# Generalized Adaptive Smoothing Method for State Estimation of Generic Two-Dimensional Flows

Yuan, Yufei; Hoogendoorn, Serge

**DOI**
10.3141/2561-03

**Publication date**
2016

**Document Version**
Accepted author manuscript

**Published in**
Transportation Research Record

# Generalized Adaptive Smoothing Method for State Estimation of Generic Two-Dimensional Flows

Yufei Yuan, PhD. (Corresponding Author)
Department of Transport & Planning
Faculty of Civil Engineering and Geosciences, Delft University of Technology
Stevinweg 1, PO Box 5048, 2600 GA Delft – The Netherlands
Phone +31 15 278 63 04 Fax +31 15 278 31 79
e-mail y.yuan@tudelft.nl

Prof. Dr. Serge Hoogendoorn
Department of Transport & Planning
Faculty of Civil Engineering and Geosciences, Delft University of Technology
Stevinweg 1, PO Box 5048, 2600 GA Delft – The Netherlands
Phone +31 15 278 54 75 Fax +31 15 278 31 79
e-mail s.p.hoogendoorn@tudelft.nl

Abstract: 235
Main text: 4262
Figures + tables: (4+1)*250 = 1250
Total: 5747

TRR Paper number: 16-2210

Submission Date: 15 Mar 2016

Paper submitted for presentation and publication at the 95th Annual Meeting of the Transportation Research Board, January 10-14, 2016

## ABSTRACT

In big cities, the proportion of slow-mode (such as pedestrian) flows in the total trip demand is steadily growing every year, along with this trend many concerns arise regarding accessibility and safety. The monitoring and management of pedestrians serves as a potential solution to maintain the transport network resilience. Thereof, the monitoring and state estimation for pedestrian flows are crucial as a foundation for a successful crowd management support system. This paper focuses on the development of pedestrian state estimation. A two-dimensional Generalized Adaptive Smoothing Method (2D-GASM) is presented, to estimate the full state of an area based on increasing amount of available pedestrian observations in practice. The 2D-GASM is developed based on similar concepts in the ASM for motorway traffic, which is based on the fact that traffic characteristics travel forward in free flow and backward in congestion. Here, the same mechanism is assumed for pedestrian flows. This extension additionally allows for comprising the two-dimensional (2D) nature of the pedestrian flow and also allows for the fusion and filtering of multi-source data (e.g., counting camera data, WiFi-sensor data, and GPS samples, etc.). Although focussing on pedestrian flow, the approach is applicable to any generic 2D flows, including bicyclist or mixed flows. This newly-developed method is validated based on the trajectory data from a walking experiment at a narrow bottleneck. The testing results present promising estimation performance and possible extensions for future applications are suggested.



**Key words**: pedestrian/bicyclist flows, state estimation, adaptive smoothing method, two dimensional flows, empirical data

# 1 INTRODUCTION AND BACKGROUND

In big cities, the proportion of "slow traffic" (such as pedestrians and bicyclists) in the total trip demand is steadily growing every year. For instance, in the city of Amsterdam, the pedestrian trip increases by 18% whereas the car demand reduces by 18% in the past few years (*1*). Although they are various reasons to welcome this trend, many concerns arise with respect to large crowds at more and more events in cities, crowded shopping areas, and traffic safety problems at places where mixed flows interact. New technical possibilities will allow developing crowd management support systems to ensure the safety of the pedestrians in the crowd. In this way, pedestrian operations can be monitored during operations, and potentially dangerous situations can be identified in a timely manner. With real-time monitoring and state estimation, the aim is to identify the traffic state (consisting of e.g., velocities, flows, densities, route choices) in the network. Based on the results, the crowd management system will give possible measures to alleviate and if possible to prevent the bottlenecks from occurring. The idea is to provide real-time pedestrian monitoring information (flow states, route choice, and potential problematic spots) to stakeholders, and to provide public useful information/instruction to improve pedestrian flow operations.

The focus of this work is on the development of the first task: pedestrian monitoring and state estimation. Limited availability of research on this topic is partially due to the absence of available real-time pedestrian flow data in reality. With the development of monitoring technologies, pedestrian flows can be observed to a certain extent. For example, aerial cameras can be used to track pedestrian movement, infrared sensors to count cross-section flow, WiFi/Bluetooth sensors to detect pedestrian flows equipped with mobile devices at specific locations, also pedestrians equipped with GPS tracker to provide real-time individual trajectories. However, this information can only reflect partial states of the network, other than the fact that errors and bias are part of the raw data. Therefore, state estimation algorithm is needed to reproduce the full picture of states in traffic network, based on the limited and coarse traffic information collected from sensors. Additionally, a combination of types of sensors will support data fusion, to make more use of the individual data sources.

