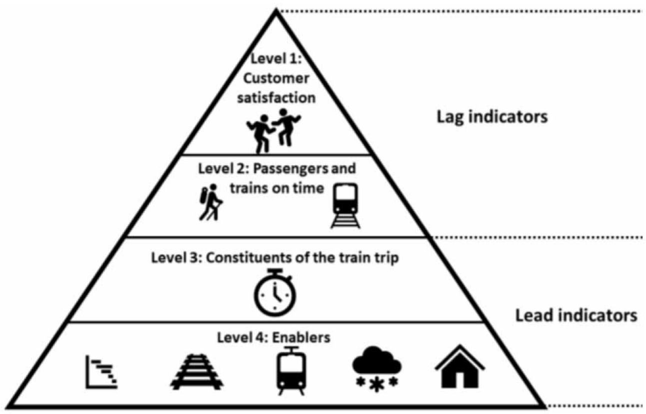

FIGURE 3. Framework developed within the project for identifying and analyzing punctuality indicators of different types.

before. The level in Norway reached a stable level of about 93% in 2012, up from about 90%, after which there has been little improvement. In Sweden we also see a slight improvement in the period 2013-2019 (89%) compared to 2001-2009 (87%). In both countries, however, progress appears to have stalled since about 2012 (Norway) and 2013 (Sweden). This suggests that in both countries, the above-mentioned systematic approach has arguably raised the level of punctuality by 2-3%-points. The target of 95% is still a long way off, however, and the approach thus needs to be modified to deliver better results.

III. A FRAMEWORK FOR MONITORING PUNCTUALITY

A. FOUR LEVELS OF INDICATORS

This paper develops a framework for indicators at different levels, as illustrated in Fig. 3. Punctuality is the outcome of a several processes at different levels and the framework can assist in sorting out which indicator to measure at which level, depending on the purpose of the measurement. A typical Level 1 indicator, what (public) operators care most about, is a customer satisfaction index, which is measured by asking travelers how satisfied they are with their train trip. A typical Level 2 indicator is punctuality, measured as the percentage of trains arriving to the destination with a delay below a certain threshold. Typically, trains that arrive less than 6 minutes behind schedule are considered punctual. Punctuality is monitored on a yearly, monthly, or even daily basis, with the purpose of determining whether punctuality is approaching a chosen goal (e.g., 95% punctual trains). Both Level 1 and Level 2 indicators are so called *lag indicators*, since they are measured *after* the train operation processes have been completed, to evaluate the performance and draw lessons for the future.

Level 3 and Level 4 indicators are, on the other hand, so called *lead indicators*. These measure the performance of the train operation before or while it is conducted, to provide insight into what can be done to decrease delays during the process. The different phases of the train trip associated with Level 3 indicators are explored in greater depth in the next section. Level 4 indicators are related to underlying factors which affect the possibility to complete a railway trip on time. Some examples of these are the timetable, infrastructure, rolling stock, station design, and weather.

Indicators on levels 1, 2 and 4 are relatively commonplace and well established in both research and practice. One of the contributions of this paper is to add Level 3 indicators, intended to bridge the large gap between Level 2 and Level 4. For that reason, the rest of this paper is centered on Level 3 indicators.

B. INVENTORY OF LEVEL 3 INDICATORS

Table 1 shows how a train journey is made up of a sequence of events and processes. Events are when the train departs or arrives at a station, which can happen before, on, or after schedule. In Table 1, each such scenario has a separate column. The runtime (Table 1b) and dwell times (Table 1 c) are processes, which can be faster, slower, or as long as scheduled, each a separate row in Table 1. This cycle is then repeated until the train reaches its destination with a final arrival time, which is often evaluated in terms of punctuality.

Level 3 indicators arise from the intersections of runtimes and arrival times (Table 1b), as well as dwell and departure times (Table 1c). There are nine combinations for both intersections, each of them with different implications, in addition to the three possibilities at the first departure (Table 1a), for a total of 21 (9+9+3) possible indicators. We have color-coded them to help distinguish between the various consequences. Three combinations in Table 1 are green, indicating that these are entirely according to *schedule*. Three of them are highlighted in red, indicating potential *delays*. Four are dark blue, indicating how delays can be *recovered*. Five are yellow, ways in which trains can come to be *ahead* of schedule. The last four, light blue, are ways in which trains come to be *less* ahead of schedule. Of course, the colors per se are not important, they only serve to illustrate the various alternatives.

Notably, we focus on the extent to which these scenarios arise and focus on frequency rather than size of any deviations. Earlier studies, as well as our own analyses [29], have shown that small delays are much more common than larger ones, and that the distributions are far from normal, with long tails of sometimes very large delays, so that averaging across size is not very meaningful or robust.

The *frequency of delay* is a simple and direct measure, and in most cases, it will correspond well to measures of delay size: large delays are more likely to cause secondary delays for other trains, resulting in a greater number of delays than would otherwise be the case. Measuring the frequency is thus a simple and direct way to sidestep the complicated issues of distributions and weighting. If the frequency of delays can be decreased, then the overall delays, the total inconvenience to passenger, and the corresponding costs to operators are all likely to decrease as well.

