22nd ITS World Congress, Bordeaux, France, 5–9 October 2015

Paper number ITS-2545

The NDW data fusion project: pilot description and results

Edoardo Felici[1*], Isabel Wilmink[2] , Diana Vonk Noordegraaf

1. National Data Warehouse for Traffic Information, the Netherlands, Griffioenlaan 2, 3526LA Utrecht, +31 631764363, edoardo.felici@ndw.nu

2. DITCM Innovations, the Netherlands

Abstract

Probe Vehicle Data (PVD) is considered a promising data source for traffic information. However, the quality and added value of this PVD and data fusion of PVD with other traffic data (e.g. loop detectors) is not evident. In the Netherlands, road authorities and service providers obtain traffic information procured by the National Data Warehouse for Traffic Information (NDW), a public-public partnership of 24 road authorities. In preparation of future procurement of traffic information, NDW, together with DITCM Innovations, has organized a pre-commercial data fusion pilot in 2014 to find out whether PVD or fused data including PVD can be deployed in the near future. Three teams, comprising 15 companies, participated in the pilot. NDW intended to use the pilot to investigate whether making more extensive use of PVD and data fusion is a good approach to enable a reduction of the number of fixed location sensors (mainly loop detectors), and to assess the added value of PVD and data fusion for NDW's end users. This paper describes the set-up of the pilot and the main conclusions and recommendations.

Keywords: DATA FUSION, PROBE VEHICLE DATA, TRAFFIC INFORMATION, FLOATING CAR DATA

Introduction: the National Data Warehouse for Traffic Information

In the Netherlands, road authorities and service providers obtain traffic information and traffic data procured by the National Data Warehouse for Traffic Information (NDW). The initial procurement of traffic data by NDW in 2009 provided speed measurements, travel times and vehicle counts with vehicle categorization measured on approximately 24.000 locations throughout the country on the main road network and a number of provincial roads. The network of measurements has expanded over the years to include higher coverage on provincial roads and main urban roads. Currently, traffic data is collected through a number of technologies, mainly induction loops, but also passive infrared (PIR) detectors, automatic number plate recognition (ANPR) cameras and Bluetooth sensors are used. The application of the different technologies depends on the required data type, data quality and possible requirements limiting the use of equipment embedded in the road surface, thus excluding induction loops from being used.

Probe Vehicle Data (PVD)[1] is considered an promising data source for traffic information. NDW expects that PVD will play an ever-growing role in the procurement process of NDW in the near future. NDW wishes to use more PVD due to the relative ease of data collection, not requiring roadside equipment implementation and maintenance, and the expected cost effectiveness of this solution. The vast installed base of roadside sensors offers NDW the opportunity to explore data fusion possibilities with PVD, with the expectation that this will create synergies in quality and optimization between the various methodologies of data collection. However, the quality and added value of this PVD and data fusion of PVD with other traffic data (e.g. loop detectors) is not evident.

Together with DITCM Innovations, NDW has organized a pre-commercial pilot in 2014 to prepare for the implementation of PVD and data fusion. Three teams, comprising 15 companies, participated in the pilot (see for more information on the approach and the teams Appendix 1). NDW intended to use the pilot to investigate whether making more extensive use of PVD and data fusion is a good approach to enable a reduction of the number of fixed location sensors (mainly loop detectors), and to assess the added value of PVD and data fusion for NDW’s end users. This paper describes the set-up of the pilot and the main conclusions and recommendations.

Goal of the data fusion pilot

The goal of the pilot was to gauge which level of quality of road traffic data (spot speed, vehicle counts and travel time) is feasible when using PVD, data from roadside equipment and data fusion, assuming that the spacing between these fixed location sensors is increased. Ideally, NDW would like to obtain data of the same quality as is delivered by the current data collection systems. After completion of the pilot, NDW intended to incorporate the insights from the pilot in the strategy for the future procurement of data traffic data.

---

[1] In this paper we use the term Prove Vehicle Data to indicate data about moving vehicles. As not only data from passenger cars but also from trucks can be used, this term is preferred over the term Floating Car Data.

