

# What Can Be Learned From (Public) Running Result Data?

Tobias Kretz

PTV – Planung Transport Verkehr GmbH, Karlsruhe, Germany,  
tobias.kretz@ptvgroup.com

Received: 4 September 2023 / Last revision received: 27 May 2024 / Accepted: 28 May 2024

DOI: [10.17815/CD.2024.173](https://doi.org/10.17815/CD.2024.173)

**Abstract** Results from running races is available in abundance. In this contribution it is shown, how this data might help to understand pedestrian dynamics in general, as well as the situation at the start and experience for runners.

**Keywords** Pedestrian flow · marathon · fundamental diagram · social force model

## 1 Introduction

Marathons and other running events have become a common global phenomenon. Distances range from five to several hundred kilometers. Some events like New York or Berlin marathon attract people from all over the world and world class runners almost annually marking new world records, while other events focus on fun. Many, even so-called “fun runs” have in common that runners’ time is measured with reliable and precise measurement technology. This means that daily new data on pedestrian dynamics is produced and published. Obviously, this data is obtained under very specific conditions and published properties are different from what one would specify as observables in a scientific study. However, the data is free and huge. These are two advantages over lab and field data. What can be learned from it?

While there is an abundance of medical literature related to long-distance running, it is a bit surprising that only few works are concerned with crowd dynamic aspects of running courses. Tomoeda and Yanagisawa discussed the optimal density at the start of a running event [1]. Rodriguez et al analyzed the evolution of the field of runners for the entire race course [2]. Treiber et al proposed a microscopic model for a long distance skiing race [3]. Pennisi et al. made use of Rome 2013 marathon start as a test ground for computer vision methods [4] Bain and Bartolo evaluated the dynamics of and within starting corrals

at various marathons and formulated a macroscopic model based on this [5]. Recently, Anagnostopoulos took a different perspective on the crossroads of running and traffic dynamics by discussing running as a commuting mode [6, 7].

## 2 Methodology, Data Source, and Nature of the Data

Here, the result list of Badenmarathon 2019 [8] is used (marathon and half-marathon). The start line was located on Hermann-Veit-Straße near the beginning of the cycling lane. At the start line the road is about 6 m wide. A few meters after the start the road widens<sup>1</sup>. Marathon and half-marathon runners started together. Fields of both races were mixed upstream of the start line. Runners were asked to position according to their speed. This was assisted with four blocks A to D marked with flags in the starting area. 5006 runners were registered for half marathon and 1000 for marathon of whom 4223 and 846 finished.

Times were measured with a UHF RFID sticker attached to the number bibs with stated precision of 0.2 s [9–12]. The result lists usually show – aside personal data of runners – the difference between the two times at the finish and the start line – this is called the “chip time” or “net time” – as well as the difference between the time at the finish line and the time of the start signal – the “gun”, hence this time is called the “gun time” or “official time”.

Empirical data will be compared with simulation results for which PTV Viswalk version 2023 [13] was used. It utilizes a combination of circular and elliptical specification II of the Social Force Model [14] with some modifications (one modification is that only a limited number of pedestrians are taken into account when calculating forced on another pedestrian, another modification is that the force contribution from elliptical variant II does not make use of the continuous  $w(\lambda)$  function, but uses a step function with forces from pedestrians in the front half space being taken fully into account and pedestrians in the rear half space being completely ignored) and extensions for example a routing layer [15] and an alternative way to compute the direction of desired velocity [16]. Relevant here is that it allows to have individual desired speeds for pedestrians and that these can be assigned automatically from numerically defined distributions as well as assigned manually copy-pasting from an external list to the list of pedestrians in the simulation. Desired speeds can be changed during simulation, which is required and realistic to model the gun’s effect.