C. RELATIONSHIP BETWEEN LEVEL 3 AND 2 INDICATORS

Ideally, operations would be executed according to schedule, and we should see a high share of movements coded as green in Table 1. If this is not the case, then categorizing the outcomes of operations according to Table 1 can help more precisely identify and target issues in the operations. For instance, it can aid in determining the extent to which delays arise at the very beginning, before the journey even begins, or when the train runs between stations, or during the dwell times at stations. It can help in identifying when and where train drivers drive faster than scheduled, perhaps suggesting that scheduled travel times are longer than required or that energy consumption could be reduced. It also indicates where margins are used to reduce delays, and where they are not. In general, it provides a more detailed picture of the operations than simply looking at punctuality.

There is a complex interplay between these Level 3 indicators which results in Level 2 indicators such as punctuality. However, the most critical factor, is the extent to which delays arise (red fields in Table 1). If delays do not occur in the first place, the trains cannot be behind schedule. If delays occur, they may at least partially be compensated by later recovery (dark blue fields in Table 1), but this is not instantaneous. The capacity to recover delays is gained by having margins on runtimes and/or dwell times. When these margins are not needed for delay recovery, trains can run faster or stop for less time at stations, allowing them to arrive ahead of schedule (yellow fields in Table 1). This is a sign of wasted capacity, sometimes realized when trains then need to slow down to return to schedule (light blue fields in Table 1). While it is preferable to recover delays than to carry them forward, the capability to do so is costly in terms of capacity and introduces further variability into the operations. If less delays arise in the first place (less red in Table 1), fewer of these margins are required, and possibilities to run ahead of time are reduced, while punctuality improves.

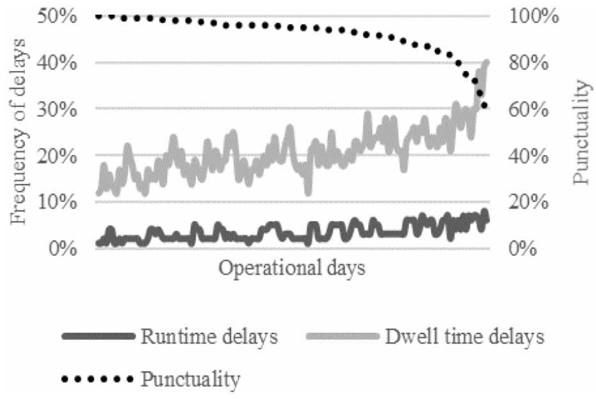

FIGURE 4. Daily variations in frequency of run and dwell time delays along with punctuality, sorted based on punctuality. A small sample of data (6 months) for illustrative purposes.

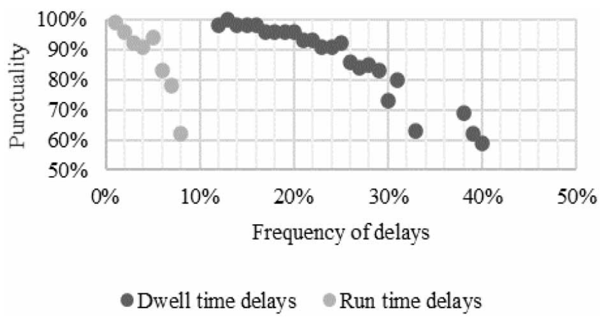

FIGURE 5. Punctuality as a function of frequency of run and dwell time delays respectively. A different presentation of the data used in Fig. 4.

Any given railway will experience varying frequencies of delays and levels of punctuality from day to day, sometimes with large variations. Fig. 4 illustrates an example, showing the shares of run- and dwell times that have been delayed each day, along with punctuality, for a relatively small sample (6 months) of data which has been sorted by punctuality, rather than date. This more clearly shows that the Level 3 indicators, both regarding the frequency of run- and dwell time delays, are inversely related to punctuality. The more delays, the lower the punctuality. We also see a positive correlation between run- and dwell time delays, although the latter is much more common and variable.

To see the relationships even more clearly, and help operationalize them, we group the data on each type of Level 3 indicator (in this case frequency of run- and dwell time delay) and take the average punctuality across these observations. Displaying these relationships in a scatter plot, we get Fig. 5. One benefit of clustering data based on the Level

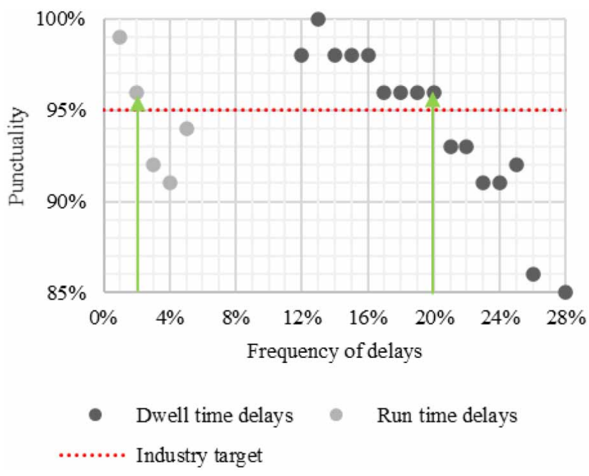

FIGURE 6. Illustration of the method to identify target levels for Level 3 indicators (e.g., run and dwell time delays) in order to reach desired goals for the Level 2 indicator (e.g., punctuality). Based on the sample data in Figs. 4–5.

