Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 52-3 (2019) 1–6

A reduced phase-space approach to analyse railway dynamics ⋆

Mark M.Dekker *,** **Debabrata Panja** *,**

* *Department of Information and Computing Sciences, Utrecht University, Utrecht, the Netherlands*

** *Centre for Complex Systems Studies, Utrecht University, Utrecht, the Netherlands*

Abstract: We introduce an approach of representing the dynamics of delays in railway systems by Eulerian aggregation of individual train delays onto the railway's network segments. The approach lends itself naturally to the analysis of multi-scale spatio-temporal evolution of delays — colloquially known as the spread of "oil-stain" on the network. To illustrate the potential of this representation, we apply it to describe the dynamics on a national (also referred to as the macro-) scale of the Dutch railways: specifically, we perform a principal component analysis of the Euler-aggregated global delay data, identify two key principal components to define a reduced phase-space, and discuss the delay evolution for the system as a moving point in this phase-space on a few example days.

Keywords: Railways; Phase space; Dynamical Systems; Delay spread

1. INTRODUCTION

The simplest description of a railway system is in terms of trains travelling on designated routes on a network following a planned timetable, acting as a reference measure for train operations. The real-time operations, however, deviate from the timetable for a variety of reasons, leading to delays. Delays can be viewed from different perspectives, depending on the source, nature and impact. See Table 1, where we denote causes of the propagation of delay on railways at different scales. Nevertheless, since delays are by definition unwelcome, naturally, a large volume of existing railways literature is devoted to the analysis and dynamics of delays.

	Micro-space	Meso-space	Macro-space
Micro-time	Single delay train activity	First-order delay propagation	Uncorrelated noise
Meso-time	Short-term advection	Delay propagation to secondary trains	Nation-wide delay propagation
Macro-time	Long-term advection	Long-term delay signature	Nation-wide large disruptions

Table 1. Scales and nature of railway delays.

Given that delay is by definition induced by the trains themselves, the existing literature is dominated by delay models associated with individual trains; for example, tracking the movement and stand-stills of individual trains (Li et al., 2016; Kecman and Goverde, 2015a,b). This approach, which is essentially the Lagrangian description (of delay dynamics) in physics parlance, has been quite successful in determining the statistics of delay on a certain route or train, leading to the enhancement of routes, infrastructure and time tables (Trap et al., 2017; Schöbel, 2012). Other examples of Lagrangian description of delay are the treatment of the trains as agents (Gambardella et al., 2002), or even neural-network prediction algorithms (Oneto et al., 2018; Cerreto et al., 2018). A large volume of simulation tools, too, focus on the agent-based railway modelling approach, and thereby take the Lagrangian approach to track delays (Middelkoop and Loeve, 2006; Middelkoop et al., 2012). In the language of Tab. 1, all of these studies can be classified as "micro-space scale"; within this spatial scale, researchers have focused on various temporal scales, e.g., local short-term delays have been analysed by considering specific delay events and the impact on the directly involved trains. Advection processes (delays moving in space by travelling along with a train) at micro-space scale have also been analysed while considering longer time-scales.

The extension of the Lagrangian approach to analyse propagation or spread of delays over the network at meso- and macro-space scales is difficult because of the (prohibitive) computational burden and the heterogeneity of interactions among trains. Note that these scales are concerned with *interactions at a higher level*, namely the spread of delay from one train towards others (possibly in different directions) that are not directly associated with the line followed by the train in question. Indeed, the analysis of interaction effects needs to involve multiple trains, since these effects are not only highly dependent on the particular configuration of infrastructure and route-planning, but are also dependent on crew-scheduling (e.g., involving train-to-train crew transfers at large railway junctions). In other words, these effects are strongly space- and time-dependent. This provides us with the key indication that it

⋆ We gratefully acknowledge financial support from the Netherlands Organisation for Scientific Research (NWO), the Dutch Railways (NS) and the Dutch railways infrastructure company ProRail. The data for this study has been provided by ProRail.

10.1016/j.ifacol.2019.06.001

would be more suitable to take a Eulerian approach — i.e., map the dynamic variables (for the case at hand, delays) on a fixed spatial grid — such that one can describe their time-evolution in space at larger length scales, rather than train by train.

