Determining the Causes of Train Delays - An Automatic Fuzzy Matching Approach

Wan-Jui Lee[1], and Bianca Mulder[2]

[1] *Maintenance Development, Dutch Railways, Utrecht, 3500GD, the Netherlands*
wan-jui.lee@ns.nl

[2] *Fleet Management, Dutch Railways, Utrecht, 3500GD, the Netherlands*
bianca.mulder@ns.nl

Abstract

In order to manage the health of assets, knowledge of conditions on the system level is often required. One of the most common approaches to determine the system conditions is to check the frequency of unplanned or corrective maintenance of considered systems. However unplanned or corrective maintenance usually means the issue has reached an intolerable degree and therefore leaves very limited room to improve the health condition of assets. To depart from the conventional approach of asset health management based on unplanned maintenance, Dutch Railways has decided to focus on operational disturbances, i.e., train delays, due to technical issues of trains in order to determine which system deterioration to investigate at an early stage.

This work introduces the framework of the cause determination system for train delays that has been implemented within Dutch Railways. The cause determination is composed of two parts. One is an automatic fuzzy matching system that will match a delay to the most probable service request assigned by the support center. The causes of 40% of delays can be automatically identified by this method with an accuracy of 95%. This significantly reduces the human hours spent in identifying delay causes. The rest of cause determination is manually carried out by the ***Delay Analysis by Calling*** project.

In this project, the calling team will first call train drivers to ask and record the encountered issues and handling procedures, and then the reliability engineers will determine the causes of delays based on these feedbacks. The calling team therefore serves as a feedback channel for drivers and support center and provides the possibility to analyze on the delay handling and advice given by the support center for continuous improvement and development. Another advantage of integrating a highly accurate and robust fuzzy matching system into an interactive cause identification framework that requires inputs from various business units is that it intrigues interests and builds up trust in data science technology within the organization. This helps the smooth introduction of culture change which often is a critical point in transforming into a data-driven organization, especially for the maintenance industry.

Based on the identified delay causes, Dutch Railways has built a Delay Analysis Dashboard which can provide a good overview of system conditions for various fleets, and also provide more possibilities to avoid operational disturbances.

1. Introduction

For railway operators the performance index in general consists of three main categories, i.e., intensity of use, quality of service, and safety (Duranton, Audier, Hazan, Langhorn, & Gauche, 2017). All these main categories rely on the reliability and availability of trains. To ensure all trains are reliable and safe to operate at the lowest cost, Dutch Railways is continuously optimizing the maintenance schedule to plan when and what to maintain (de Vos & van Dongen, 2015). The focus of this study is to determine the causes of train delays, i.e., loss in punctuality of passenger services which is a major indicator for the quality of service due to train defects (Lee, Tax, & Duin, 2016). This includes an analysis on which components and subsystems are more often causing train delays when encountering a problem. However, train delays due to technical issues are sometimes beyond component/system defects. In fact, one can define the impact of delays as the aggregation of (1) number of delays, (2) primary delay minutes and (3) secondary delay minutes (Lindfeldt, 2012). While the number of delays is relevant to the technical conditions of trains, primary delay minutes, on the other hand, are more often related to the actions taken by the drivers and the ad-

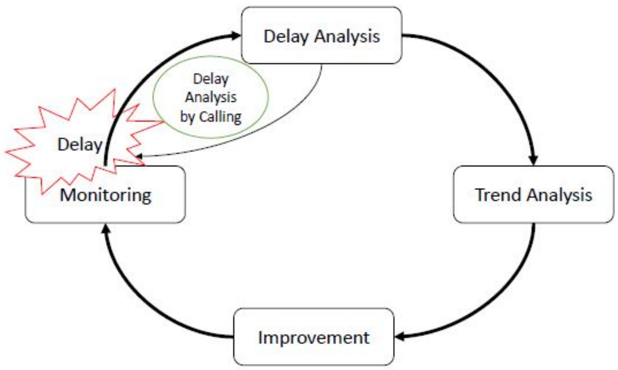


Figure 1. Iterative Delay Management Method.