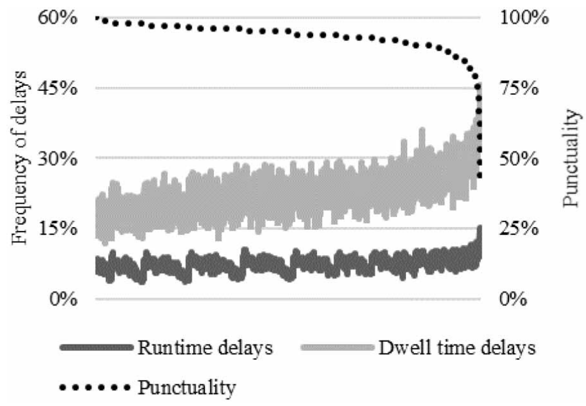

FIGURE 7. Daily variations in frequency of run and dwell time delays along with punctuality, sorted based on punctuality. Data from commuter train operations for the years of 2001-2020 across the three metropolitan regions in Sweden: Stockholm, Gothenburg, and Malmö.

3 indicator and calculating the punctuality within this cluster, rather than the reverse, is that the direction of causality is preserved: we want to reduce the frequency of delays to improve punctuality, not the reverse. On the horizontal axis we have the frequency of delay (for dwell and run times separately), or some other Level 3 indicator, while the vertical axis shows the punctuality (or a Level 2 indicator of choice).

D. METHOD FOR CALCULATING TARGET LEVELS

To make practical use of these relationships, we propose that railway operators set targets for Level 3 indicators (i.e., the frequency of run and dwell time delays), such that the Level 2 indicator of choice (i.e., punctuality) will in turn reach a desired level. In Sweden, for instance, many railway operators aim at a punctuality level of 95%. As we have seen, this has (strict) implications in terms of the frequencies of delays that can be accepted. To illustrate this, please observe the horizontal red, dotted line in Fig. 6, indicating a desired punctuality of 95%. We can then see that up to 2% of run times *can* be delayed while achieving a punctuality of 95% (arrow to the left), while this is *not* true at 3%. For dwell time delays, we see that a frequency of up to 20% is acceptable (arrow to the right), but not 21%. The exact numbers here are only intended as an example, based on a relatively small sample, to illustrate the idea. But in this example, we would advise that the operator in question should aim at consistently keeping the frequencies of run- and dwell time delays at or below 2% and 20%, respectively, to achieve the desired punctuality of 95%.

IV. DEMONSTRATION OF THE METHOD IN A CASE STUDY

A. DATA AND CASE DESCRIPTION

To demonstrate the application of the indicators discussed above, we use a large dataset of commuter train operations. Our data covers the years of 2001-2020 across the three metropolitan regions in Sweden: Stockholm, Gothenburg, and Malmö. In total, this contains 63 million observations of train movements. In detail, there are about 3.4 million unique train runs across 3,290 days, with an average of 1024 per day, and an average of 19 observations made per train run. These data include scheduled and actual departure and arrival times at stations (including technical stations). Our analysis will both cover the pooled dataset and discuss regional differences. The latter is to demonstrate how conditions vary across different railways, and the importance of basing analyses on context-specific data. In the analysis below, we have aggregated the data so that every observation corresponds to a day. This is well aligned with how operators currently track punctuality, and only requires the additional calculation of the Level 3 indicators, which can easily be automated.

B. RELATIONSHIP BETWEEN LEVEL 2 AND 3 INDICATORS

We begin with the pooled data, which is more generally applicable and gives a good overview of commuter train operations in Sweden's three metropolitan regions. Fig. 7 shows punctuality (a Level 2 indicator) and the frequency of run- and dwell time delays (two Level 3 indicators), sorted by punctuality. We can see that run- and dwell time delays are positively correlated, and that these are negatively correlated with punctuality, although there can be substantial volatility in each indicator for each level of punctuality.

Fig. 8 instead groups the data based on the Level 3 indicators (just as Fig. 5 did) and displays the average punctuality at each point. This is a much cleaner way to illustrate the relationships and makes it easy to set target levels. In each of these three figures we see that dwell time delays are much more frequent than runtime delays, but that punctuality seems more sensitive to variations in runtime delays. These are on the pooled data, however, so some caution is warranted when applying these to specific operational contexts.