There are many existing state estimation techniques available for car traffic flows. For example, model-based state estimation approaches can be applied, which rely on three components: *a*) dynamic traffic flow models (e.g., first- or second- order traffic flow models in Eulerian and Lagrangian coordinate systems (*2*, *3*)) to predict the evolution of the state variables; *b*) a set of observation equations (e.g., fundamental diagrams) relating sensor observations to the system state; and *c*) a data-assimilation technique (e.g., Kalman filter (*4*) and/or its more advanced variants) is adopted to combine the model predictions with the sensor observations. The estimation quality of model-based methods depends on many factors, such as the representativeness of traffic flow model and observation models, size of discretized units, observation qualities, data assimilation methods, calibrated model input and parameters. This poses a challenge to achieve good estimation performance. Alternatively, the Adaptive Smoothing Method – (one dimensional 1D-ASM) (*5*), extended and generalized in (*6*), can be deployed. It is in essence an approach that interpolates over space and time. The ASM takes explicitly into account car traffic dynamics, namely the propagation of traffic characteristics, and it interpolates the data along two propagation speeds. In contrast to model-based methods, the ASM is data driven, and there is no complex traffic flow model involved which is easier and faster to implement and maintain while providing adequate estimates.

Compared to car traffic, pedestrian flows are two-dimensional and multi-directional. They possess many typical behaviours and phenomena, such as lane formation (in bidirectional flows), diagonal strip formation (in crossing flows), phase transition from laminar to turbulent flow, mildly anisotropic, etc.(*7, 8*). Since the previous techniques for car traffic are not directly transferrable, new state estimation methods are needed to consider the specific characteristics of pedestrian flows. In this paper, we propose a two dimensional generalization of the Adaptive Smoothing Method (2D-GASM), based on predefined available real-time (historical) data sources for pedestrian flows. The extension of the 1D-ASM for pedestrian traffic is twofold: 1. to comprise the two-dimensional (2D) nature of the pedestrian flow; 2. to allow for the fusion and filtering of multi-source data (counting camera data, WiFi-sensor data, and GPS samples, etc.). The new approach is served as a candidate in the development of a crowd monitoring system on behalf of the Amsterdam SAIL nautical event 2015, where visitors walk along a route, while watching and visiting tall ships, eating and drinking (*9*).

This paper is organized as follows. The proposed methodology is firstly presented, with detailed mathematical formulation and the application scope. Next, a validation case study is setup based on the trajectory data stem from a walking experiment. Different types and resolutions of observations are considered. Then, simulation results and discussion are presented, and they are followed by conclusion and recommendations.

## 2 METHODOLOGY

State estimation aims at estimating full (macroscopic) states of pedestrian traffic (densities /number of pedestrians, speeds, and flows) at defined areas (called reservoirs/cells), based on real-time/historical pedestrian data (counts, WiFi-sensor data, GPS), and route choice information (directional field, flow/route share). Note that, although this paper focuses on pedestrian state estimation, the related analysis and formulation can be easily generalised to other 2D flows, including bicyclist and mixed flows.

As indicated, the 2D-ASM that will be presented in this paper is developed based on concepts similar to the 1D-ASM for motorway traffic. The ASM was originally designed for processing single data sources and reconstructing spatiotemporal traffic plots, which is based on spatiotemporal characteristics of car traffic flow, that is, perturbations travel forward in free flow (with the free flow speed $v_0$) and backward in congestion (with the jam propagation speed $\omega$). The ASM estimates the traffic state variable by determining a weighted average of the "neighbouring" data in space and time. The weights for the averaging are determined by a so-called kernel, where the weights depend on the direction and distance of the included data points relative to the center of the kernel (where closer means higher weight). Two kernels are used, one for free flow conditions and one in congestion. Typically Gaussian or negative exponential functions are used to define the shape of the kernel, and by setting the parameters for three dimensions (one for temporal dimension, two for spatial dimension), the size of the kernels can be adjusted, as will be described in detail below.

**Generalised Adaptive Smoothing Method in three dimensions**

The method assumes knowledge about the following pedestrian flow quantities:

- The pedestrian free speed $v_0$ (with a magnitude of 1.5 m/s)
- The pedestrian shockwave speed $\omega$ (opposite of the walking direction)
- The expected walking direction $\vec{\gamma}(t,\vec{x})$ for each time instant and location; specifically, $\vec{\gamma}(t,\vec{x})$ denotes directional vector in a 2D spatial plane.

Although the shape of pedestrian fundamental diagram depends on many factors, such as the size/shape/function of an area, the heterogeneity/composition of pedestrian flows, trip purposes (*10*), for demonstration purposes, we assume a triangular fundamental relation, similar to that in car traffic with two constant wave speeds. Note that, the assumption with a pedestrian shockwave speed is proposed to accommodate the algorithm formulation and its applications. In reality, we might observe shockwaves in big festivals and events (e.g., the aforementioned SAIL event) or queuing experiments (e.g., a Sugiyama-type circuit experiment (*11*)) with dense crowds. To check the plausibility of this assumption and to observe a clear shockwave speed is still an interesting topic to investigate for future.

The general idea entails applying the ASM in two separate directions (vertical and horizontal directions). The parameters of the ASM are adapted given the expected walking directions. Note that, the generalisation to 2D flows is not trivial because of local route choice processes that cause "lateral" flows influencing the longitudinal flows. Hence, knowledge needs to be added about the walking directions / local route choice of the elements in the pedestrian flow.