In Tab. 1 we have split the large(r) length scales into meso- and macro-spaces. The former relates to processes that, although affecting other lines than the spatial location of the primary delay source, spread to regional scales (for instance, problems in regions around a specific station, or a specific track — while all trains that pass through this region are being delayed because of it, and spreading it across their lines), the delay generally "melts" further away from the delay source. In contrast, macro-space problems concern nation-wide disruptions, which generally relate to longer time-scale problems.

2. EULERIAN AGGREGATION OF DELAY

We now describe the aggregation process of delays (defined as the difference between the realised and the planned times of each train activity) on the infrastructure network. Delay is referred to as 'pressure' on the system, affecting its functionality of the system (Monechi et al., 2018). Note also that delay alone does not contain all information on the system's behaviour w.r.t. the timetable: e.g., on severely disrupted days, a lot of train activities get cancelled, meaning that train activities are removed from the schedule, putting the delay contribution of those activities to zero. This effect reduces the delay, while the severity of the disruption could actually be worse.

The data we use for our analysis has been obtained from the Dutch railway network (July 1st 2017 - June 30th 2018). The Dutch railways consist of 801 sensors in the tracks (near passenger stations, but also in other parts of the tracks) that keep record of all trains passing by, while logging the planned and realised time. In the Dutch railway system, not only the amount of delay is logged, but also whether it concerns a departure, passing or arrival. To make the data uniform over the entire network, at any given time, we aggregate all logging activities on to the segments (tracks between sensors, 1438 in total) that the trains should be travelling on. To be specific, we aggregate the departure and arrival activities respectively on the segment that the designated train moves to and where it comes from, in the following manner (see Fig. 1).

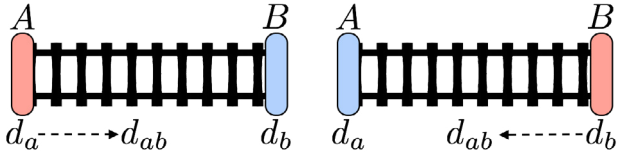

Fig. 1. Transformation of logged delay at nodes to segmental delay. Consider a train travelling from node A and to node B, where delays d_A and d_B are logged. Left: departure logging at A, meaning that we 'forward' this logged delay in space towards segment AB. Right: arrival logging at B, we aggregate this event also to segment AB.

The logs are discrete events. To get a real-time progression of delays on the network, we need to process the logs to generate continuous time series data. At the logging events (e.g., a departure event at node A in Fig. 1), both the planned (t_p) and realised (t_r) times are recorded. At the moment of logging, the delay carried by this train is obviously $t_r - t_p$. But between t_r and t_p, the train is already delayed and consequently affecting the segment, although the train is yet to enter the segment. This leaves us with two choices to deal with the delay: (a) keep the delay with the train, i.e., keep it on the segment where the train presently is, or (b) put the delay on the segment where it is planned to be. We use the latter, since it more accurately indicates the effect of the delay on the network: e.g., when a train is an hour late at a given segment, it already causes problems before it actually gets there (an hour later than scheduled). Mathematically, we define the delay contribution $d_i^j(t)$ of train j on each segment i at time t:

$$d_i^j(t) = t - t_p \tag{1}$$

for $t_p < t < t_r$ ($d_i^j(t) = 0$ otherwise), where t_p and t_r are the planned and realised time of the train activity. Intuitively, the above definition involves the build-up of delay (one second per second) when a train should be at the segment while it is late, and disappears from that segment when the train exits it, giving rise to a sawtooth pattern of the delay. (Note also that once the delay disappears from a given segment, unless it happens to be the last segment of service for that train, the delay simply continues on the next segment.) Note that a segment can have multiple tracks, so that multiple trains can travel on the same segment at the same time. Moreover, multiple trains can build up delay on the same track at the same time, as they can build up delay on a segments without being there (if they should have passed that track already). We therefore compute the total delay $d_i(t)$ on segment i at time t, by summing d_i^j over all trains j (both directions) as

$$d_i(t) = \sum_j d_i^j(t) \tag{2}$$

In this form, the equation for $d_i(t)$ contains a discontinuity when t reaches t_r. This is unwanted, since for prediction purposes, we need to treat the system as a continuous dynamical system. We therefore use a Gaussian weighted running window. This means that the aggregation is done on a 5-second time resolution, which, for computational reasons, is further reduced to 1-minute time resolution (although the results do not differ much as one can intuitively already estimate from the definition of delay in Eq. 1).