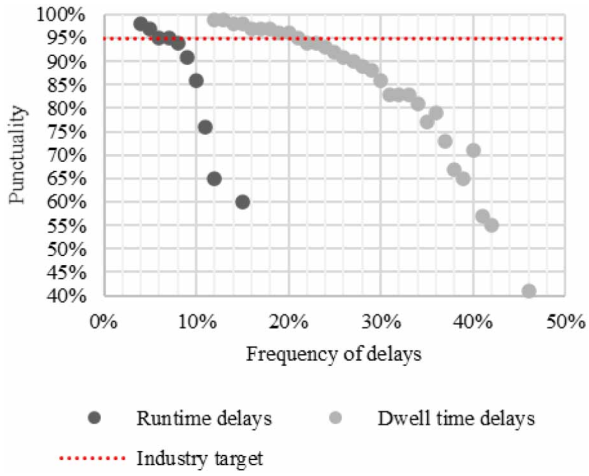

FIGURE 8. Punctuality as a function of frequency of run and dwell time delays respectively. Data from commuter train operations for the years of 2001-2020 across the three major metropolitan regions in Sweden.

C. REGIONAL DIFFERENCES

Fig. 9 gives an overview of delay frequencies (level 3 indicators) and punctuality (level 2 indicators) for commuter trains in the three metropolitan regions of Stockholm, Gothenburg, and Malmö. A quick glance reveals that each region struggles with quite different challenges.

Stockholm (pictured in Fig. 9, top) comes out in the middle in terms of punctuality, with an average of 92%, but it has the largest day-to-day variation with the most frequent and severe dips. It has the lowest level of runtime delays (2%) and struggles mostly with dwell time delays (which happen in about 24% of cases). Gothenburg (Fig. 9, center) serves as a reminder that there is a complex interplay between the processes that result in punctuality: it has more runtime delays (7% of cases) than Stockholm, and even more problems with dwell time delays (36% of stops), but still has the highest punctuality at about 93%.

This is possible because trains in Gothenburg have more margins and are, therefore, in a better position to recover any delays that arise. The cost of this is that time and capacity is wasted when delays do not arise, and that trains often operate ahead of schedule (which can be captured by other Level 3 indicators). Commuter trains in the region around Malmö are the least punctual of the three regions, at 89%. Fig. 9, bottom, shows that these trains are also different from the others in that runtime and dwell time delays are just as likely to occur, at an average of 17%. The frequency of dwell time delays is the lowest across the three regions, i.e., fewer trains are delayed at stops around Malmö than in the other cities.

These differences highlight the importance of analyzing a particular railway based on its own circumstances and challenges, these often vary substantially from place to place. They also help show what is possible. If it is possible to only have a 2% risk of runtime delays in Stockholm, it should be possible to lower that risk substantially in the other regions as well. Operators in Stockholm can perhaps, on the other hand, also learn from those in Malmö regarding dwell times. As we see in that region, there is no law that dwell time delays *must* be more frequent than runtime delays.