We assume that we have data available, that consist of the quadruples $(t_i,\vec{x}_i,\vec{v}_i,\vec{z}_i)$. Here, $i$ denotes the index of the data point, $(t_i,\vec{x}_i)$ the time and the location it pertains to, $\vec{v}_i$ the velocity and $\vec{z}_i$ the actual measurement that we aim to reconstruct. Note that in many cases, $\vec{z}_i=\vec{v}_i$ (e.g., the measurement we aim to reconstruct is actually the velocity). However, $\vec{z}_i$ can also refer to other observable quantities that we want to reconstruct, such as flow $\vec{q}_i$ and/or density. In this case, the corresponding speed measurement pertain to the targeted reconstructed quantity needs to know. Note that, both the targeted quantity (e.g., flow) and its related speed measurement should follow the walking direction (align with the directional line).

The aim of the approach is to estimate the value $\vec{z}=\vec{z}(t,\vec{x})$ for any value of $(t,\vec{x})$ using the available measurements. There are mainly four steps: a. construction of directional fields; b. projecting data on the directional line; c. determine the free flow and congestion weights of each observation data point; d. determining regime and overall weights for the free flow and congestion filters.

**Step 1**: constructing the directional field. Given the infrastructure, destinations, measurements, etc., we aim to reconstruct the directional field of the pedestrians that we are estimating the traffic variables for. Multiple approaches can be used here, such as modelling (*12*), estimating from the (historical) data, etc. The idea is that the outcome of the estimation approach will be relatively robust against errors determined when determining the directional field. In the end, we assume that $\vec{\gamma}(t,\vec{x})=(\gamma_1,\gamma_2)$ can be determined for all relevant values of $(t,\vec{x})$, where $\gamma_1$ and $\gamma_2$ are the horizontal and vertical components of the direction vector in the Euclidean space (${\gamma_1}^2 + {\gamma_2}^2 = 1$).

**Step 2**: projecting the observation data $\vec{z}_i$ on the directional line. Note that in the current version, the same type of observations (speed or flow) from different data sources can be non-discriminatively put into the observation set (implicitly of the same reliability level). The pedestrian flow will move approximately along the parameterized line:

$$\vec{x}(s) = \vec{x} + \vec{\gamma}(t, \vec{x}) \cdot s$$

Here, $s$ denotes the updating time step, and $\vec{x} = (x_1, x_2)$, where $x_1$ and $x_2$ are the horizontal and vertical components of vector $\vec{x}$.

We can easily show that we can rewrite this line as follows:

$$ax_1 + bx_2 + c = 0$$

with $a = \gamma_2$, $b = -\gamma_1$ and $c = \gamma_1 x_2 - \gamma_2 x_1$. Using this expression, we can compute the distances $\delta_i$ to this line (orthogonal projection) for all data points as follows:

$$\delta_i = \frac{\left|ax_1^{(i)} + bx_2^{(i)} + c\right|}{\sqrt{a^2 + b^2}}$$

Next to the orthogonal distance, we can compute the distance $d_i$ from the projection on the line to the point $(t, \vec{x})$. This distance is equal to:

$$d_i = \sqrt{D_i^2 - \delta_i^2}$$

Here, $D_i$ is the distance between the observation point $\vec{z}_i$ and position $\vec{x}$. FIGURE 1 provides a graphical description of Step 2.

Since we need to include the direction as well, we define:

$$\lambda_i(t, \vec{x}) = sign(\vec{\gamma}(t, \vec{x}) \cdot (\vec{x}_i - \vec{x}))$$

**Step 3:** determining the free flow and congestion weights of each data point. For each data point *i*, we will determine a weight expressing the contributions to the estimate at location $(t, \vec{x})$ - under the assumption that traffic flow is either free (information moves along $\vec{\gamma}(t, \vec{x})$ with the free speed) or congested (information moves in the opposite direction $-\vec{\gamma}(t, \vec{x})$ with the shockwave speed). We first define the weighing function (negative exponential function) that expresses the impact of time $s$ and (generalized) distances $d$ and $\delta$:

$$\phi_0(s, d, \delta) = \exp\left(-\frac{|s|}{\tau} - \frac{|d|}{\sigma} - \frac{|\delta|}{\eta}\right)$$

where $\tau$, $\sigma$ and $\eta$ are parameters, determining the kernel weight of each data point. $\tau$ relates to the temporal dimension, and the other two count for the spatial dimension. Note that, a Gaussian type function can also be used to determine the shape of the kernel:

$$\phi_0(s, d, \delta) = \exp\left(-\frac{s^2}{2\tau^2} - \frac{d^2}{2\sigma^2} - \frac{\delta^2}{2\eta^2}\right)$$

The contribution of the data point $i$ in case of free flow conditions is now computed as follows:

$$\phi_i^{free} = \phi_0(s - \lambda_i \cdot d_i / v_0, d_i, \delta_i)$$

where $v_0 > 0$ is the free speed, while the contribution of the data point in case of congested conditions would be equal to:

$$\phi_i^{cong} = \phi_0(s - \lambda_i \cdot d_i/\omega, d_i, \delta_i)$$

where $\omega < 0$ is the shockwave speed.