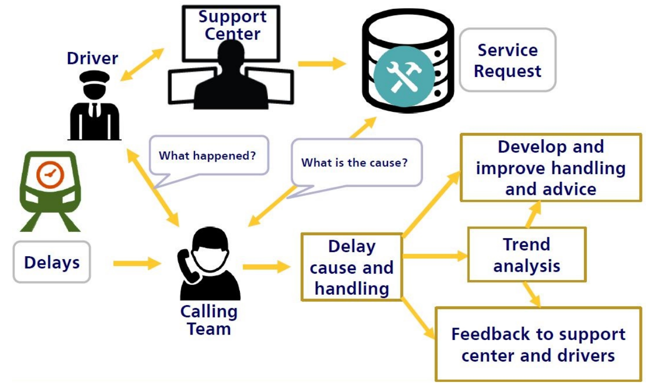


Figure 2. Illustration of the ***Delay Analysis by Calling project***.

vice given by the support center. Secondary delay minutes are mostly related to the logistic choices and less relevant to system conditions, and therefore are considered out of scope in this study. Therefore, by using primary delay minutes, the focus is shifted from a mainly technique-oriented approach (aiming to prevent failures of systems and trains) to a more process-oriented approach (aiming to prevent total impact on passenger). In which the impact is defined as

$$\text{Impact= [number of failures]} \times \text{[primary delay minutes per incident]}.$$

In Figure 1 is the iterative management method implemented by Dutch Railways for the control and continual improvement of delay handling processes. In order to better identify the delay causes and evaluate the handling procedures, the Fleet Management department of Dutch Railways initiated the ***Delay Analysis by Calling*** project in 2016. An illustration of the procedures within this project is given in Figure 2. Members of this project team are in charge of contacting the drivers by phone if a train delay has occurred due to technical issues to figure out which issues had exactly happen and what was the action taken by the driver. When the driver had contacted the support center, the advice given by the support center was also registered. All these efforts of communication can help in developing better advice for the support center, give proper training to the personnel and also perform more accurate trend analysis and health management for fleets on the train component/system level.

To facilitate the communication between the calling team and train drivers about a specific train delay, the calling team will first look for service requests within maintenance records for the specific train on the date that a delay has occurred. Service requests are usually assigned by the support center if a driver has reported some issues and a judgment is made by the support center that a service to the train might be required. However, technicians at the service or maintenance sites can also assign service requests if they have discovered some additional issues when carrying out maintenance tasks. Service requests can provide the calling team with more information about the possible issues before talking to the drivers. However, this is a rather time-consuming step since querying the database of maintenance records usually costs time and it is difficult to filter records on a detailed level. Insufficient filtering of records will also result in several matches and human decisions need to be made to choose the most possible match.

To automatize the matching between train delays and service requests, the Maintenance Development department also participates in this project to develop advanced techniques. Together with Fleet Management, a set of matching rules utilizing the fuzzy logic (Zadeh, 1996) has been developed considering the train number, time range, the type of reporter and the criticality of components/systems to score each possible match. Fuzzy logic has been chosen over other AI techniques due to its direct applicability in translating expert knowledge into an inference system. For each delay, only the match with the maximum score will be automatically reported. Moreover, if the maximum score for a certain delay does not exceed a pre-defined threshold, it will be determined as no match found for this delay. After the implementation of the automatic matching system, for delay logs from January 2017 to February 2018, 40% of delays are automatically matched with service requests in the maintenance records. Considering 41% of delays are not even manually matchable with service requests due to lack of information, the actual detection rate can be considered as $\frac{40}{59} = 67.8\%$ instead of 40% . This significantly reduces the human hours spent in querying maintenance records. Also the automatic fuzzy matching and scoring system is proven to be effective with an accuracy of 95%.