A sub-goal for the pilot was to encourage participants to research this in an in-kind and open-innovation setting. This meant that teams did not receive any financial compensation and would ideally not be competing with each other – as there was no prize to be won – but would share acquired knowledge and results with all participants. This is why NDW decided to run the pilot in collaboration with the DITCM Innovations.

NDW

The National Data Warehouse for Traffic Information (NDW) in the Netherlands is a public-public partnership that procures traffic data from various sources, processes these into one database, and delivers these data to road authorities and service providers. A number of governments are partners in NDW: the national road authority Rijkswaterstaat, the twelve provinces of the Netherlands, the cities Amsterdam, Utrecht, Rotterdam and The Hague and a number of regional cooperations. NDW started in 2009 with the goal to provide a one-stop-shop for national traffic data procurement and distribution and act as a shared service organization on behalf of the partners. (www.ndw.nu)

DITCM Innovations

DITCM Innovations is an open innovation organization in which government, industry and knowledge institutes work together on the successful introduction of cooperative systems to sustainably support mobility and accessibility. DITCM Innovations has almost 30 partners who jointly operate a development and test environment for new forms of intelligent vehicles and the associated intelligent roadside systems. DITCM offered operational support and coordination during the duration of the pilot, as well as a platform to publicize the pilot and its results. (www.ditcm.eu)

Scope of the data fusion pilot, research questions and research steps

Scope

Given the short 4-month period in which the pilot had to be carried out, and the in-kind contribution of the teams, the scope of the pilot was limited and a small set of scenarios covering relevant situations was defined. The following points of departure were formulated:

- Motorways as well as the underlying road network were considered.
- The focus was on speed measurement and vehicle counts, as data fusion was thought to have the most added value for those data when considering a reduction of the

number of fixed location sensors. More is known already about the use of PVD to generate travel time data.

A comparison was to be made between the fused data and the loop detector data. The loop detector data were considered as the reference in this case, even though NDW was aware none of the utilized technologies provides a ground truth.

Research questions

The pilot aimed to answer the following main research question:
Which level of quality of road traffic data (spot speed, vehicle counts and travel time) can be achieved by using PVD, data from fixed location sensors and data fusion, assuming that the spacing between fixed location sensors is increased?

Underlying research questions were:

- What is the added value of PVD and/or data fusion?
- Can the investigated solutions be scaled up (and widely deployed) easily?
- Which other lessons learned can be concluded from the pilot?

Research steps

Six steps were determined for the successful completion of the pilot:

1. Determine indicators on a per-minute basis
   Spot speed and vehicle counts had to be determined using PVD for those locations that were removed in the different scenarios (see Appendix 2 for more details). Examples of the different scenarios are: removal of one location, removing of all locations between intersections except one, removal of 50% of the locations on either motorway or underlying road, etc. For each scenario the teams had to demonstrate how well they were able to produce the traffic data in the situations where data from fixed measurements (i.e. mainly loop detectors) was not available.

2. Assess results on a per-minute basis using NDW quality criteria
   The quality criteria that had to be assessed were based on: number of minutes for which data (either travel time, vehicle count of spot speed) could be generated in an hour, derived inaccuracy per period of two hours compared to the original fixed measurement data, derived unreliability for every period that data is compared for between data fusion and fixed measurement data and the deviation in percentage between the averages for the measurement period.

3. Analyze the results in relation to traffic engineering variables
   The results had to be analyzed further to determine if the deviation from the original measurement data was small or large looking at various traffic engineering variables like the traffic volume, presence of convergence or divergence points, load, moment of congestion creation, shockwaves and incidents. By looking at these criteria, a more qualitative analysis was made of the pilot results.

4. Determine the key factors to achieve high quality data
   Teams were asked to determine which aspects of the PVD (accuracy of positioning and time intervals between measurements, number of observations per minute/hour during different times of the day and influence of differences in vehicle categorization and locations) where key to achieving the results. Also, the data fusion algorithms had to be evaluated on the way they functioned, their self-learning capabilities and their response in different traffic conditions (congestion, traffic lights, at intersections and during incidents).

5. Assess the added value of PVD and data fusion
   Teams were asked to openly assess the added value of their solutions in relation to different road types, also outside of the research area, and the possibility to intensify measurements or generate different types of data.