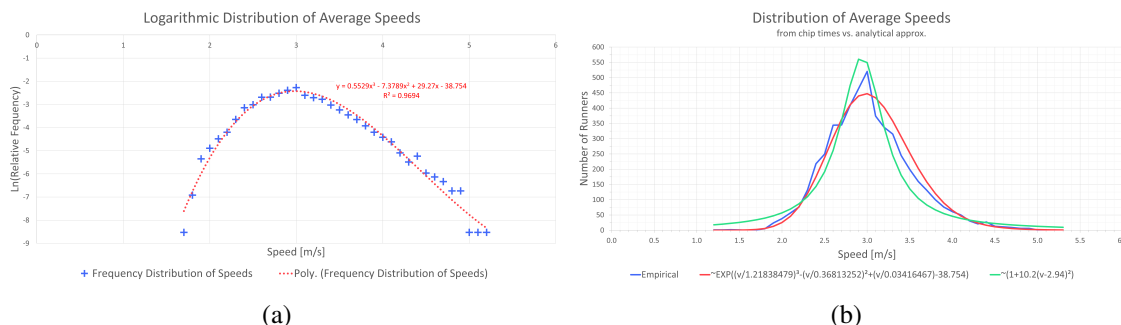
## 3 Analysis of Empirical Data

The first step is to calculate from chip times the average speeds of runners. The difference of gun time and chip time allows to reconstruct the start time of a runner (this time is stored in the timing system database anyway, hence this method in fact is a bit circuitous, however, this is the way doing it using only publicly available data).

---

<sup>1</sup><https://www.google.com/maps/@48.99576,8.3819104,69m/data=!3m1!1e3>

The frequency distribution of average speeds  $f(v)$  can be approximated reasonably with a function of the shape  $f(v) \propto \exp(av^3 + bv^2 + cv)$ , see Fig. 1. Certain deviations from an analytical shape can be expected to remain, as staying below certain finishing times like 4 hours (or 2 hours in half-marathon) define special aims for runners.



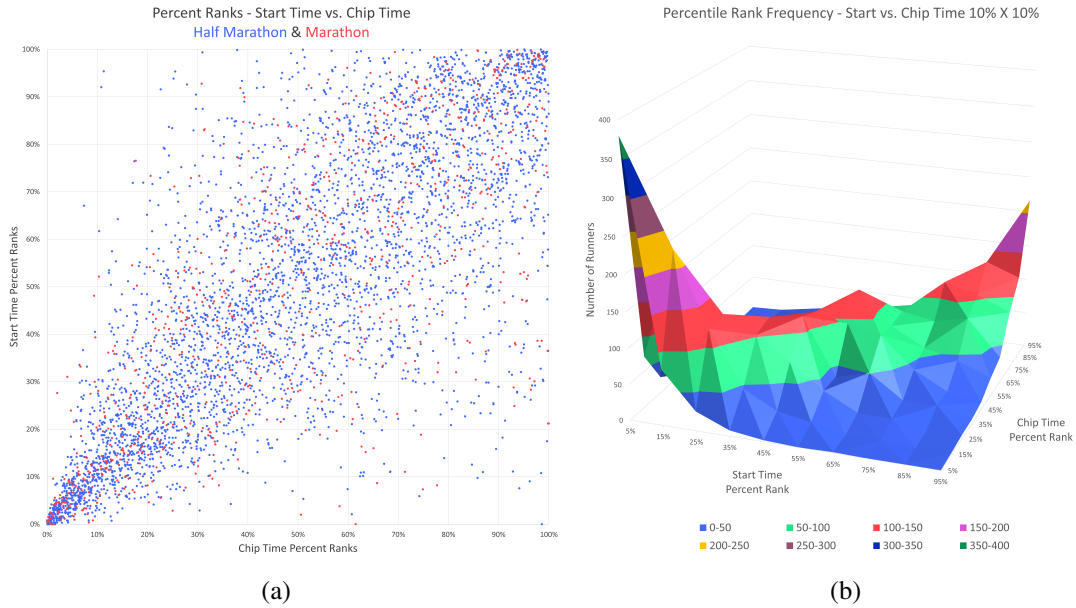
**Figure 1** Frequency distributions for average speeds [m/s] of runners from their chip times. Left: log-scale of relative frequencies (0.1 m/s granularity) with approximation function. Right: absolute frequencies of real distribution and two analytical approximations