In Section 2, data flow and software components of the ***Delay Analysis by Calling*** project will be introduced and the automatic fuzzy matching and scoring system for matching delays with service requests will be explained in Section 3. The performance of the automatic fuzzy matching and scor-

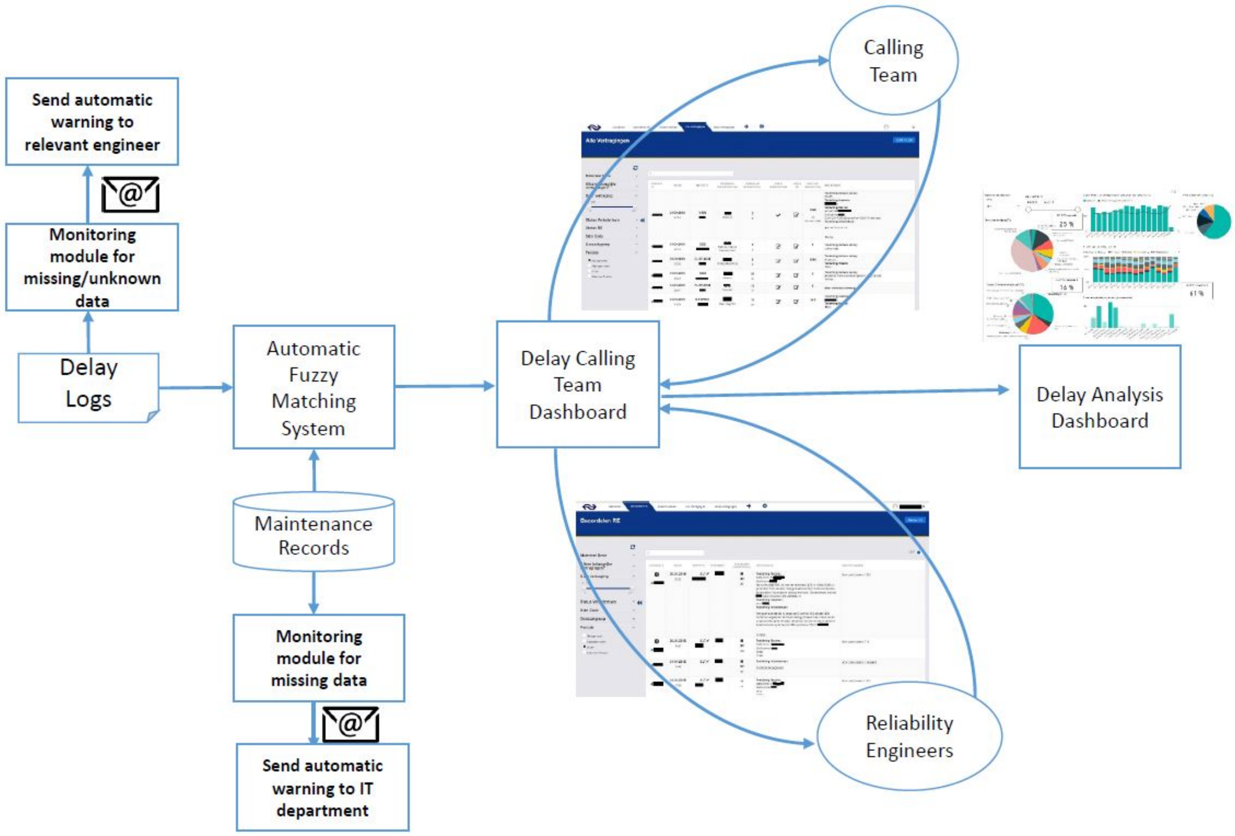


Figure 3. Data flow and software components of the ***Delay Analysis by Calling*** project.

ing system will also be given in the same section. Further analysis on the delay causes for different fleets will be given in Section 4. Conclusions will be given in Section 5.

2. FRAMEWORK OF DELAY CAUSE DETERMINATION

In the previous section, the processes, procedures and goals of the ***Delay Analysis by Calling*** project are presented. In this section the focus will be the data flow and software components of this project as illustrated in Figure 3. Two types of data inputs are available, one is automatic data source which includes delay logs and maintenance records, and the other is manual data which are entered by the calling team and the reliability engineers. Since automatic data source is fixed and cannot be modified, it is important to ensure this type of data source is reliable and continuous. Therefore, as indicated in Figure 3, there are monitoring modules for delay logs and maintenance records separately to automatically detect unknown data or discontinuity in data. If any of these situations has occurred, emails will be sent to the responsible persons automatically. These additional monitoring modules ensure the robustness of data flow and avoid unexpected data blackout to the calling team.