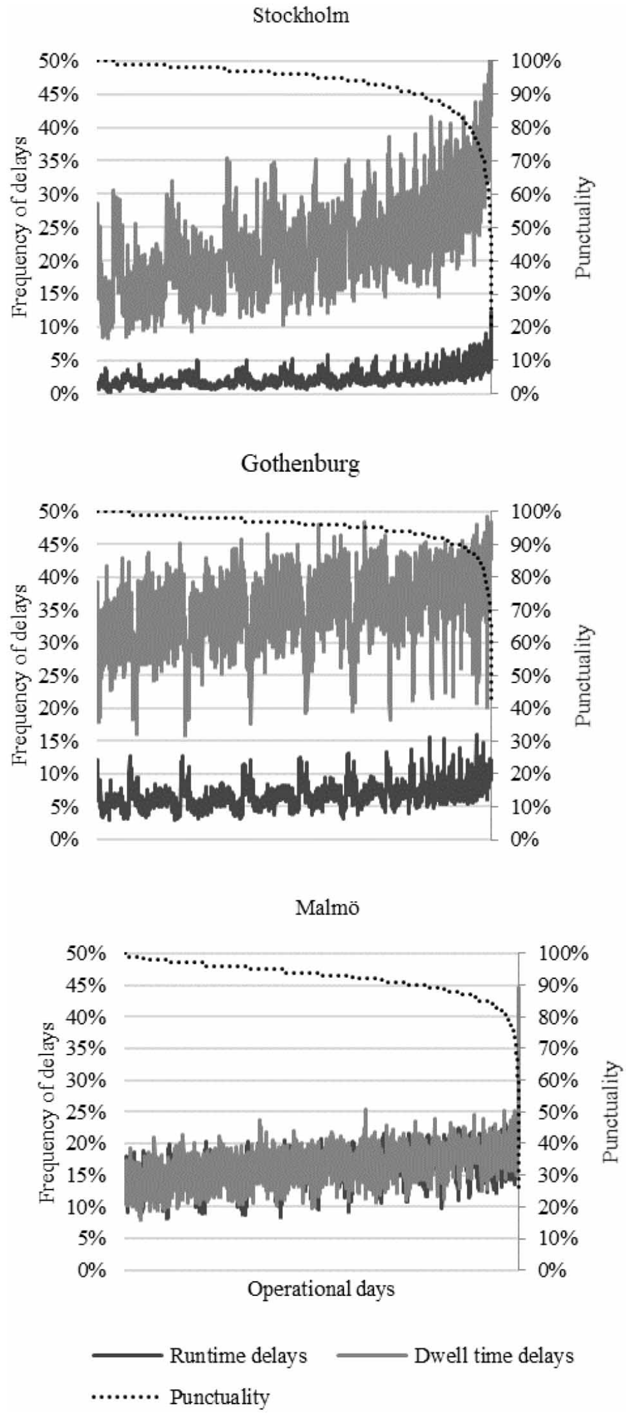

FIGURE 9. Distribution of runtime delays, dwell time delays and punctuality in Stockholm (top), Gothenburg (middle), and Malmö (bottom). Days sorted by punctuality in descending order. Data covers commuter trains in the regions, between 2001 and 2020.

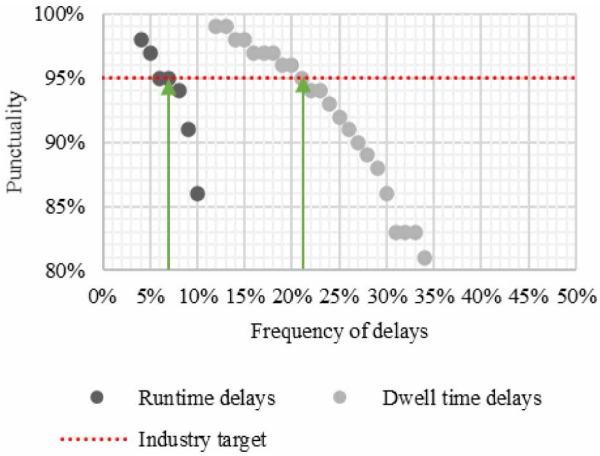

FIGURE 10. Target levels (green arrows) for Level 3 indicators in the pooled data covering all three Swedish metropolitan regions.

D. TARGET LEVELS

To illustrate the target levels for the frequency of run- and dwell time delays that we find in the studied data, we first use the pooled data, and then follow with each of the three regions separately. Figs. 10–11 illustrate how punctuality varies with these indicators, and contains vertical arrows to indicate the target levels, or thresholds, which are acceptable for punctuality to reach the desired level of 95%. Fig. 10 indicates that if punctuality is to reach or exceed 95% across the three regions, then at most 7% of runtimes and 21% of dwell times can be delayed.

Circumstances vary by region, however. Fig. 11, top, shows that commuter trains in Stockholm need to be stricter than average with runtime delays and should instead target a level of at most 2%, while the required maximum frequency of dwell time delays is the same as in the pooled data, at 21%. Fig. 11, center, shows that a target for runtime delays in Gothenburg is 7%. As we saw in the previous section, both these levels happen to correspond to the historical averages. The main challenge in both regions is instead to reduce the frequency of dwell time delays: in Stockholm from an average of 24% to 21% or less, and in Gothenburg from 36% to 19% or less. The case is different in the Malmö-region. Fig. 11, bottom, indicates that at most 11% of runtimes and 14% of dwell times can be delayed if punctuality there is to reach or exceed 95%. Both levels are lower than the average of 17% shown in Fig. 9, so the operator in Malmö would need to address *both* run- and dwell time delays. Unlike in the two other regions, the gap is thus larger for runtimes than dwell times.

E. EFFICIENT FRONTIERS BETWEEN INDICATORS

As we have seen, railways in each of the three regions above have different combinations of targets. This reinforces that one needs to take local circumstances into account when giving recommendations. It also shows that there are several ways to achieve the same target, in this case of 95% punctuality: if there is more of one type of delay, there must