The function  $f(v) \propto \exp(av^3 + bv^2 + cv)$  has unfavorable mathematical properties: there is no analytical form of the integral function and cumulative distribution  $F(v)$  and no inverse functions  $v(f)$  and  $v(F)$ . Therefore in freffig:speeds a second analytical function is given (with speeds  $v$  in [m/s], units omitted):  $f(v) = \sqrt{10.2}/(1 + 10.2(v - 2.94)^2)/n$ ,  $F(v) = \arctan(\sqrt{10.2}(v - 2.94))/n + c$ , and  $v(F) = 2.94 + \arctan(n(F - c))/\sqrt{10.2}$  with the constants  $n = 2.8392$  and  $c = 0.4922$  to get  $F(v = 1.15) = 0$  and  $F(v = 5.35) = 1$ . 2.94 m/s is the position of the maximum of the distribution and  $10.2/(m/s)^2$  a measure for its width. With this, it is possible to determine approximately the appropriate starting position for a runner depending on their speed.

With a correlation coefficient of 78% between chip time and start time ranks, runners mostly succeed in aligning according to planned average speed as Fig. 2 shows. However, deviations from a perfect order are not negligible. They are the result of a combination of uncertainty about one’s own running speed and not knowing precisely about the position associated with a particular speed. Note that already upstream of the start line overtaking can occur. This means that the start time percent rank is only an approximation for the spatial distribution of runners before the gun.

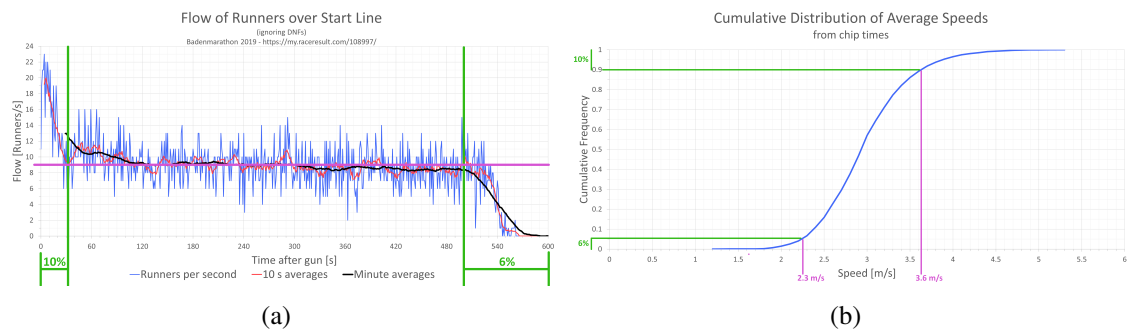
Figure 3 shows the flow of runners over the start line. Interestingly, the flow is about constant from 10th to 94th percentile of the field at around 9/s corresponding to average speeds of 3.6 and 2.3m/s respectively. These limits were chosen as at 10% the 10 seconds average flow for the first time falls below 9/s and at 94% it is for the last time above. A slight decline of the flow can be recognized, but in reality, flow may have been even more constant, as it is plausible that in the later field more runners gave up the race and did not get listed in the results.

The fact that a group of people who all aspire and most of them actually can finish at least a half marathon do not achieve a flow (much) larger than about 1.5 to 1.7/s/m in a race is remarkable. This suggests that it is realistic to assume – as it is done for



**Figure 2** Left: Start time percent ranks vs chip time percent ranks. Right: since the density of dots on the left cannot well be evaluated visually, this 3d histogram is added for assistance. It visualizes that most runners position themselves approximately correct.