When no issue is found in the automatic data source, delay logs and service requests in the maintenance records will be matched by the automatic fuzzy matching system, which will be explained in detail in the next section. The matching system will link each delay log to 1 or 0 service request to facilitate the delay analysis. Necessary information for contacting drivers and understanding delays such as driver and conductor information, train composition, train type, timestamps, matched service request, reporter, etc., will also be retrieved to feed into the Delay Calling Team Dashboard. The Delay Calling Team Dashboard allows manual inputs and adjustments. It also shows the task status for each delay to remind the calling team and reliability engineers which delays are not processed yet. When the dashboard detects new delays, new entries with necessary information will be automatically generated. For each new entry, the calling team will first call train drivers to ask and record the encountered issues and handling procedures, and then the reliability engineers will determine the causes of delays based on these feedbacks.

All the processed delays together with their identified causes will then be analyzed in the Delay Analysis Dashboard. It can provide health conditions of different systems over various fleets within an adjustable period. A few examples of the reports generated by the Delay Analysis Dashboard will be given in Section 4.

The main contribution of our delay cause determination system includes

- focusing on operational disturbances instead of failure causes to prioritize investigations on system deteriorations at an early stage.
- automatizing the matching between delay logs and service requests with a set of expert-defined fuzzy rules can reach a high accuracy of cause identification and save human hours.
- determining delay causes allows more detailed delay analysis and health monitoring for fleet management.
- building a feedback loop for drivers and support center enables them to be more aware of the delay handling and advice/action to take.
- aggregating the feedbacks to develop and improve advice and trigger personnel to learn from delays.
- embedding an interactive and robust software into different aspects of business units to intrigue interests and build up trust in data science technology within the organization at various levels.

3. FUZZY DELAY MATCHING AND SCORING SYSTEM

If a delay due to technical issues occurs and the driver has contacted the support center to ask for handling procedures, the support center might establish a service request for maintenance to fix the suspected technical issues. Work orders will be constructed in accordance with these service requests and both work orders and service requests will be stored in the maintenance databases.

Note that drivers might not contact the support center when a delay occurs if they consider the issues are directly solvable or can disappear after a certain handling. Also service requests would not always be generated when a driver contacted the support center, especially if the issue was already handled properly and it was decided no further check-up is necessary. On the other hand, service quests can also be generated by technicians who carry out the planned or unplanned maintenance at the service or maintenance sites if they have discovered some additional issues during the maintenance.

Therefore, for a delay log, there might be none, some or several service requests being retrieved from the database of maintenance records. In order to find out which service request is the best match to a delay, a scoring mechanism is required to realize the degree of matching. From experience, the most important information to match between delays and service requests are the train number, time range, the type of reporter and the criticality of components/systems. Following the procedures as described in Figure 4, a set of 8 fuzzy rules are developed by the reliability engineer and data scientist with a few iterations of improvement to establish a reliable scoring system. The design of these procedures follows the guidelines introduced in (ISO 17359:2011, 2011) for condition monitoring but with modifications to fit our application. The designed fuzzy rules are listed as follows:

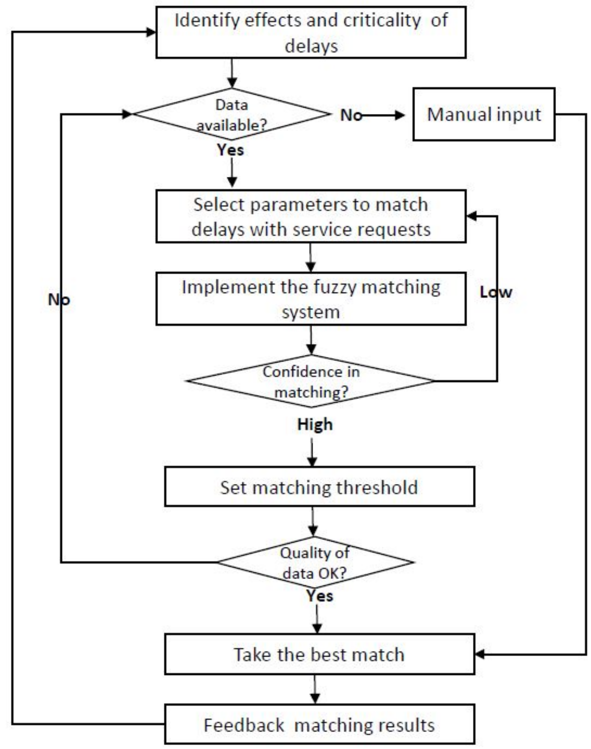


Figure 4. Procedure flowchart for designing the fuzzy matching system.

- Rule r_1: If the train numbers of service request and delay on the same date are the ***same***, then the matching score is ***extremely high***.
- Rule r_2: If the timestamps of service request and delay of a train are ***close***, then the matching score is ***very high***.
- Rule r_3: If the reliability of the reporter is ***high***, then the matching score is ***high***.
- Rule r_4: If the reliability of the reporter is ***medium***, then the matching score is ***medium***.
- Rule r_5: If the reliability of the reporter is ***low***, then the matching score is ***low***.
- Rule r_6: If the criticality of the component is ***high***, then the matching score is ***medium***.
- Rule r_7: If the criticality of the component is ***medium***, then the matching score is ***low***.
- Rule r_8: If the criticality of the component is ***low***, then the matching score is ***very low***.

For rules r_1 to r_8, center of area is used as the defuzzification method to derive a single score for a match aggregated from

Table 1. Number of delays with identified causes and the accuracy of identification

Cause Identification	Number	Accuracy
Automatic Fuzzy Matching	40%	95%
Delay Calling Team Dashboard	59%	N/A

these 8 rules. For example, if k service requests are found on the same date of a delay log for a specific train, the score of service request j, S_j, can be computed for all k service requests.

Please note that the identical train number on the same date will have the largest contribution to the score, and a close proximity in time will also lead to a high score. The type of reporter and the criticality of components/systems will contribute less to the score but they will be helpful in case there are several service requests with the identical train number occur around the same time.

For each delay, only the match with the maximum score will be automatically reported. Moreover, if the maximum score for a certain delay does not exceed a pre-defined threshold θ, it will be determined as no match found for this delay. That is,

$$\text{Best Match} = \begin{cases} \arg\max_{\forall j} S_j, & \text{if } \max_{\forall j} S_j \geq \theta \\ \emptyset, & \text{otherwise.} \end{cases} \quad (1)$$

The value of θ is chosen in such a way that when only rules r_5 and r_8 are fired, the matching score will not be higher than θ. Please note that when there are no rules fired at all, the default matching score will be zero.

The performance of the automatic fuzzy matching system for a total of 20,244 delay logs from January 2017 to February 2018 is given in Table 1. From Table 1 one can see that the automatic fuzzy matching system can identify the causes for 40% of delays, and the rest of 60% will be manually determined by the reliability engineers based on the information filled in by the calling team after communicating with drivers. In the end only for 19% of delays can the causes be further manually identified and the rest of 41% remains undecidable due to lack of train information, unreachable drivers, incorrect train/personnel data and so on. Considering 41% of delays are not even manually matchable with service requests due to lack of information, the actual detection rate can be recalculated as $\frac{40}{59} = 67.8\%$. To summarize, the time saved by the automatic fuzzy matching system is twofold:

1. saving time spent by the calling team for 100% of all delay logs to query and locate the most possible match to a service request, and
2. saving time spent by reliability engineers for 40% of all delay logs to assign the delay cause to a specific system.

In the 40% of delay logs whose causes are automatically identified by the automatic fuzzy matching system, the accuracy of cause identification is 95%. The accuracy is assessed by the calling team and the reliability engineers and concluded as highly effective for operational use.