6. Determine the scalability of the solution
   NDW has an ambition to implement a data fusion solution on a national scale. The teams were asked to determine if their solutions could be scaled geographically to include larger areas, and if it could be used in a real-time environment.

Team solutions

Team ARS Traffic and Transport Technology and Here

Team ARS used aggregated Here PVD in a statistical Singular Value Decomposition. With this method, temporal-spatial variations in vehicle counts and spot speeds are divided into components. The first components show the largest and most important structural variations. Other components show lesser variations. The method depends on the historical data that was provided by NDW to make a calculation of the most important components, where all the data provided for the month of March was used, except for the day a prognosis was made for. This method proved useful to predict spot speed patterns and can be used to generate traffic information. Adjustments to the filtering method will make it possible to detect and follow shockwaves through traffic. ARS recommended NDW to define which applications would make use of data fusion – statistical goals, traffic management and traffic information all cater to different needs and require different approaches.

Team Be-Mobile, Goudappel Coffeng, MAPtm and VORtech;

Team Be-Mobile made use of the Treiber-Helbing filter (adaptive smoothing method) and the Kalman filter combined with a traffic model (Omnitrans). The PVD originated from various sources that have been contracted by Be-Mobile. Different PVD-aggregations were used, with a penetration of between 2% and 4%. The Treiber-Helbing filter combined data from the PVD and nearby fixed measurement locations, where the weight of the two sources could be varied within the equation. The Kalman filter used together with Omnitrans could estimate the missing values for spot speeds and vehicle counts around existing measurement locations. Their conclusions were that 'raw' PVD provided significantly better results compared to aggregated PVD. The quality of PVD, in terms of coverage but also the way the GPS-positions are translated to spot speeds, is a determining factor in the quality of data fusion. The amount of fixed measurement locations can be significantly reduced, without loss of quality. Integrating PVD ensures a continuous overview of traffic conditions that would not be available based solely on fixed measurement location data.

Team CGI, SAS, Grontmij, Accenture, HP, TU Delft and TomTom

Team CGI used an adaptive smoothing method in an extended generalized Treiber-Helbing filter. This ensures that data of different granularity in space and time can be combined. The individual deviation in quality and reliability of each data source can be taken into account. The method constructs detailed location- and time-dependant vehicle count and spot speed patterns, based on the measurements of various sources and taking into account the propagation of patterns during different traffic states: downstream during free flow, upstream during congestion. The quality of PVD is a determining factor for the quality of the data fusion solution. Team CGI also recommends to determine which purpose should be served by NDW with data fusion, as to determine the most suitable quality criteria for the specific purpose.

Conclusions and recommendations

The insights that resulted from carrying out the data fusion pilot have been summarized in a set of conclusions and recommendations, as given below.

Market readiness of PVD and data fusion

One of the questions with which NDW started the data fusion pilot was whether the industry was ready to deliver PVD and data fusion to NDW. This question was comprised of two parts:

- Can PVD and data fusion be delivered with the desired level of quality?
- What do the participating teams consider feasible in the short term?

From the comparison with loop detector data it could not be concluded that replacing loop data with other data (such as PVD/fused data) is feasible while maintaining the level of quality as required by NDW – if the loop data are used as a reference case, the accuracy and reliability of the data obtained with PVD and data fusion are too low in many of the scenarios that were evaluated. It has to be noted, however, that the comparison with loop data does not give a complete picture, because there is noise in the loop detector data as well (i.e. it is not the “ground truth”). The participating teams are confident that PVD and data fusion can be deployed soon. It seems therefore useful to adapt the quality criteria currently used by NDW, in order to (a) enable quality checks of PVD/fused data and (b) have more diversified quality criteria that better match the goal for which the data are eventually used (policy monitoring, traffic information, evaluation of measures etc.). This can have consequences for the temporal resolution of the data storage (not necessarily per minute anymore).
PVD and data fusion appear to have added value in specific situations (such as incidents and shockwaves), and can probably also deliver good information for policy monitoring purposes, but it is not a fully developed product yet. To actually implement it, a period of further development of about six months is likely needed to meet all wishes of NDW’s users of traffic information.