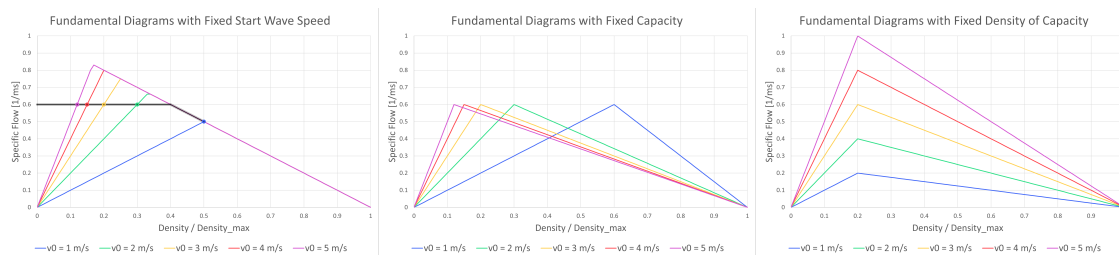
example in [17] – for large groups of people and for an average of the population that a flow larger than  $1.3/s/m$  cannot be maintained for longer times. At the same time, Fig. 3 also shows that for shorter times or particularly fit populations larger flows can occur, compare [18–21].



**Figure 3** Left: Flow of runners over start line per second with running averages for 10 and 60 s. The range where flow is about constant is marked as well as the approximate average value. Right: Cumulative distribution of average speeds with the same percentiles marked as on the left side.

Given that flow is almost constant for average speeds over a relatively wide range from 2.3 to 3.6 m/s one may ask for the reason which could be

- capacity is not reached at the start or it is reached only for a part of the field.
- capacity is independent of (desired) speed.
- a self-synchronization process homogenizes speed and capacity.



**Figure 4** Schematic representation of concepts of FD and their dependence on free speed  $v_0$  if it is assumed that parameters are independent. Left: constant start-wave speed, Center: constant capacity, Right: Constant density of capacity.

- runners have different (desired) speeds at the start than on average of the course.

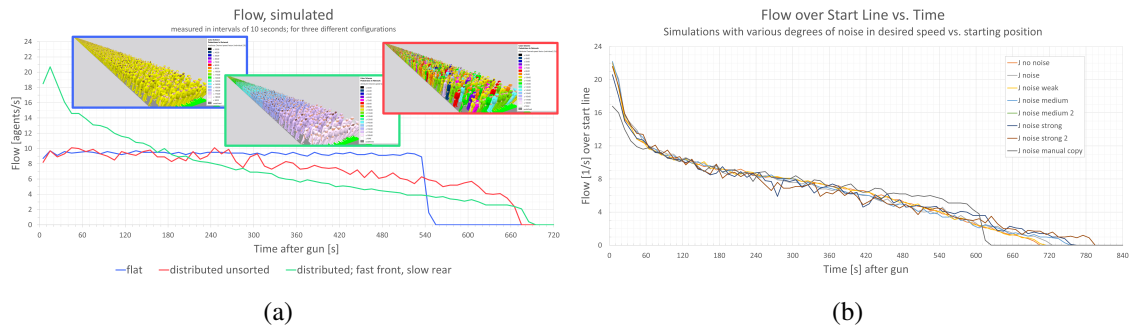
This raises follow-up questions: if runners would not reach capacity at the start of a race who else would? Does it make sense at all to have a notion of capacity? What follows for the fundamental diagram (FD), if capacity was independent of (desired) speed? Concerning the latter question a few thoughts are made in the following. A thorough discussion would go beyond the scope of this contribution. Fig. 4 shows schematically three ways the FD can depend on free speed  $v_0$ . Provided that it is assumed that parameters are independent, in literature one can find examples for all of these (mostly proposed for road traffic, however)<sup>2</sup>. Examples (the referenced property may be implemented only approximately) for constant start-wave speed (left side) are [22–24] and for constant density of capacity (right side) [25–29]. The FD by Daganzo [30] (with parametrization free speed, capacity flow, start wave speed) would belong to the middle as well as to the right category. Similarly, the van Aerde FD [31] is an example for a FD which cannot be categorized clearly. Interestingly, [29] uses the same functional form as [22] but with a re-parametrization through which here it falls in a different category. The FD in Fig. 4 which has the invariant capacity for varying free speeds – the one in the middle – that would fit schematically with the data presented here, has flow-density and speed-density functions for various free speeds which intersect. Usually, the assumption is that larger free speeds lead to equal or larger speeds and flows at any density such that FD functions for various free speeds would not intersect. However, recently from observations at San Fermín festival (running of the bulls) it was hypothesized that the speed-density functions for running are qualitatively different from those of walking, that the former intersect with the latter and that for running functions for all free speeds unify to one function at a density around  $2/m^2$ , see Figure 8 of [32].<sup>3</sup>