For the variable estimate to be determined, we now determine the weighted averages:

$$\vec{z}^{free}(t,\vec{x}) = \frac{\sum_i \phi_i^{free} \cdot \vec{z}_i}{\sum_i \phi_i^{free}}$$

and

$$\vec{z}^{cong}(t,\vec{x}) = \frac{\sum_i \phi_i^{cong} \cdot \vec{z}_i}{\sum_i \phi_i^{cong}}$$

**Step 4:** determining regime and overall weight. In the final step, we determine the overall weight by considering which regime is present. This is based on the estimate of the speed at the location we are considering. We have:

$$w = \frac{1}{2}\left(1 + \arctan\frac{V(t,\vec{x}) - V_c}{\Delta V}\right)$$

Here, $V_c$ is the critical speed and $\Delta V$ denotes a smoothing parameter. $V(t,\vec{x}) = \min(\vec{v}^{cong}(t,\vec{x}), \vec{v}^{free}(t,\vec{x}))$, this indicates a “congestion-win” principle, which means the speed regime is always bounded by the lowest speed.

The resulting value of the estimate of variable $z_i$ then equals:

$$\vec{z}(t,\vec{x}) = (1-w)\vec{z}^{cong}(t,\vec{x}) + w\vec{z}^{free}(t,\vec{x})$$

This algorithm falls in the category of parametric data-driven methods. Indeed, for optimal accuracy, each site should probably have a (slightly) different set of model parameters. These parameters, including kernel weight parameters ($\tau$, $\sigma$ and $\eta$), parameters $V_c$ and $\Delta V$ in the weighting function, and pedestrian wave speeds would need to calibrate site-specifically, although they should be of the same magnitude. Perhaps ground truth data within very short periods would be adequate for calibration, via for example from cameras at that site.

**Application of the GASM for different types of pedestrian data**

Pedestrian flows can be observed and monitored by various surveillance techniques. Aerial cameras can be deployed to track pedestrian movement and to obtain trajectories via image processing, but sensor coverage is normally limited. This method is not in prevailing usage since it is costly and different to maintain and interpret. Local counting cameras and infrared sensors can be used to count cross-section flows (and speeds). This type of sensors has been used to monitor pedestrians at selected important spots. Similarly, WiFi/Bluetooth sensors can be placed at fixed points to detect (partial) pedestrian flows equipped with mobile devices, but in an economic and loose way. Moreover, pedestrians equipped with GPS trackers (GSM signals, dedicated GPS devices) can provide real-time individual trajectories, although the penetration rate is modest. As mentioned, errors and bias are part of the raw data obtained from different techniques in reality. In the remainder, we assume that the same reliability level applies to various data sources. Although this may not be generally the case, since the data with a higher quality should be weighted more in the model; for the sake of this method presented, the error that is made due to this assumption does not influence the main message of this contribution. To assign different predefined reliability weights for various data sources will be considered in the extension of the current work.

In this paper, two types of observations are assumed to be available to the model, namely flow and speed information from local count cameras at a cross-section level – in the form of data quadruples as $(t_i^{Cam}, \vec{x}_i^{Cam}, \vec{v}_i^{Cam}, \vec{q}_i^{Cam})$, and individual pedestrian trajectories from GPS trackers – in the form of $(t_i^{GPS}, \vec{x}_i^{GPS}, \vec{v}_i^{GPS}, \vec{v}_i^{GPS})$. The two data sources can be exclusively or inclusively input into the framework, of an equal reliability level and a refined data quality (no noise). The GASM can be used in a straightforward manner to fuse different data sources (refer to step 2). Based on these observations at discretized points and the proposed GASM, we aim at estimating the complete flow and speed states for the whole surveyed area. Note that, the directional field (route choice) is assumed to be known.

## 3 APPLICATION EXAMPLE AND MODEL VALIDATION

For validation purposes, the considered situation is a narrow bottleneck scenario. The data stem from a large scale walking experiment which was used to analyze the pedestrian walking behaviour in case of a narrow corridor of 1 m width (*13*). FIGURE 2 shows the full set of pedestrian trajectories and a snapshot of the walking experiment. Thereof, all the trajectory information of individual pedestrians is available, and it can be used to provide ground truth data and to emulate any type of observation data (e.g., GPS data and local camera data). The surveyed area is a 10m x 4 m rectangular area. The total experiment time period is 920s.

The main purpose of this validation study is to apply the GASM using limited measurement to reproduce traffic states in the whole area. For demonstration purposes, the surveyed area is subdivided into small equidistant x-y grids in size of 0.25m x 0.25m. The updating interval is set as 10s. Note that, ground truth data can be derived from the trajectory information with respect to this spatiotemporal setting. The parameter settings in the GASM: the free flow speed $v_0 = 1.5\text{m}/s$; the congested wave speed $\omega = -0.25\text{m}/s$; the critical speed $V_c = 0.7m/s$, the smoothing parameter $\Delta V = 0.5$, and the kernel parameters $\tau = 10s$, $\sigma = 0.5m$ and $\eta = 0.1m$. This choice is based on the experience accumulated in literatures, and the calibration of the parameter set on a site-by-site basis remains as future work. The directional field of the pedestrians was directly estimated from the trajectory data.