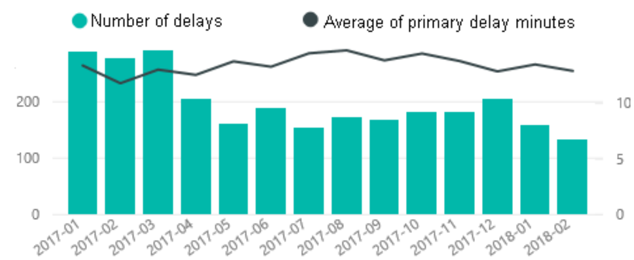


Figure 5. Number of delays and the average of primary delay minutes for the FLIRT trains per month.

4. DELAY ANALYSIS FOR FLEET MANAGEMENT

Based on the cause identification by the automatic fuzzy matching system and the additional manual input as described in the previous sections, sufficient information about delays are obtained to carry out analysis in order to have a better control of train fleets. In the following, some examples of delay analysis that have been included in the Delay Analysis Dashboard are given. Please note that the Delay Analysis Dashboard provides very comprehensive delay analysis and only a small part of it has been shown in this work. The following examples contain the analysis results of different train series such as SLT (Sprinter Light Train) and FLIRT (Fast-moving Llightweight Innovative Regional Train).

(A) Trends of each fleet: Figure 5 gives an example on the trending in delay occurrences for a specific fleet. Particularly, the numbers of delays and the average of primary delay minutes for the FLIRT trains in each month are demonstrated by histograms and a black solid line, respectively. It is clear to see that for this specific fleet, the number of delays has decreased from the second quarter of 2017. On the other hand the average of primary delay minutes remains at the same level.

(B) Difference among different fleets: In the Delay Analysis Dashboard, the difference among different fleets are also investigated. Figure 6 shows the total primary delay minutes among different train types per month. It can be observed that the total primary delay minutes of the SLT train series is usually higher than the other train series.

While example **(A)** and example **(B)** give more general analysis on performance of train fleets with a measure of delays, more detailed analysis on delay causes which is the main purpose of this work can also be carried out. Two examples **(C)** and **(D)** about delay causes analysis are given in the following.

(C) Delay causes of each fleet: In Figure 7(a) and Figure 7(b),

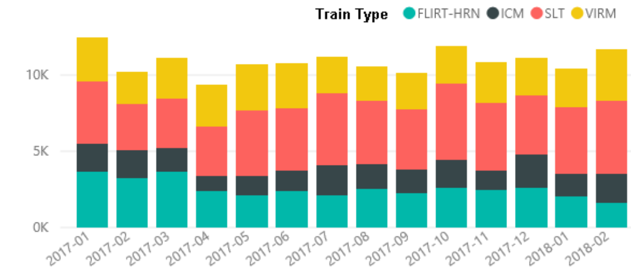
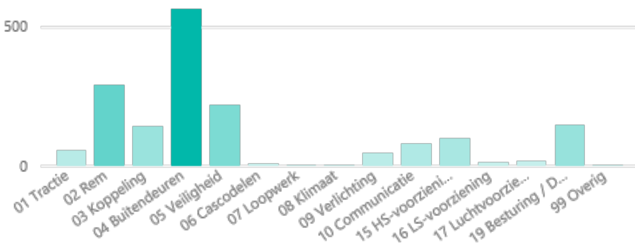
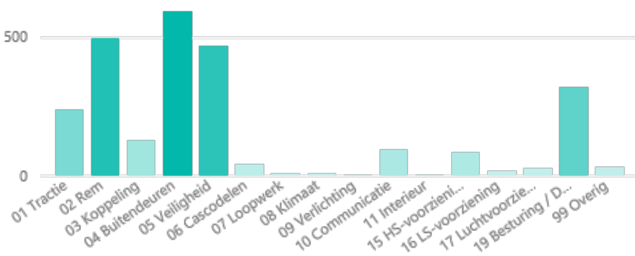


Figure 6. Primary delay minutes among different train types per month.