We illustrate the above procedure in Fig. 2. Panel (a) shows the delay on the segment from station Utrecht Vaartsche Rijn towards Utrecht Central Station on July 1st 2017. There are no planned train services in early morning until shortly after 06:00 AM, for which we define the delay to be zero. Indeed, as soon as trains start running on this segment, sawtooth signatures start appearing - for a delay of one minute this already happens. It is visible that no delay buildup occurs at larger timescales on this segment of the network. Note that there are only few

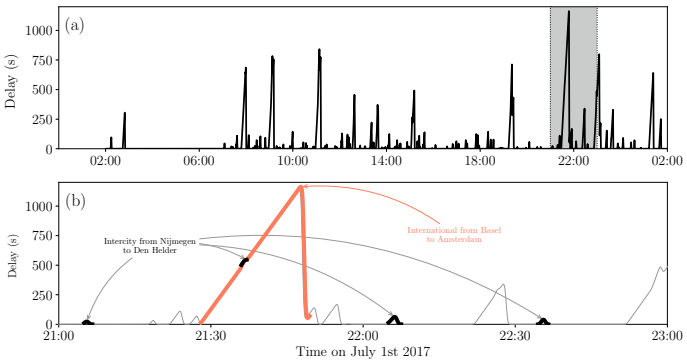

Fig. 2. (a) Time series showing the aggregated delay on the segment from station Utrecht Vaartsche Rijn towards Utrecht Central Station on July 1st 2017. Panel (b) shows a zoomed version (as depicted by the greyish area in the upper panel) between 21:00 and 23:00, wherein the sawtooth delay signatures are clearly visible and several spikes are interpreted.

'follow-up' on large spikes, indicating that there is little 'memory' on this segment.

Panel (b) shows a zoomed-in version of panel (a), revealing the sawtooth patterns. This indeed allows for the interpretation of each sawtooth in terms of individual trains. In black, we highlight small delay 'bumps' caused by a domestic train line that is scheduled on this trajectory every 30 minutes, travelling from Nijmegen to Den Helder. The bumps are small because every time during this period, trains from this line where only a few minutes late. In contrast, the red-highlighted sawtooth reflects a train that is about 20 minutes late, being the maximum attained delay on this day for this segment. This was caused by an international train that travelled from Basel to Amsterdam.

2.1 Aggregated delay at the global scale

Once the delays have been accumulated per segment, they can be pieced together to provide the 'delay snapshot' at the global (in this context, national) scale from the Eulerian perspective. An example is shown in Fig. 3: at every timestep, the system has a delay value at every point in space. This allows for a fast evaluation of where delays in the network are located.

Figure 3 serves as a classic illustration of how the (Eulerian) aggregation of delay on the network segments lends itself naturally to propagation and spread of delays in time (or, in colloquial terms, 'spread of the oil-stain'): the chain reaction of delayed trains spreading delays onto new trains by means of incapacitating railway tracks and delaying crew members. On February 3rd 2012, the Dutch railways suffered a nation-wide railway disruption. Around 10:00 in the morning, small failures in the infrastructure led to the delay of trains near Rotterdam and Amsterdam, which quickly spread and amplified up to point that trains in the whole country were strongly delayed. (The delays lasted until the following day.)

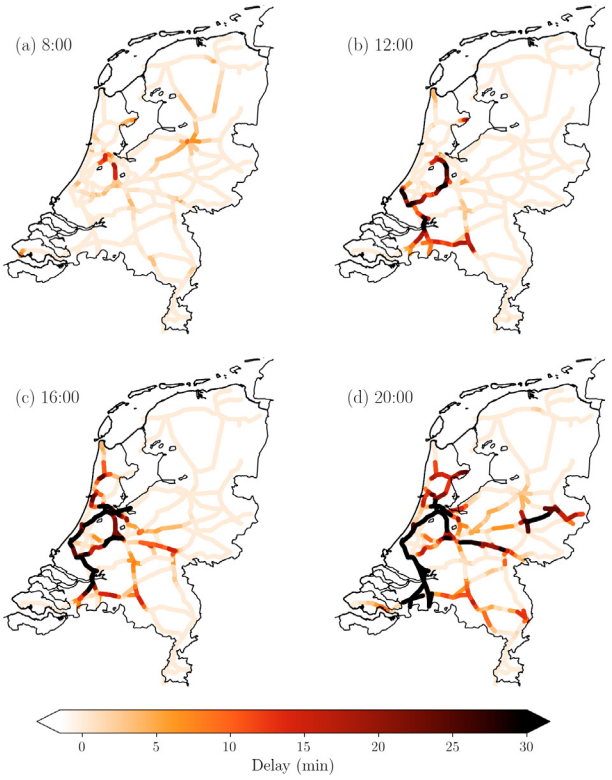

Fig. 3. Aggregated delays on the Dutch railway network, at four different timestamps on February 3rd 2012. Thin black lines depict the coastal and country borders of the Netherlands. Figure adapted from Dekker et al. (2018).