Whether the fused data are good enough, cannot be concluded from the results of the pilot alone. The decision about this depends on what exactly will be tendered (in terms of data supply contracts). The pilot provides little insight into whether data fusion (with PVD) can be deployed for real-time traffic information applications. There need to be requirements for data latency for these types of applications.

Quality of traffic data generated with PVD and data fusion

When the data generated with PVD and data fusion are compared to loop detector data, the NDW quality criteria can be met for spot speed data outside peak hours. Traffic patterns can generally be reproduced well, as visualizations show. It is recommended to judge the quality of vehicle counts and speed data generated with PVD and data fusion in a different way than with the current NDW quality criteria, by for instance looking at distributions of speeds and traffic volume (and to compare those to distributions derived from loop detector data). A quality indicator expressed in just one number does not do justice to data generated with PVD and data fusion.

Factors influencing data quality

From visualizations of fused data it has become clear that the granularity of PVD is important in specific situations (such as congestion caused by incidents, and shockwaves) – a higher resolution in space and time gives better insight into what exactly is happening in such situations. An important issue for future tendering of PVD is how to determine the quality of the data. The pilot showed that this entails more than just penetration rate – both the source of the data and certainly also how the data are processed are important.

Added value of PVD and data fusion

Visualizations of fused data clearly show the added value of (disaggregated) PVD and data fusion, where it concerns for example incident congestion and shockwaves. The visualizations give a more complete picture of traffic flow conditions (there is also information for the sections between the loop detectors). This type of visualizations gives more insight into the quality and plausibility of the data. There also seem to be opportunities for the underlying road network (urban and rural roads), for which there are much less loop detector data. For judging the potential of data fusion (also in the context of future tenders), it is important not to just check the data quality at cross sections. Data for road sections or routes should be considered too – see also the remark above about generating a more complete image of traffic (this brings with it that another way of storing the data is needed.) Data fusion also has potential as a tool for continuous plausibility checks, because a coherent picture is formed from the various data sources. This offers, on the one hand, the possibility to give a higher quality score to estimates based on data sources all pointing in the same direction; on the other hand there is the possibility to identify a data source that systematically deviates from the ‘complete’ (fused) picture.

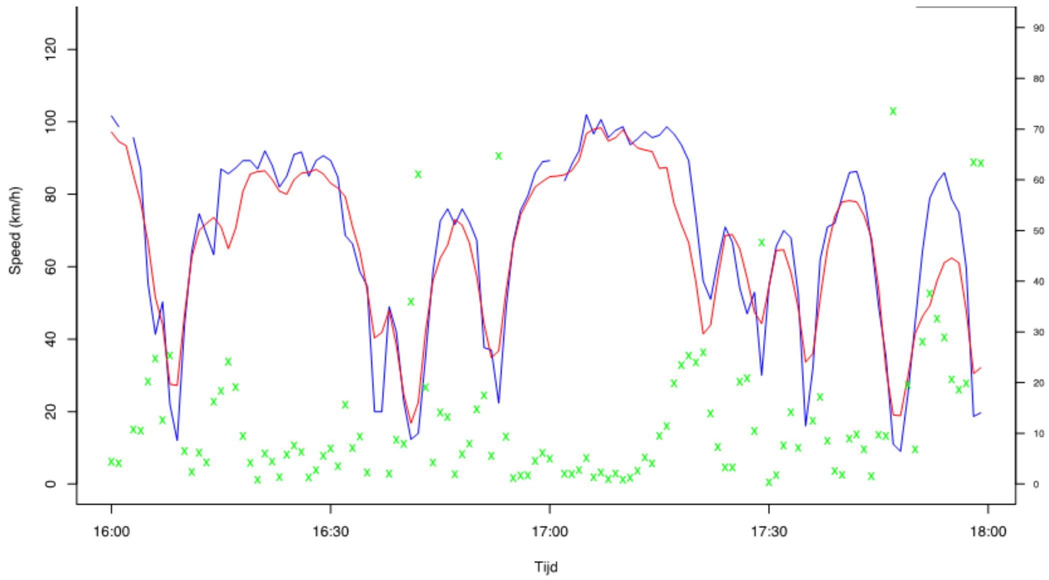

Figure 5 – Comparison between fixed measurement data (blue) and fused data (red) by team Be-Mobile.