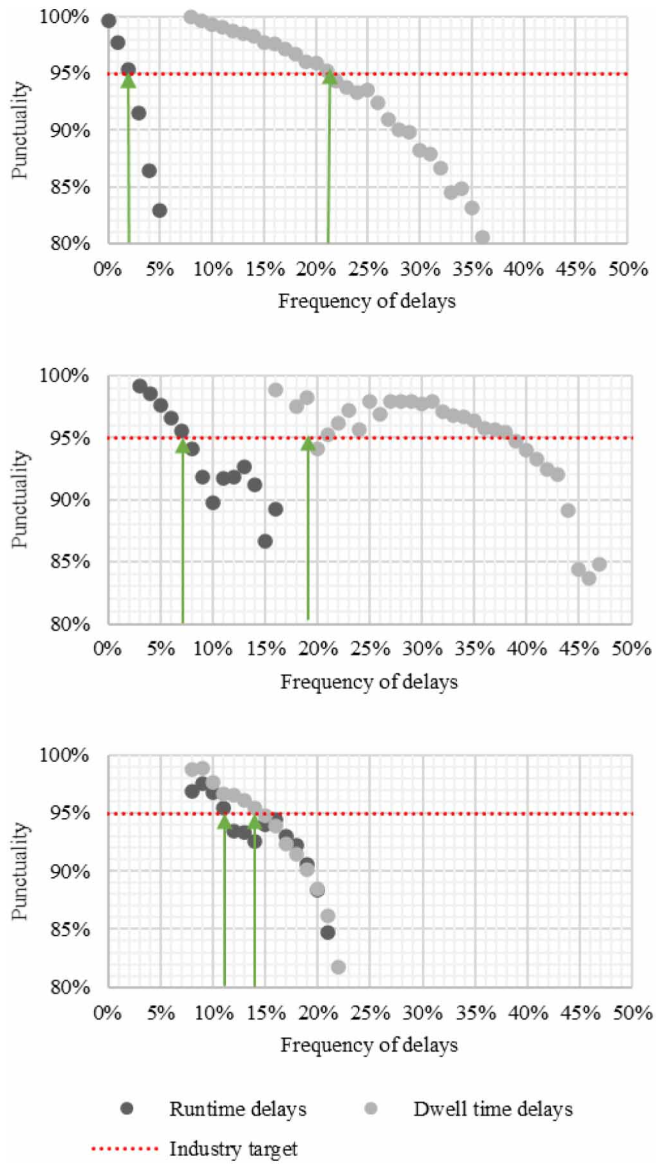

FIGURE 11. Target levels (green arrows) for Level 3 indicators when looking at Stockholm, Gothenburg, and Malmö data separately.

be less of the other, or punctuality will suffer. This is illustrated in Fig. 12, by plotting the three sets of suggested target levels above, with the frequency of run time delays on the horizontal axis, and dwell time delays on the vertical axis.

Together, these recommendations form a sort of efficient frontier between the two indicators. We have also added corresponding levels for 94% and 96% punctuality, for higher levels there is insufficient data, and we would have to extrapolate. In all cases captured in Fig. 12, a given reduction in the frequency of runtime delays seems to go further (on the order of about 40%) in terms of improving punctuality than the corresponding reduction of dwell time delays. On the other hand, dwell time delays are typically (but not always) more common. Which approach is more promising

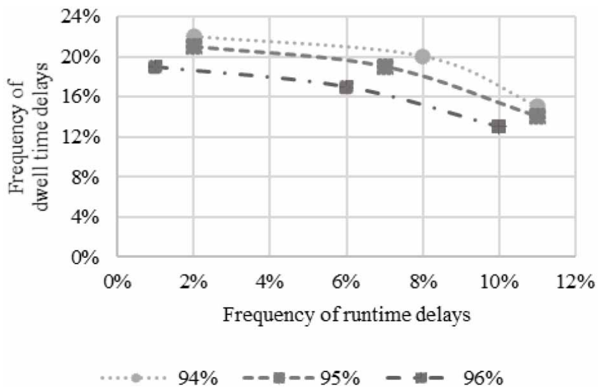

FIGURE 12. Interconnection between frequency of dwell time delays and frequency of runtime delays in reaching different punctuality levels.

will depend on local circumstances and will change over time.

The aim of this paper is not to provide a comprehensive list of measures to reduce the frequency of delays, or to compile evidence for such measures. Still, some examples of how to reduce delays can be enlightening. For runtime delays measures include more targeted maintenance of critical infrastructure components, homogenizing the speed of trains in critical sections, and development of driver advisory systems. For dwell time delays, example measures might instead include platform markings indicating where passengers should stand, more appropriate scheduling of dwell times, and a clear policy to depart precisely on time rather than waiting for late passengers.

Fig. 12 could be extended by adding observations and target levels from railways in other regions, or from other operators, and to other sets of indicators, to get an even better picture of what is required to reach a certain level of punctuality. The same could also be done with another Level 2 indicator, such as passenger punctuality, or punctuality measures defined in other ways (i.e., including intermediate stops or different delay thresholds), and so on. In principle more variables can be considered simultaneously, but an obvious problem is that the number of dimensions makes it difficult to visualize and interpret the results. We believe that it is more practical to consider each of the trade-offs in pairs, as we have done here.

V. CONCLUSION

In this paper we have investigated how the train punctuality improvement work can be taken one step further to close in on punctuality goals set by the railway industry. While substantial efforts have been made to systematize quality work in the railway sector, previous quality work has struggled with a loose connection between the overarching goals (such as high punctuality and customer satisfaction), what has been measured (such as the number of delay hours due to infrastructure failures), and what has been done (such as increased maintenance spending). This loose connection has made it difficult to identify and evaluate concrete improvement efforts.