3. REDUCED PHASE-SPACE FOR DESCRIBING DELAY EVOLUTION

The analysis of the 1438 time series of delay is burdensome for computation and interpretation purposes. We therefore reduce the dimension of the time series by applying principal component analysis (PCA) (Pearson, 1901).

We have performed the PCA on the data for the 'disrupted days' (as classified by ProRail) between July 1st 2017 - June 30th 2018, revealing that, indeed, only a few of the principal components (PCs) are relevant, as presented in Fig. 4 (the results have been checked for robustness under various time steps, time periods (seasons) and selections of days based on severity, and are not discussed here further). The explained variance for every PC in the year-long dataset is shown in panel (a), revealing that the first two PCs stand out in terms of the variances they carry. Similarly, panel (b) shows the characteristic decay times for the autocorrelation function of each PC, indicating their persistence in time. The combination of the two panels provides us with the cue that a two-dimensional time-series, based on PC1 and PC2, is a minimalistic, and yet optimal vehicle for describing the delay evolution in global scale.

Stated differently, PC1 and PC2 can be used to define a reduced (two-dimensional) phase-space, wherein the amplitudes of the principal components, as a function of time, describe the evolution of the global delays. These amplitudes are calculated by taking the dot product of the global

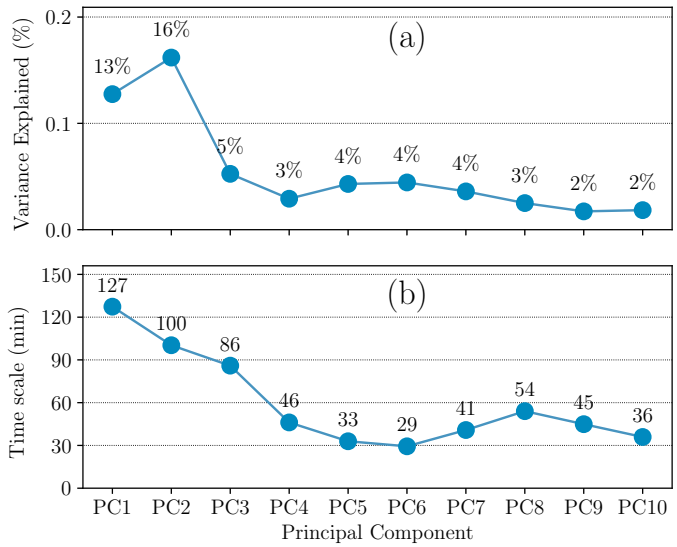

Fig. 4. Variance explained (a) and timescale of autocorrelation decay rate (b) for the first ten principal components.

delay snapshot (a 1438-dimensional vector) with PC1 and PC2 (also 1438-dimensional vectors). In this way we can track and monitor the railway system's delay behaviour in the (reduced) phase-space at all points in time (next section). Since by definition there is no delay outside the service hours, the system's delay position starts and ends at the origin (0,0) in this phase-space.