<sup>2</sup>It is assumed here that maximum density is not depending on free speed, FDs where this is not the case would have to be discussed separately

<sup>3</sup>Daniel Parisi also presented this in his keynote at PED 2023.

## 4 Comparison with Simulation Results

With the flow over the start line being surprisingly constant, a natural next question is if this would be reproduced by simulations of pedestrians. This can not necessarily be expected as the simulation models usually do not include the start of races among their use cases. If one tries nevertheless, from all parameters obviously at least desired speeds must be adapted from default settings to the higher values of runners. Fig. 5 (left) shows the result for three variants of (localized) speed distribution: 1) all pedestrians having the same desired speed (equal to the average speed of all runners over the whole race), 2) matching approximately (with the “ATAN distribution”) the real distribution of speeds (over the whole race) but distributed randomly over runners in the start field and 3) matching approximately the real distribution of speeds and strictly ordered with fast runners at the front and slow runners at the back end of the field. On average the flow is reproduced fairly well, but none of the three cases captures all properties of the real flow: either the spike at the beginning is not reproduced or the flow is not stable enough over time.

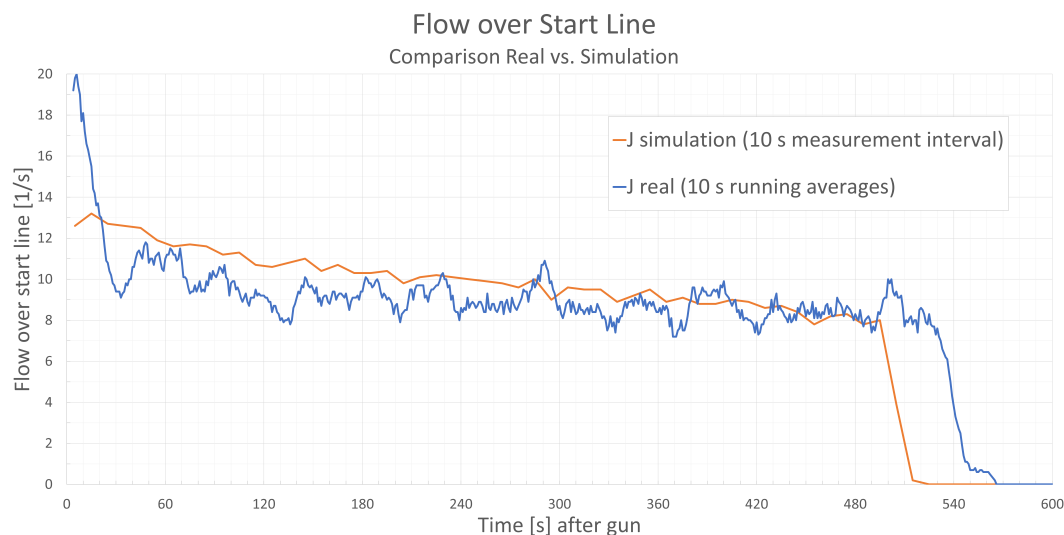


**Figure 5** Left: simulated flow over start line with equal desired speeds (blue), and distributed desired speeds once randomly placed (red) and once strictly ordered (green). Right: Simulated flow with distributed desired speeds basically ordered, but with various degrees of “noise” i.e. slightly misplaced runners with respect to their desired speed.