Figure 5 shows how patterns in traffic are quite successfully reproduced using fused data. This can lead to a number of cost-effective applications in traffic management centres, even though the data does not adhere to the current NDW quality standards – these have been formulated to comply with strict statistical purposes as well.

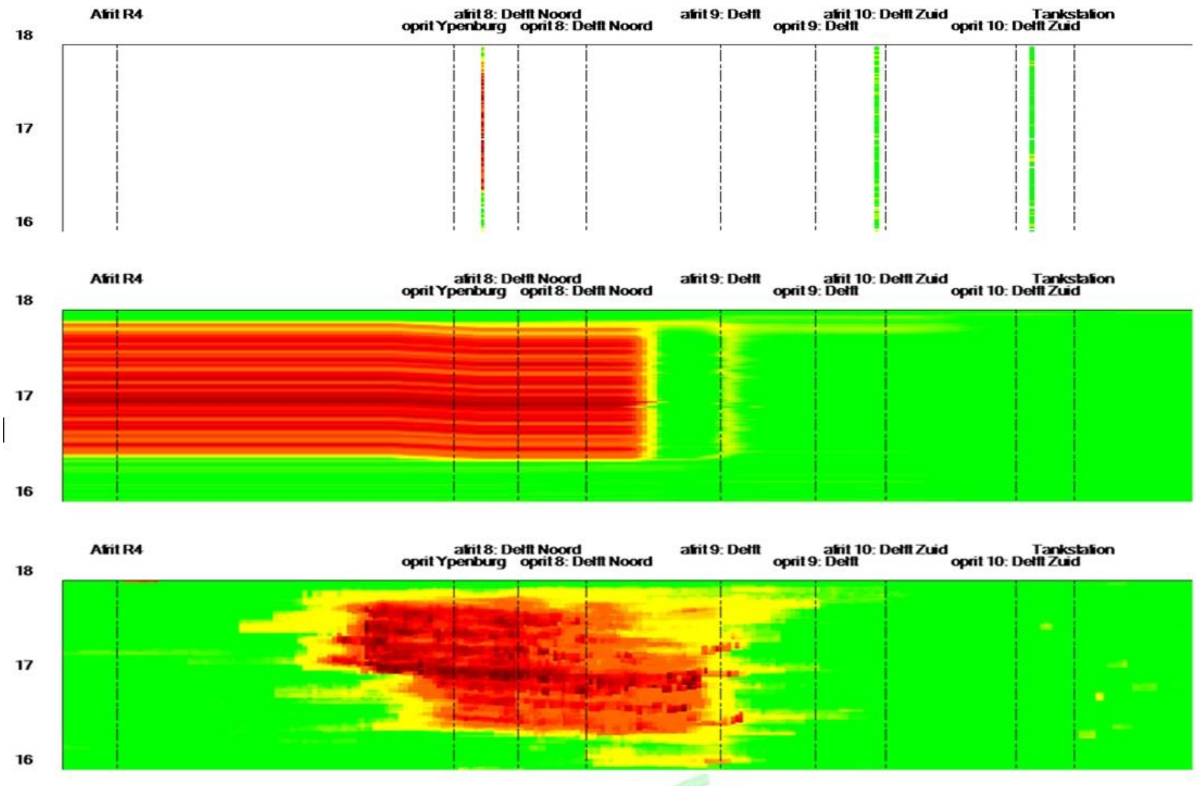

Figure 6 – Congestion data from induction loops (top), data smoothed out between fixed measurement locations (middle), fused data (bottom), by team Be-Mobile.

The three diagrams in figure 6 show the potential of fused data. Congestion can be spotted in a much more accurate manner using fused data than by exclusively using induction loop data.

Scalability

The methods applied appear to be suitable for large scale deployment. Note that this may not be the case when traffic models are used, as this makes the methods more complex. It is expected, however, that with the current capabilities of computers this complexity can be handled.