Several scenarios have been performed to test the validity of the method with diverse detection resolutions. Two data sources are considered: GPS data providing instantaneous speeds at reporting instants with six varying penetration rates, local counting camera providing aggregated flows and speeds at fixed positions (cross-section data relating to a small longitudinal grid) with three varying spatial (longitudinal) resolutions. In total, there are 17 data scenarios in three categories, two scenarios with sole counting camera data, five with pure GPS data, and the rest ten cases regarding data fusion of both sources, see

**TABLE 1.**

The quality of the filter result is measured by the Root Mean Squared Error (RMSE) and the Mean Absolute Percentage Error (MAPE), regarding the speed ($V$) and flow ($Q$) estimations at equidistant spatiotemporal grids (0.25m x 0.25m x 10s). Both error indicators can provide absolute and relative performance of estimation scenarios.

# 4 RESULTS AND ANALYSES

FIGURE 3 presents the performance of all the data scenarios listed in

**TABLE 1.** Clearly, the estimation accuracy increases with detection resolutions (high penetration rate and dense camera spacing).

The scenario with only 5% exclusive GPS data (where solely speed estimate is available) shows a considerable improvement in terms of both the RMSE and the MAPE, compared to the case of 1~3% GPS data (moving from 1 to 5% GPS data improves the RMSE for $V_x$ about 30%, and a clear inflection point can be observed at the performance curve in FIGURE 3(a)). This indicates that small amount of GPS data would entitle a satisfactory estimation performance, although the exact value of the GPS data penetration for this performance is still subject to verify with more empirical tests. The main contribution of the GPS data is to capture the movement trend for the whole population. Instead, cross-section cameras measure the whole population locally at fixed locations (similar to loop detections for car traffic), and the estimation based on this local information provides acceptable results (refer to the error values intersected with the vertical axes in FIGURE 3(a~c)). With limited speed samples, the 2D-GASM is able to reconstruct the two-dimensional speed information to a satisfactory level. If the local cross-section camera data with additional *flow* and speed information together with the GPS data are put into the filtering framework, speed estimation can provide better results. More importantly, flow estimation becomes available for the entire area.

However, the increasing of GPS data input has very limited influence on the flow estimation, since GPS data only provide one-off (isolated) position-speed measures which may not be representative of the flow component, see FIGURE 3(d). Meanwhile, it is noticed that flow estimation result is not as good as the speed estimation (flow MAPE error is in general larger than 40%). Because the underlying model cannot capture the dynamics of flow evolution (e.g., the flow conservation principle), it is suggested to deploy a model-based estimation approach for improvement (e.g., a Kalman type of state estimation based on a macroscopic pedestrian model (*14*)).

It is interesting to see that the percentage error (MAPE) for speed in lateral (y) direction is considerably high (larger than 100% thus not included here), even when its absolute sense (RMSE) is in a small magnitude (see FIGURE 3(b)). This can be partially explained by the fact that there exist both positive and negative values along the lateral direction and both values are relatively small, and thus sign error would lead to substantial MAPE results.

FIGURE 4 gives the graphic presentations of both the ground-truth data and the estimation at a certain time instant. From the graphic results, the estimation with pure GPS data (or exclusive cross-section data) already presents very promising results for speed estimation, see FIGURE 4(b) for the GPS case. In FIGURE 4(c), the estimation performance is enhanced with data fusion of additional local camera data, and details at network boundaries are able to reproduce (refer to red dashed circles). The flow estimation (see FIGURE 4(d)) is actually a reasonable representation of the ground truth. Compared to the ground-truth data, the 2D-GASM is able to provide a full coverage of states at the entire surveyed area, with relatively low error indicators. Note that, the estimation values at some network grids (the right upper and lower parts – denoted by rectangular boxes in the figure) that do not exist in reality, can be neglected from the results. Equivalently, spatial-operating domain can be easily included to consider the existence of obstacles when applying the GSAM method.

If we apply a naive method, for example, the speed estimate is calculated by a local arithmetic average of the surrounding GPS samples at a network grid, instead of spatiotemporal neighbouring average. The result is quite limited, only obtaining speed estimates at a few grids of the entire area. Flow estimation is not even possible.

## 5 CONCLUSION AND RECOMMENDATIONS

This paper pioneers to investigate the possibility of pedestrian state estimation for crowd management systems, in the context of the increasing amount of available pedestrian observations in practice, and the increased need for such crowd management systems. It presents a novel two-dimensional generalized Adaptive Smoothing Method, incorporating pedestrian flow nature and data fusion concept. The methodology can be generalized to any generic 2D flows, including bicyclist or mixed flows. The model validation study based on trajectory data from a walking experiment has demonstrated that the 2D-GASM is adequate to reconstruct pedestrian speed and flow field of the network, although flow estimation can be further improved. The data fusion concept embedded in the filtering framework can improve estimation quality successfully incorporating both GPS data and cross-section data.

In this paper, the 2D-GASM shows good performance at a simple pedestrian motion base case: unidirectional flow. Future work is needed to test the performance of the GASM at more complex situations: crossing flows and/or bidirectional flows which require applying the GASM separately for each of the flow groups (leftward, rightward, upward and downward), and for other generic 2D flows (e.g., bicyclists). Meanwhile, the sensitivity analysis in terms of different model parameters (free-flow/congested wave speeds, kernel parameters) and observation data qualities remains as future work. In the current application, all the data sources are combined under the assumption of the equal reliability level. Further extension is needed to allow separate data input with predefined reliability weight for each of the data sources.