What might cause this discrepancy? A possible cause may be that in reality runners were neither strictly ordered nor completely unordered. In a first simulation attempt, simply some noise was added, e.g. if a runner is positioned at 65% of the total length of the start field their desired speed was drawn from the “ATAN distribution” near but not exactly at 0.65. By and by with further simulation runs the amount of noise was increased. It can be seen in Fig. 5 (right) that these attempts had near to no effect, except for the simulation run when the “ATAN distribution” was dropped and the exact distribution of the real runners was applied (the rank over time when crossing the start line of real runners was used for the rank of simulated runners in the start field). This resulted in a more but still not realistically stable flow over time.

Therefore, this last method was combined with another change: while so far only desired speeds of the runners were distributed, now also the value of parameter  $\tau$  – the relaxation parameter of the driving force of the Social Force Model [33] – now also was varied for runners in a simple manner: the value of  $\tau$  is coupled to desired speed such that

$v_0/\tau = 4.17 \text{ m/s}^2$ . This led to the result as shown in Fig. 6. The match with the real data is still not perfect. However, this could not be expected as the desired speeds at the start were taken from the average speed over the whole race and second, the fixed coupling of the value of  $\tau$  to desired speed almost surely is – although it brings an improvement – an oversimplification. Similarly, results from a lab experiment [34] recently indicated that it produces more realistic simulation results for bottleneck situations when  $\tau$  is varied with  $v_0$  compared to having an identical value for  $\tau$  for all pedestrians.



**Figure 6** Simulated flow with  $\tau$  set such that  $v_0/\tau = 4.17 \text{ m/s}^2$  for each runner compared to real flow.

## 5 Summary, Discussion, Conclusions, and Outlook

It was found in data from a running event that the flow over the start line is almost constant for about 85% of runners and at 1.5 to 1.7/s/m. Given the wide variety of average running speeds in the race from which indicate different physiological abilities, this result could not be expected right away. Runners' positions before the start correlated well but not perfectly with their whole race average speed. Judging based on simulations the amount of position mismatches would play a rather minor role for the stability of the flow. The simulations produced the best match when the real distribution of whole-race-average speeds was used and combined with identical  $v_0/\tau$  for all runners.

That flow over the start line appears to not depend on desired speed, has odd and difficult to interpret consequences for the FD, if the flow is interpreted as capacity.

For race organization it may be interesting to evaluate more of the many available data sets as simulation results indicate that a high degree of speed-position mismatches would lead to a different time evolution of the flow. Turning the argument around, the time evolution of flow could be an indicator how annoying fast runners might experience having to overtake slower ones during the start phase and that it might pay-off to invest in a better

match of positioning at the start and average running speed (envisioned finishing times). However, it is not clear if this could help beyond reacting on feedback of participants.

A limitation of this study is starters who did not finish (DNF) not being included in the data. With 6006 runners having been registered and 5069 who have completed the race the upper estimation is that 937 contributed to flow but were not measured. However, a rule of thumb for no-show-up (DNS) at ticketed events is 10% such that one would rather have to think of about 300 DNFs. Future works may therefore consider the original raw data where DNFs are included.

With the interesting recent results from the running of the bulls events, the start phase of running events could be seen – in terms of stress of the observed – as being in the middle between a lab experiment and running with bulls, possibly close to evacuation from a building with an unannounced alarm, but no fire products affecting occupants. This can serve as a motivation and justification to also think of extra effort for evaluation, for example the start could be video recorded to allow to measure both speed and density (instead of only their product) and analyze relative spatial positioning of runners, i.e. properties of their Voronoi cells. Analysis of audio recordings could give hints about possible emergent synchronization of steps and effects on desired speeds from that.

**Ethics Statement** This study re-purposes openly published data which was fully anonymized for this study. The author serves as technical product manager for the simulation software PTV Viswalk which was used in the study.