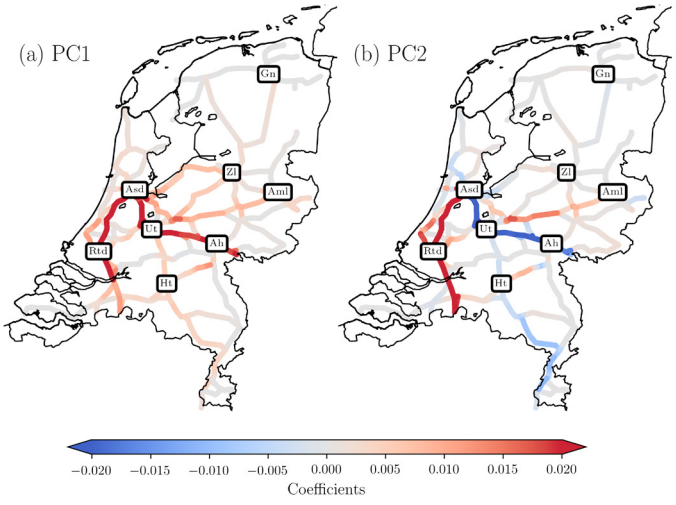

Fig. 5. The spatial maps of the PC1 and PC2 (also called empirical orthogonal functions or EOFs) for the Dutch railway system, explaining respectively 13% and 16% of the variance over a full year. A running spatial average-smoothening is applied for visualisation reasons. Abbreviations refer to important passenger stations: Amsterdam Central (Asd), Rotterdam Central (Rtd), Utrecht Central (Ut), Arnhem (Ah), Groningen (Gn), 's-Hertogenbosch (Ht), Amersfoort (Amf) and Almelo (Aml).

The PCs can be interpreted by considering their 'spatial maps', also called Empirical Orthogonal Functions of EOFs, and are shown in Fig. 5. Both PCs contain large-magnitude coefficients (related to strong covariance) on important international lines from Amsterdam towards Belgium and Germany. All the elements of PC1 are positive, while PC2 has negative elements in the middle of the Netherlands, i.e., between Amsterdam, Utrecht towards Arnhem and Germany. For every point in the phase-space, the coordinates are the amplitudes of these two PCs, and they denote how much of the EOFs are present in a delay snapshot; using these the evolution of the global delay can be deduced back on the physical rail network. This allows for the analysis of the macro-dynamics in a strongly reduced space.

4. PHASE-SPACE DYNAMICS

Figure 6 shows four examples (days) of the system's evolution in this phase-space. The smaller sub-panels in the top right corner of each sub-figure indicate the evolution of the total (summed) delay on the *entire* network. Delays in panels (a) and (b) are much weaker than in (c) and (d). Every minute of any day is represented by a dot in the phase-space. As the exact combination of the PCs can be given, by inference from Fig. 5, as estimate of where delay is occurring, the phase-spaces as in Fig. 6 provide a temporal map of the dynamical evolution of the buildup, peak moments and decay of delay of the particular days shown. We can simultaneously plot a subpanel of the total delay on these days, and color the dots with the same total delay information, to get an idea of the severity of each area in the phase-space. These plots can, for example, be used to make predictions, or to filter statistically significant trajectories in the phase-space.

The first panel (Figure 6a) shows a so-called 'green' day, classified as a day on which the railway system in its entirety suffered barely any delay, which is reflected in the fact that the system stayed relatively close to the origin in the phase space. Two (minor) events in the phase-space evolution are however noticeable, coinciding with the small delay maxima in the delay evolution subpanel. They are highlighted in the figure (numbered I and II). Event I occurred around 10:00 AM and involved a movement of the system towards the upper-right side of the phase space, indicating larger positive-PC1 and positive-PC2 values. This area in the phase-space (PC1 positive and PC2 positive) can be interpreted by looking at the coefficients in Fig. 5. It turns out that this area is related to delay between Amsterdam, Rotterdam and towards Belgium. The second event (II) relates to the movement of the system towards the upper-left side of the phase-space. This area can be attributed to delay in the center/east of the Netherlands: between Amsterdam, Utrecht and towards Germany (again by looking at the coefficients in Fig. 5). Going back-and-forth from the phase-space to physical space allows for the usage of these diagrams to analyse railway dynamics in two dimensions.

Moving over to another example of a day's dynamics in the phase-space, Fig. 6(b) shows a more representative 'average' day. There is delay at multiple occasions, but they are not exceptional. This results in the state being kept roughly in the center of the phase-space.

In contrast, in Fig. 6(c) the system quickly moves away from the origin. Apart from a relatively average episode