To tackle this issue, we propose a new four-level framework for indicators relevant for punctuality improvement work. A new third level of indicators measuring the different parts of train operation, such as the frequency of run- and dwell time delays, is intended to bridge the gap between punctuality (for both trains and passengers) and influencing factors (such as quality of infrastructure, rolling stock and timetables, weather, etc.). Using 20 years of operational data from commuter trains in the three large metropolitan areas of Sweden (Stockholm, Gothenburg, and Malmö), we show that there is a clear link between two such indicators, the frequency of run and dwell time delays, and punctuality. We also demonstrate how a railway can use such operational data to identify target levels for these indicators, which must be met if it is to achieve a desired level of punctuality. Finally, we have shown how efficient frontiers between indicators can be estimated, so that railway operators might make informed trade-offs when they set their target levels.

The methods and indicators proposed in this paper are simple and use data that is already collected by railway operators on a daily basis, so that they are easy to implement in practice. If implemented, they should provide a better way to identify and evaluate punctuality improvement efforts and help the already systematic work make faster progress. In addition to improving punctuality, this framework can also help identify instances where trains have unused margins, and ways in which to increase the capacity of the network.

The focus of this paper has been on method development for improved punctuality monitoring. The developed method is applicable to many areas, not only commuter train punctuality, for which the method was demonstrated in this paper. Future work could apply the method also to punctuality monitoring of long-distance rail and rail freight and investigate if there are any principal differences compared to monitoring of commuter train punctuality. We have also focused on two of the 21 possible Level 3 indicators, future work can investigate more of these, considering for instance the ability to recover delays, and the ways in which trains come to be ahead of schedule. Simply adding data from more railway systems from across the world would also help create a better picture of the efficient frontiers and trade-offs between different types of indicators (such as dwell- vs. runtime delays, delays vs. margins, and so on), and what is required to reach higher levels of punctuality.

Rather than arguing for any specific delay reduction measures, our hope with this paper is to facilitate an environment where many small and hands-on improvement efforts can be tried and evaluated in a constructive fashion. The expected impact of any given measure on punctuality, as conventionally measured, is naturally very small, often too small to detect. The indicators proposed in this paper should be better suited to detect such marginal improvements, and thus encourage, rather than discourage, an atmosphere of continuous experiments and improvements from the bottom up.

REFERENCES

[1] J. Bates, J. Polak, P. Jones, and A. Cook, "The valuation of reliability for personal travel," *Transp. Res. E, Logist. Transp. Rev.*, vol. 37, nos. 2–3, pp. 191–229, 2001.

[2] P. Rietveld, F. R. Bruinsma, and D. J. Van Vuuren, "Coping with unreliability in public transport chains: A case study for Netherlands," *Transp. Res. A, Policy Pract.*, vol. 35, no. 6, pp. 539–559, 2001.

[3] M. Wardman, "A review of British evidence on time and service quality valuations," *Transp. Res. E, Logist. Transp. Rev.*, vol. 37, no. 2, pp. 107–128, 2001. Accessed: Sep. 20, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1366554500000120

[4] M. Wardman and R. Batley, "Travel time reliability: A review of late time valuations, elasticities and demand impacts in the passenger rail market in Great Britain," *Transportation*, vol. 41, no. 5, pp. 1041–1069, 2014.

[5] J. Högdahl, M. Bohlin, and O. Fröidh, "A combined simulation-optimization approach for minimizing travel time and delays in railway timetables," *Transp. Res. B, Methodol.*, vol. 126, pp. 192–212, Aug. 2019.

[6] Y. Lee, L.-S. Lu, M.-L. Wu, and D.-Y. Lin, "Balance of efficiency and robustness in passenger railway timetables," *Transp. Res. B, Methodol.*, vol. 97, pp. 142–156, Mar. 2017.

[7] P. Sels, T. Dewilde, D. Cattrysse, and P. Vansteenwegen, "Reducing the passenger travel time in practice by the automated construction of a robust railway timetable," *Transp. Res. B, Methodol.*, vol. 84, pp. 124–156, Feb. 2016.

[8] A. D. McTavish and D. Maidment, "Quality monitoring in British Rail," *Rail Int.*, vol. 20, nos. 4–5, pp. 17–24, 1989.

[9] J. Fahlén and T. Jonsson, "Train punctuality in a new perspective," *Eur. Railway Rev.*, vol. 11, no. 1, 2005. Accessed: Sep. 18, 2018. [Online]. Available: https://www.globalrailwayreview.com/article/2583/train-punctuality-in-a-new-perspective/

[10] M. Veiseth, P. Hegglund, I. Wien, N. Olsson, and Ø. Stokland, "Development of a punctuality improvement method," *TQM J.*, vol. 23, no. 3, pp. 268–283, 2011. Accessed: May 19, 2021. [Online]. Available: www.emeraldinsight.com/1754-2731.htm

[11] A. Graham, S. Shiba, and D. Walden, *Four Practical Revolutions in Management: Systems for Creating Unique Organizational Capability*. Portland, OR, USA: CRC Press, 2001.

[12] C. M. Tam and M. Y. Hui, "Total quality management in a public transport organisation in Hong Kong," *Int. J. Project Manage.*, vol. 14, no. 5, pp. 311–315, 1996.