## References

- [1] Tomoeda, A., Yanagisawa, D., Imamura, T., Nishinari, K.: Propagation speed of a starting wave in a queue of pedestrians. *Physical Review E* **86**(3), 036113 (2012). [doi:10.1103/PhysRevE.86.036113](https://doi.org/10.1103/PhysRevE.86.036113)
- [2] Rodriguez, E., Espinosa-Paredes, G., Alvarez-Ramirez, J.: Convection–diffusion effects in marathon race dynamics. *Physica A: Statistical Mechanics and its Applications* **393**, 498–507 (2014). [doi:10.1016/j.physa.2013.09.051](https://doi.org/10.1016/j.physa.2013.09.051)
- [3] Treiber, M., Germ, R., Kesting, A.: From drivers to athletes: modeling and simulating cross-country skiing marathons. In: *Traffic and Granular Flow’13*, pp. 243–249. Springer (2015). [doi:10.1007/978-3-319-10629-8\\_29](https://doi.org/10.1007/978-3-319-10629-8_29)
- [4] Pennisi, A., Bloisi, D.D., Iocchi, L.: Online real-time crowd behavior detection in video sequences. *Computer Vision and Image Understanding* **144**, 166–176 (2016). [doi:10.1016/j.cviu.2015.09.010](https://doi.org/10.1016/j.cviu.2015.09.010)
- [5] Bain, N., Bartolo, D.: Dynamic response and hydrodynamics of polarized crowds. *Science* **363**(6422), 46–49 (2019). [doi:10.1126/science.aat989](https://doi.org/10.1126/science.aat989)
- [6] Anagnostopoulos, A.: The rise of run-commuting as a form of transportation: research on the characteristics and spatial needs of these trips. In: *Advances in*



- Mobility-as-a-Service Systems: Proceedings of 5th Conference on Sustainable Urban Mobility, Virtual CSUM2020, June 17-19, 2020, Greece, pp. 684–693. Springer (2021). doi:10.1007/978-3-030-61075-3\_67
- [7] Anagnostopoulos, A.: Designing runnable cities. *Transportation Engineering* p. 100238 (2024). doi:10.1016/j.treng.2024.100238
- [8] MyRaceResult: Fiducia gad baden-marathon karlsruhe. website (2019). [https://my.raceresult.com/108997/#1\\_95459E](https://my.raceresult.com/108997/#1_95459E)
- [9] Nikitin, P., Rao, K., Lazar, S.: An overview of near field uhf rfid. In: 2007 IEEE international conference on RFID, pp. 167–174. IEEE (2007). doi:10.1109/RFID.2007.346165
- [10] Klohr, N.: RACE RESULT :: Tech Deep Dive :: UHF RFID Timing Fundamentals. youtube (2018). <https://youtu.be/MCQe3kwb76k>
- [11] Klohr, N.: race—result tech deep dive: UHF Detection Algorithm. youtube (2018). <https://youtu.be/AXjWsJM0sTE>
- [12] Result, R.: Hardware configurator. website (2023). <https://www.raceresult.com/en-us/systems/passive.php>
- [13] PTV Planung Transport Verkehr GmbH, Haid-und-Neu-Str. 15, D-76131 Karlsruhe: Manual – PTV Vissim 2023 (2023). Section 4.2
- [14] Johansson, A., Helbing, D., Shukla, P.: Specification of the social force pedestrian model by evolutionary adjustment to video tracking data. *Advances in complex systems* **10**(supp02), 271–288 (2007). doi:10.1142/S0219525907001355
- [15] PTV Group, Karlsruhe: PTV Vissim 2020 Manual (2020)
- [16] Kretz, T., Große, A., Hengst, S., Kautzsch, L., Pohlmann, A., Vortisch, P.: Quickest paths in simulations of pedestrians. *Advances in Complex Systems* **14**(05), 733–759 (2011). doi:10.1142/S0219525911003281
- [17] Maritime Safety Committee (MSC), S.C.o.F.P.F.: Revised guidelines on evacuation analysis for new and existing passenger ships. Circ 1533, International Maritime Organization (IMO), 4 Albert Embarkment, London, UK (2016)
- [18] Schadschneider, A., Klingsch, W., Klupfel, H., Kretz, T., Rogsch, C., Seyfried, A.: *Evacuation Dynamics: Empirical Results, Modeling and Applications*, pp. 3142–3176. Springer New York (2009). doi:10.1007/978-0-387-30440-3\_187
- [19] Zhang, J.: Pedestrian fundamental diagrams: Comparative analysis of experiments in different geometries. Dr., Universität Wuppertal, Jülich (2012). URL <https://juser.fz-juelich.de/record/128157>. Universität Wuppertal, Diss., 2012