Process

All parties involved have gained many insights, both content-wise and process-wise. In addition, all involved parties in the pilot worked well together. The interactions with and between the teams were pleasant and there was a large amount of goodwill. Participants indicated that working together in a precompetitive setting is something they would like to do again. Also, the teams see opportunities for further collaborations in the future.

Acknowledgements

NDW would like to thank all participating teams for their effort and availability to participate in this in-kind pilot, and DITCM Innovations for their expertise and professional guidance throughout the pilot.

References

1. Be-Mobile (2014), Traffic Flow QBench.
2. Connekt (2013), Beter geïnformeerd op weg. Routekaart 2013-2023 – Hoofddocument. Delft, Connekt, In opdracht van het Ministerie van Infrastructuur en Milieu, oktober 2013.
3. DITCM (2014), About DITCM Innovations, Frequently Asked Questions http://www.ditcm.eu/images/Contact/FAQ_DITCM_Website.pdf.
4. NDW (2014), Organisatie, http://www.ndw.nu/pagina/nl/3/organisatie.

Appendix 1 Pilot scenarios

During the pilot a number of scenarios were researched where the number of fixed sensor locations was reduced in different quantities and positions. PVD and data fusion were then used to extrapolate the missing data on the locations that had been removed. The following illustrations clarify the setup:

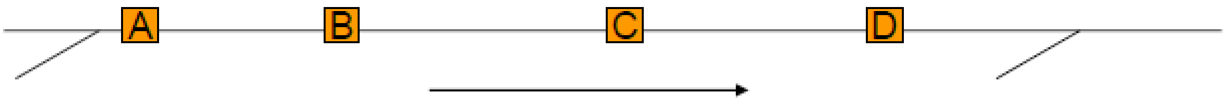

Figure 1 – Basic measurement setup

In figure 1, four measurement locations named A through D are placed on a road section between two intersections or on-/off-ramps.

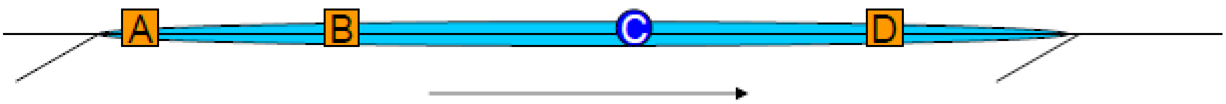

Figure 2 – Removal of a measurement location

In figure 2, measurement location C has been removed from the dataset. Teams must now extrapolate the data that was measured in location C using PVD and data fusion.

Various scenarios were used for the pilot: rush hour and non-rush hour periods on both the motorways and the underlying road network.

Research area

The location that was chosen for the pilot was the road network comprising the A4 and A13 motorways and the N470 (Kruithuisweg) around Delft. The location was chosen on the basis of the availability of data, the combination of different road types and the occurrence of congestion.

Figure 3 – Locations of fixed measurements on the pilot road network

Figure 4 – Research area (red line), additional data availability (green line)

Data from the fixed measurement locations was supplied to the teams covering two time periods: from 01/03/2013 to 31/03/2013 and from 19/05/2013 to 05/06/2013.

Appendix 2 Approach and teams

Approach of the datafusion pilot

During the pilot a collaborative approach was used with respect to both content and process. The main components of this approach were:

- A project kick-off open to all interested parties at Intertraffic.
- A more content-oriented kick-off with the teams that had registered to participate in the pilot.
- Telephone interviews and site visits conducted by impartial NDW/DITCM experts to discuss the teams' results.
- A meeting with the teams to discuss the final results.
- Data analyses and reporting by the teams and NDW/DITCM.
- An assessment of the process with the teams.
- A symposium during which the results of the pilots were presented to a wide audience. This covered the results made public; the teams decided which of the results they wanted to share.

Participating teams in the datafusion pilot

Three teams successfully completed the pilot:

- Team ARS: ARS Traffic and Transport Technology, Here;
- Team Be-Mobile: Be-Mobile, Goudappel Coffeng, MAPtm and VORtech;
- Team CGI: CGI, SAS, Grontmij, Accenture, HP, TU Delft and TomTom.