## ACKNOWLEDGEMENT

This research was supported by the European Research Council advanced grant for unrAveLLing slow modE travellinG and tRaffic with innOvative data to a new transportation and traffic theory for pedestrians and bicycles (ALLEGRO).


## REFERENCES


1 Amsterdam-manicipality. *Jaarverslag 2011 (2011 annual report)*. Gemeente Amsterdam - Dienst Infrastructuur Verkeer en Vervoer (Amsterdam manicipality - Department of infrstructure, traffic and transport), 2011.

2 Wang, Y., and M. Papageorgiou. Real-time freeway traffic state estimation based on extended Kalman filter: A general approach. *Transportation Research Part B: Methodological,* Vol. 39, No. 2, 2005, pp. 141-167.

3 Yuan, Y., J. W. C. Van Lint, R. E. Wilson, F. Van Wageningen-Kessels, and S. P. Hoogendoorn. Real-Time Lagrangian Traffic State Estimator for Freeways. *IEEE Transactions on Intelligent Transportation Systems,* Vol. 13, No. 1, 2012, pp. 59-70.

4 Kalman, R. E. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering,* Vol. 82, No. 1, 1960, pp. 35-45.

5 Treiber, M., and D. Helbing. Reconstructing the spatio-temporal traffic dynamics from stationary detector data. *Cooper@tive Tr@nsport@tion Dyn@mics,* Vol. 1, 2002, pp. 3.1-3.24.

6 Van Lint, J. W. C., and S. P. Hoogendoorn. A robust and efficient method for fusing heterogeneous data from traffic sensors on freeways. *Computer-Aided Civil and Infrastructure Engineering,* Vol. 25, No. 8, 2009, pp. 596-612.

7 Hoogendoorn, S. P., and W. Daamen. Self-organization in walker experiments. In *Proceedings of the 5th Symposium on Traffic and Granular Flow*, 2004.

8 Duives, D. C., W. Daamen, and S. P. Hoogendoorn. State-of-the-art crowd motion simulation models. *Transportation Research Part C: Emerging Technologies,* Vol. 37, 2013, pp. 193-209.

9 SAIL Amsterdam 2015. https://www.sail.nl/en/. Accessed Jul. 2015.

10 Helbing, D., A. Johansson, and H. Z. Al-Abideen. Dynamics of crowd disasters: An empirical study. *Physical Review E,* Vol. 75, No. 4, 2007, pp. 046109.

11 Sugiyama, Y., M. Fukui, M. Kikuchi, K. Hasebe, A. Nakayama, K. Nishinari, S.-i. Tadaki, and S. Yukawa. Traffic jams without bottlenecks - experimental evidence for the physical mechanism of the formation of a jam. *New Journal of Physics,* Vol. 10, No. 3, 2008, pp. 033001.

12 Hoogendoorn, S. P., and P. H. L. Bovy. Pedestrian route-choice and activity scheduling theory and models. *Transportation Research Part B: Methodological,* Vol. 38, No. 2, 2004, pp. 169-190.

13 Hoogendoorn, S. P., and W. Daamen. Pedestrian Behavior at Bottlenecks. *Transportation Science,* Vol. 39, No. 2, 2005, pp. 147-159.

14 Hoogendoorn, S. P., F. L. M. van Wageningen-Kessels, W. Daamen, and D. C. Duives. Continuum modelling of pedestrian flows: From microscopic principles to self-organised macroscopic phenomena. *Physica A: Statistical Mechanics and its*

*Applications,* Vol. 416, 2014, pp. 684-694.

## LIST OF TABLES



## LIST OF FIGURES

**TABLE 1 Experimental scenarios with GPS data of six testing levels and local camera data of three testing levels**

| Source | Input/Output | Resolution (Testing level) (penetration rate or spacing) | No. Scenarios |
|---|---|---|---|
| GPS | Speed (instantaneous) | 0%, 1%, 3%, 5%, 7% 9% | 6 x 3 - 1 = 17 |
| Camera | speed and flow (cross-section aggregated) | Void, 4m (1,5,9), 2m (1,3,5,7,9) | |