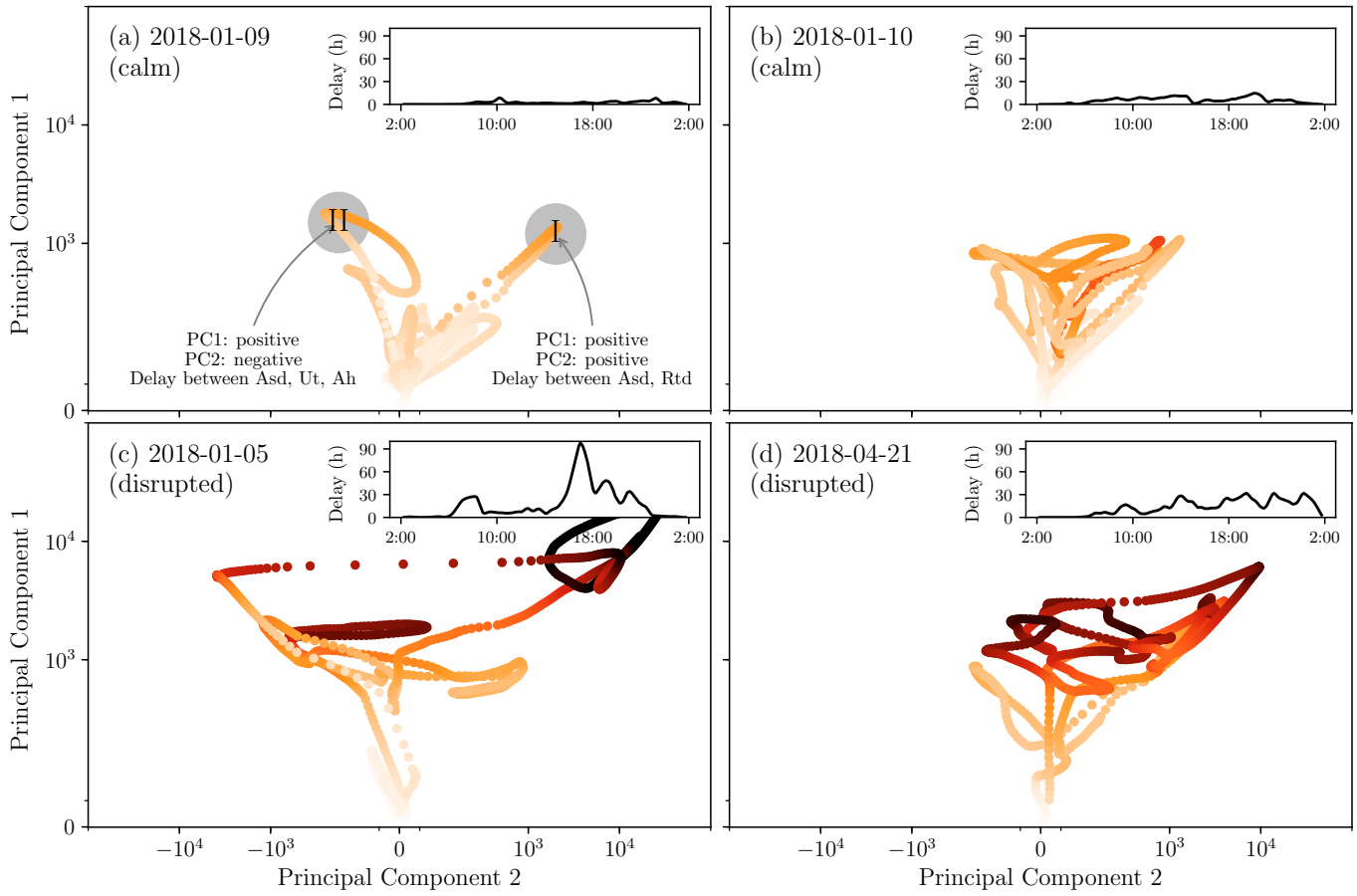

Fig. 6. Phase-spaces with amplitude of principal component 1 versus amplitude of principal component 2 for (a) January 7th 2018, (b) January 8th 2018, (c) January 3rd 2018, and (d) April 19th 2018. Dots refer to the position of the system within the phase-space every minute. The dot colouring indicates the total delay. Subpanels show the evolution of the total delay (summed across the whole network) throughout the day. Two events related to somewhat stronger delays are denoted by I and II in panel (a). Abbreviations refer to the cities of Amsterdam (Asd), Utrecht (Ut), Rotterdam (Rtd) and Arnhem (Ah).

between 10:00 and 16:00, the system attains high values of delay, moving far away into both upper right and left corners of the phase-space. The movements in the phase-space are also fast, covering many different parts of the phase-space, pointing to the fact that delays occurred many locations on the network.

Another disrupted day is shown in Fig. 6(d). Here the system's movements remain in the positive-PC2 side, reflecting delays in the southwest of the Netherlands, around Amsterdam and Rotterdam.