(a)

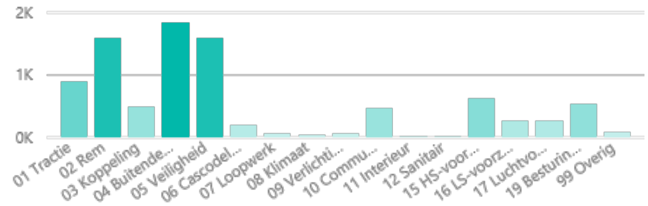


(b)

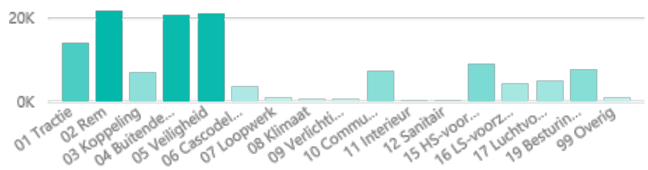
Figure 7. Number of delays caused by different systems in (a) FLIRT and (b) SLT train series, respectively.

are the numbers of delays caused by different systems for FLIRT and SLT train series, respectively. The x-axis are the codes and names of systems in Dutch. These systems will be referred by their codes which are the first 2 digits such as ”01”, ”02”, ..., ”99” as shown in the figures. For FLIRT trains most delays are caused by ”02” and ”04” systems, but for SLT trains the impact of ”05” system is also significant in addition to these two systems. This analysis gives a better insight about various system conditions for different fleets.

(D) Delay impact of each system: By analyzing the delay impact of each system, more insights are provided on which system to focus on. In Figure 8, an interesting example is given. It is clear from Figure 8(a) that the number of delays caused by system ”04” is higher the number of delays caused by system ”05”. However, in Figure 8(b) it suggests the total primary delay minutes caused by system ”04” is slightly lower than those caused by system ”05”. This implies that when a delay is caused by system ”05” issues, it lasts longer compared to system ”04” issues.



(a)



(b)

Figure 8. The total (a) number of delays and (b) primary delay minutes caused by different systems of all trains.

5. CONCLUSIONS

This work introduced the framework of the cause determination system for train delays that has been implemented within Dutch Railways. The cause determination is composed of two parts. One is an automatic fuzzy matching system that matches a delay to the most probable service request assigned by the support center. It was proven to be highly effective and can significantly reduce the human hours spent in identifying delay causes. The rest of cause determination is manually carried out by the ***Delay Analysis by Calling*** project. This work demonstrates the effectiveness and benefits of introducing advanced data science technologies for cause analysis and fleet management. The processes, procedures, data flow and software architecture are also introduced in this work to show the necessity of intensive cooperation among various business units for implementing such a project.

From the cause and delay analysis, the benefits and effectiveness of the automatic fuzzy matching system and the ***Delay Analysis by Calling*** project can be evaluated and validated. Currently Dutch Railways is using the information to monitor the performance of domestic fleets. Due to the success of this project, this framework is planned to be implemented for international fleets as well within the coming few months.

REFERENCES

Condition monitoring and diagnostics of machines - general guidelines. (ISO 17359). (2011). International Organization for Standardization.

de Vos, J., & van Dongen, L. (2015). Performance centered maintenance as a core policy in strategic maintenance control. In *Procedia cirp* (Vol. 38, p. 255-258).

Duranton, S., Audier, A., Hazan, J., Langhorn, M. P., & Gauche, V. (2017). *The 2017 european railway performance index* (Tech. Rep.). The Boston Consulting Group.

Lee, W. J., Tax, D. M. J., & Duin, R. P. W. (2016). Beyond condition-monitoring: Comparing diagnostic events with word sequence kernel for train delay prediction. In *Proceedings european conference of the prognostic and health management society 2016* (p. 162-167).

Lindfeldt, A. (2012). *Congested railways influence of infrastructure and timetable properties on delay propagation*. Stockholm: KTH Royal Institute of Technology.

Zadeh, L. A. (1996). *Fuzzy sets, fuzzy logic, fuzzy systems*. World Scientific Press.