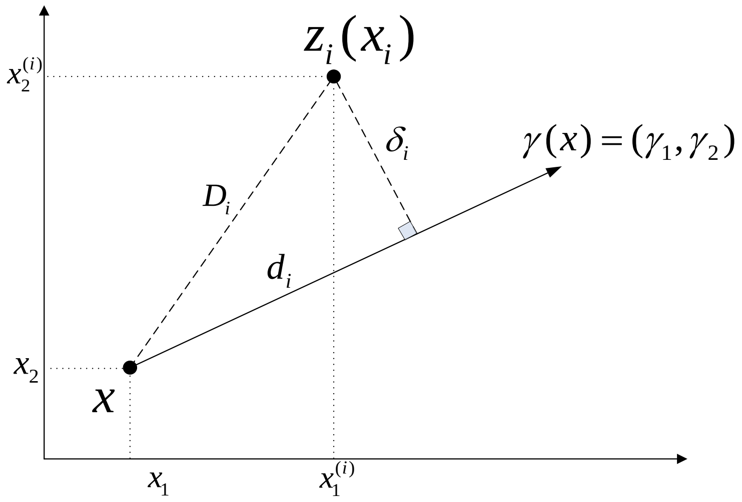



**FIGURE 1 Projecting observation data to the directional line.**

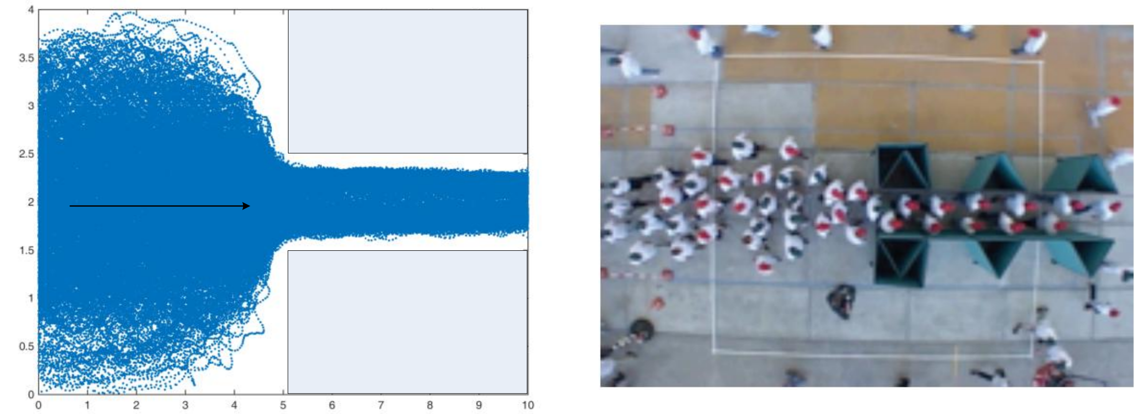
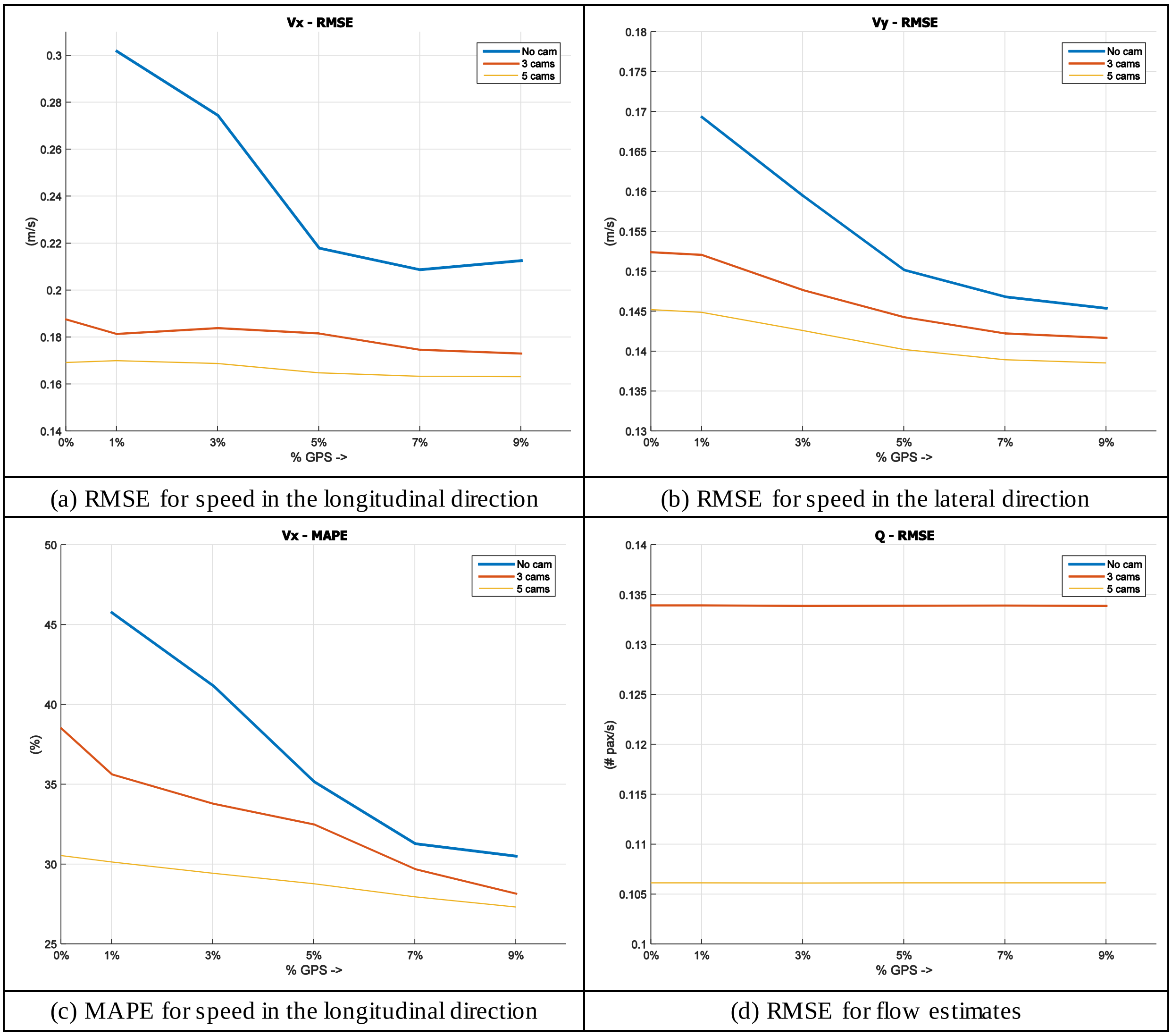