[13] D. Wealleans, *The Organizational Measurement Manual*, Gower Publ., Aldershot, U.K., 2001.

[14] R. Sreedharan V, V. Sunder M, and R. Raju, "Critical success factors of TQM, Six Sigma, Lean and Lean Six Sigma: A literature review and key findings," *Benchmarking Int. J.*, vol. 25, no. 9, pp. 3479–3504, 2018.

[15] F. Talib and Z. Rahman, "Identification and prioritization of barriers to total quality management implementation in service industry: An analytic hierarchy process approach," *TQM J.*, vol. 27, no. 5, pp. 591–615, 2015.

[16] B. Aquilani, C. Silvestri, A. Ruggieri, and C. Gatti, "A systematic literature review on total quality management critical success factors and the identification of new avenues of research," *TQM J.*, vol. 29, no. 1, pp. 184–213, 2017.

[17] B. Nyström, "Aspects of improving punctuality: from data to decision in railway maintenance," Ph.D. dissertation, Dept. Civil Min. Environ. Eng., Luleå Tekniska Universitet, Luleå, Sweden, 2008. Accessed: Sep. 29, 2021. [Online]. Available: https://www.diva-portal.org/smash/get/diva2:999787/FULLTEXT01.pdf

[18] A. Økland and N. O. E. Olsson, "Punctuality development and delay explanation factors on Norwegian railways in the period 2005–2014," *Public Transp.*, vol. 13, no. 1, pp. 127–161, 2021, doi: 10.1007/s12469-020-00236-y.

[19] N. O. E. Olsson and H. Haugland, "Influencing factors on train punctuality—Results from some Norwegian studies," *Transp. Policy*, vol. 11, no. 4, pp. 387–397, 2004, doi: 10.1016/j.tranpol.2004.07.001.

[20] C.-W. Palmqvist, N. O. E. Olsson, and L. Hiselius, "Some influencing factors for passenger train punctuality in Sweden," *Int. J. Prognost. Health Manage.*, vol. 8, no. 3, pp. 1–13, 2017.

[21] R. B. van der Kooij, A. D. Landmark, A. A. Seim, and N. O. Olsson, "The effect of temporary speed restrictions, analyzed by using real train traffic data," *Transp. Res. Procedia*, vol. 22, pp. 580–587, Dec. 2017.

[22] M. Veiseth, N. Olsson, and I. A. F. Saetermo, "Infrastructure's influence on rail punctuality," *WIT Trans. Built Environ.*, vol. 96, no. 10, pp. 481–490, 2007, doi: 10.2495/UT070451.

[23] Y. Xia, J. N. Van Ommeren, P. Rietveld, and W. Verhagen, "Railway infrastructure disturbances and train operator performance: The role of weather," *Transp. Res. D, Transp. Environ.*, vol. 18, pp. 97–102, Jan. 2013.

[24] G. Zakeri and N. O. E. Olsson, "Investigating the effect of weather on punctuality of Norwegian railways: A case study of the Nordland line," *J. Modern Transp.*, vol. 26, no. 4, pp. 255–267, 2018, doi: 10.1007/s40534-018-0169-7.

[25] "Monthly Punctuality Report Online." TTT. Trafikverket. Sep. 2021. [Online]. Available: https://www.trafikverket.se/resa-och-trafik/jarnvag/jarnkoll–fakta-om-den-svenska-jarnvagen/jarnkoll-pa-punktlighet/ (Accessed: Oct. 19, 2021).

[26] "TTT—Tillsammans för Tåg i Tid—Årssammanfattning 2020 (TTT—Together for Trains on Time—Yearly Summary 2020)." JBS. Trafikverket. [Online]. Available: https://www.trafikverket.se/resa-och-trafik/trafikinformation/punktlighet-i-tagtrafiken/Tillsammans-for-tag-i-tid/ (Accessed: Aug. 17, 2021).

[27] I. Kristoffersson. "Indikatorer för Ökad Punktlighet På Järnväg (Indicators for Improved Railway Punctuality): VTI Rapport 1008." May 2019. [Online]. Available: http://vti.diva-portal.org/smash/get/diva2:1314131/FULLTEXT02.pdf (Accessed: Sep. 9, 2019).

[28] S. Brisère, "Samverkan Mot Ökad Punktlighet: Analys av TTT (collaboration towards improving punctuality: An analysis of TTT)," M.S. thesis, Skolan Arkitektur Samhällsbyggnad, KTH Royal Inst. Technol., Stockholm, Sweden, 2021. Accessed: Aug. 17, 2021. [Online]. Available: https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1581027&dswid=-6197

[29] C.-W. Palmqvist, "Delays and timetabling for passenger trains," M.S. thesis, Dept. Technol. Soc., Lund Univ., Lund, Sweden, 2019. Accessed: Jan. 30, 2020. [Online]. Available: https://lup.lub.lu.se/search/ws/files/70626078/Carl_William_Palmqvist_web.pdf