- [20] Daamen, W., Hoogendoorn, S.: Emergency door capacity: influence of door width, population composition and stress level. *Fire technology* **48**, 55–71 (2012). doi:10.1007/s10694-010-0202-9
- [21] Kretz, T.: An overview of fundamental diagrams of pedestrian dynamics (2019). doi:10.13140/RG
- [22] Newell, G.: Nonlinear effects in the dynamics of car following. *Operations research* **9**(2), 209–229 (1961). doi:10.1287/opre.9.2.209
- [23] Wahle, J., Neubert, L., Esser, J., Schreckenberg, M.: A cellular automaton traffic flow model for online simulation of traffic. *Parallel Computing* **27**(5), 719–735 (2001). doi:10.1016/S0167-8191(00)00085-5
- [24] Johansson, A.: Constant-net-time headway as a key mechanism behind pedestrian flow dynamics. *Physical review E* **80**(2), 026120 (2009). doi:10.1103/PhysRevE.80.026120
- [25] Greenshields, B., Bibbins, J., Channing, W., Miller, H.: A study of traffic capacity. In: Highway research board proceedings, vol. 14, pp. 448–477. Washington, DC (1935)
- [26] Greenberg, H.: An analysis of traffic flow. *Operations research* **7**(1), 79–85 (1959). doi:10.1287/opre.7.1.79
- [27] Underwood, R.: Speed, volume and density relationships, in quality and theory of traffic flow. *Bur. Highway Traffic, Yale Univ.* pp. 141–188 (1961)
- [28] Drake, J., Schofer, J., May, A.: A statistical analysis of speed-density hypotheses. *Traffic Flow and Transportation* (1965)
- [29] Weidmann, U.: Transporttechnik der fußgänger: transporttechnische eigenschaften des fußgängerverkehrs. *IVT Schriftenreihe* **90** (1993). doi:10.3929/ethz-a-000687810
- [30] Daganzo, C.: The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation research part B: methodological* **28**(4), 269–287 (1994). doi:10.1016/0191-2615(94)90002-7
- [31] Van Aerde, M., Rakha, H.: Multivariate calibration of single regime speed-flow-density relationships [road traffic management]. In: Pacific Rim TransTech Conference. 1995 Vehicle Navigation and Information Systems Conference Proceedings. 6th International VNIS. A Ride into the Future, pp. 334–341. IEEE (1995). doi:10.1109/VNIS.1995.518858

- [32] Parisi, D., Sartorio, A., Colonnello, J., Garcimartín, A., Pugnali, L., Zuriguel, I.: Pedestrian dynamics at the running of the bulls evidence an inaccessible region in the fundamental diagram. *Proceedings of the National Academy of Sciences* **118**(50), e2107827118 (2021). [doi:10.1073/pnas.2107827118](https://doi.org/10.1073/pnas.2107827118)
- [33] Johansson, F., Duives, D., Daamen, W., Hoogendoorn, S.: The many roles of the relaxation time parameter in force based models of pedestrian dynamics. *Transportation Research Procedia* **2**, 300–308 (2014). [doi:10.1016/j.trpro.2014.09.057](https://doi.org/10.1016/j.trpro.2014.09.057)
- [34] Kretz, T., Eckes, L., Diz, J., Knappik, L., Lipp, D., Müller, S.: Using empirical travel time distributions for calibration of a model of pedestrian dynamics. submitted (2023). Presented as poster 51 at PED 2023.