**FIGURE 2 Trajectories and snapshot from a walking experiment at a narrow bottleneck.**



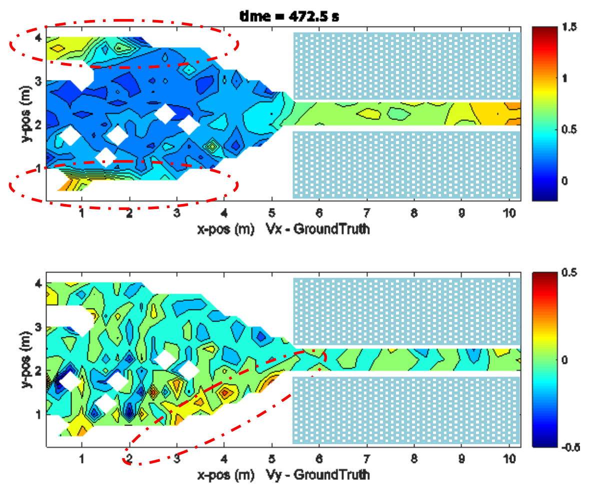

(a) RMSE for speed in the longitudinal direction

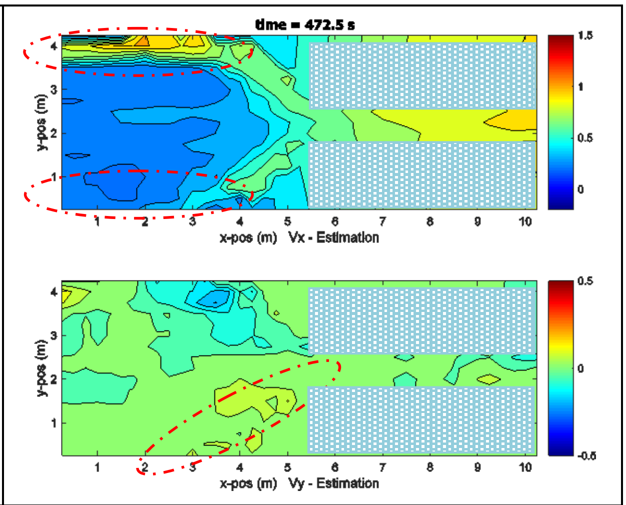
(b) RMSE for speed in the lateral direction

(c) MAPE for speed in the longitudinal direction

(d) RMSE for flow estimates

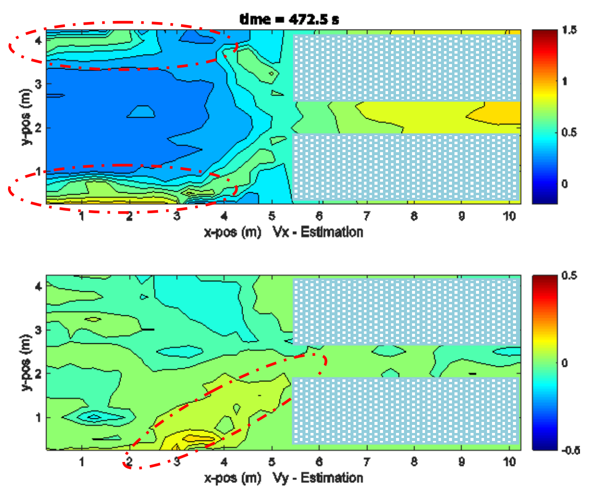
**FIGURE 3 Error performance of all the data scenarios regarding RMSE and MAPE.**

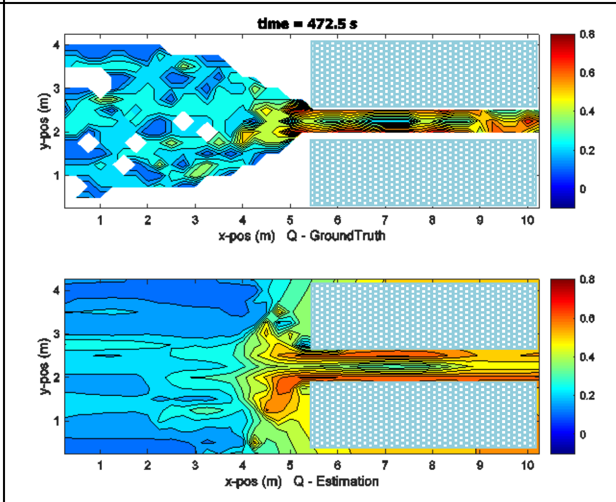



(a) Ground truth - speed

(b) Estimation with GPS (5%) - speed




(c) Estimation with data fusion (5%-GPS, 2m-camera) - speed

(d) Flow estimation comparison with cross-section cameras (2m)

**FIGURE 4 The ground-truth data with the reconstructed velocities (in x and y direction) and flows at one (random) selected time instant (472.5s).**