5. CONCLUDING REMARKS

In this paper we have introduced the Eulerian perspective of delays for railway systems. As an example, we have considered the macro-scale delays for the Dutch railway system, and have identified a reduced two-dimensional phase-space to capture the dynamics. We have further analysed four example days using this framework, distinguishing the more severely disrupted days from days that are calm in terms of delay, and providing the corresponding real-space interpretations of the delays in this reduced phase-space.

The principal components analysis provides statistically robust covarying patterns that specifically allow the analysis at larger spatial scales. However, this comes at the price of less specific localization, as the PC-coordinate system only gives a rough estimate of where the delay is situated. To have more specific localization, one could focus on lines or areas to apply the dimensional reduction to, rather than to the whole country. It is also important to note that not all delay variance is captured within only the first two principal components (see Fig. 4). Nevertheless, the proposed perspective of delay and dimensional reduction potentially opens up the analysis of delay dynamics at a multitude of length and time-scales. The phase-space perspective can be used for diagnostic analysis of the past (e.g., to identify parts in the phase-space that are prone to delay amplification), and for predictions of how the system will evolve.

ACKNOWLEDGEMENTS

The authors thank Henk Dijkstra and Stefan Dekker for the discussions on the reduced phase-space analysis, and the Dutch railway organisations NS and ProRail for the data and further valuable information.

REFERENCES

Cerreto, F., Nielsen, B.F., Nielsen, O.A., and Harrod, S.S. (2018). Application of Data Clustering to Railway Delay Pattern Recognition. *Journal of Advanced Transportation*, 2018, 1–18. doi:10.1155/2018/6164534.

Dekker, M.M., Lieshout, R.v., Ball, R.C., Bouman, P.C., Dekker, S.C., Dijkstra, H.A., Goverde, R.M., Huisman, D., Panja, D., Schaafsma, A.A., and Akker, M.v.d. (2018). A next step in disruption management: combining operation research and complexity. *Conference on Advanced Systems in Public Transport and TransitData proceedings*. URL `https://repub.eur.nl/pub/109054/`.

Gambardella, L.M., Rizzoli, A.E., and Funk, P. (2002). Agent-based Planning and Simulation of Combined Rail/Road Transport. *SIMULATION*, 78(5), 293–303. doi:10.1177/003754970207800551.

Kecman, P. and Goverde, R.M.P. (2015a). Online data-driven adaptive prediction of train event times. *IEEE Transactions on Intelligent Transportation Systems*, 16(1), 465–474. doi:10.1109/TITS.2014.2347136.

Kecman, P. and Goverde, R.M.P. (2015b). Predictive modelling of running and dwell times in railway traffic. *Public Transport*, 7(3), 295–319.

Li, D., Daamen, W., and Goverde, R.M. (2016). Estimation of train dwell time at short stops based on track occupation event data: A study at a Dutch railway station. *Journal of Advanced Transportation*, 50(5), 877–896. doi:10.1002/atr.1380.

Middelkoop, A.D. and Loeve, L. (2006). Simulation of traffic management with FRISO. In *WIT Transactions on the Built Environment*, volume 88 of *WIT Transactions on The Built Environment, Vol 88*, 501–509. WIT Press, Southampton, UK. doi:10.2495/CR060501.

Middelkoop, D., Steneker, J., Meijer, S., Sehic, E., and Mazzarello, M. (2012). Simulation backbone for gaming simulation in railways: A case study. In *Proceedings - Winter Simulation Conference*, 1–13. IEEE. doi:10.1109/WSC.2012.6465195.

Monechi, B., Gravino, P., Di Clemente, R., and Servedio, V.D. (2018). Complex delay dynamics on railway networks from universal laws to realistic modelling. *EPJ Data Science*, 7(1), 35. doi:10.1140/epjds/s13688-018-0160-x.

Oneto, L., Fumeo, E., Clerico, G., Canepa, R., Papa, F., Dambra, C., Mazzino, N., and Anguita, D. (2018). Train Delay Prediction Systems: A Big Data Analytics Perspective. *Big Data Research*, 11, 54–64. doi:10.1016/j.bdr.2017.05.002.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559–572. doi:10.1080/14786440109462720.

Schöbel, A. (2012). Line planning in public transportation: models and methods. *OR spectrum*, 34(3), 491–510.

Trap, M.L., Huisman, D., and Goverde, R.M.P. (2017). Assessment of alternative line plans for severe winter conditions in the netherlands. *Public Transport*, 9(1-2), 55–71.