A METHODOLOGY TO DISAGGREGATE SPEED DATA COLLECTED BY ROAD DETECTORS

Catani L., Tremblay J.-M., Bassani M., Cirillo C.

POLITECNICO DI TORINO
Department of Environment, Land and Infrastructure Engineering

UNIVERSITY OF MARYLAND
Department of Civil and Environmental Engineering

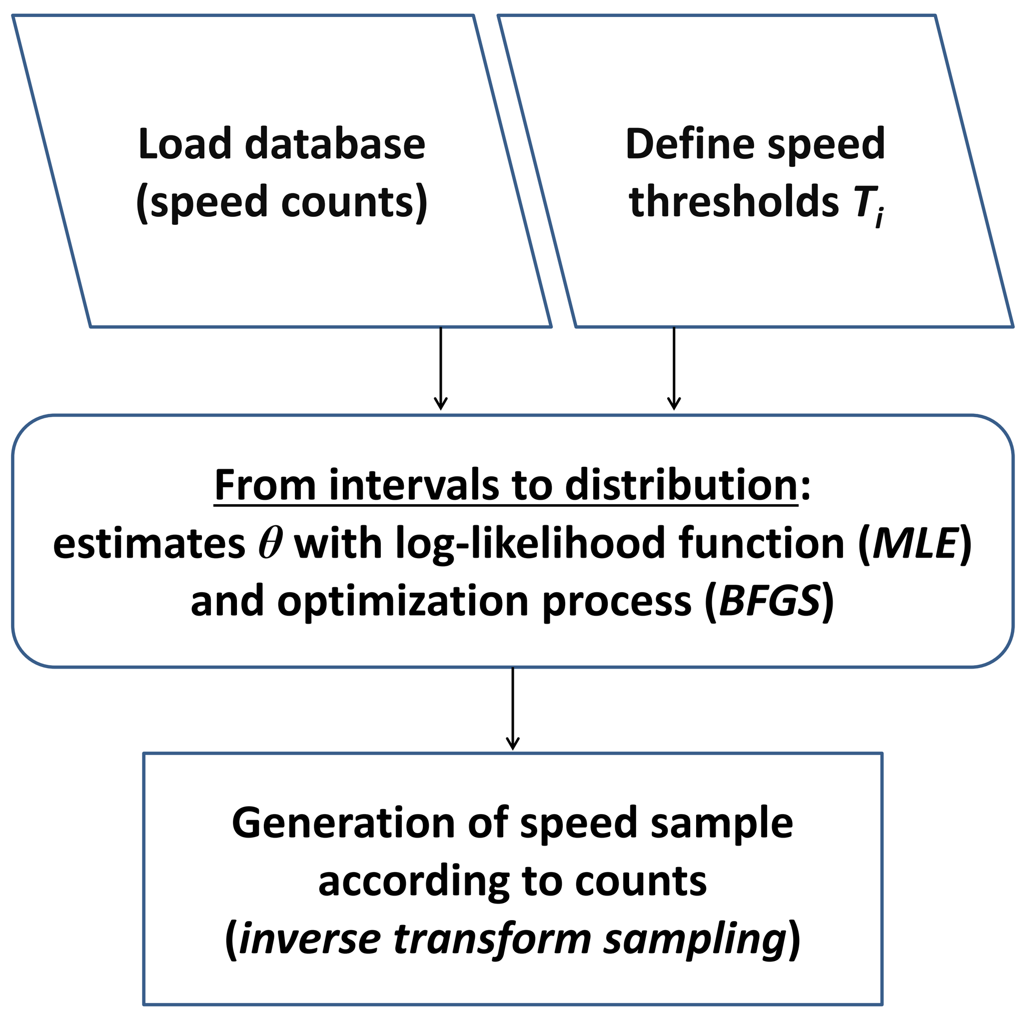
BACKGROUND, MOTIVATIONS AND OBJECTIVES

- **Speed** is the fundamental parameter in **road design**, **traffic operation and control**, and **road safety**.
- **Speed data acquisition** are performed by **several methods** (magnetic, pneumatic, laser, video, etc.) that may collect data in **aggregate** or **disaggregate** form; each method has different **functionalities**, **costs**, and **accuracies**.
- Aggregate data **conceal information** about the real trend of discrete values, that are more powerful for operating purposes.

Algorithm that computes disaggregate values from aggregate speed samples by considering **different statistical distributions**.
Such **individual speeds** are necessary to operate with **continuous distribution functions** and to derivate **basic descriptive measures**.

METHODOLOGY

The algorithm has been provided for three statistical distributions (Normal, Lognormal, and Gamma) usually employed for speed modeling. It was developed by using *R* software v3.1.1.



Definition of counts. Assuming that V_j are observed speeds that follow a particular distribution f, characterized by a parameter θ

$$V_1, V_2, \dots V_j, \dots, V_n \ iid \ f(\theta) \qquad \text{for } j = 1, 2, \dots, n$$

Defining the speed thresholds T_i, the counts of vehicles c_i are defined as:

$$c_i = \sum_{j=1}^{n} I\left\{V_j \in [T_i, T_{i+1}]\right\} \qquad \text{for } i = 1, 2, \dots, k$$

$$I(V_j) := \begin{cases} 1 \ if \ V_j \in [T_i, T_{i+1}] \\ 0 \ if \ V_j \notin [T_i, T_{i+1}] \end{cases}$$

Maximum Likelihood Estimation method. The Log-Likelihood (*LL*) function and the estimated parameter $\hat{\theta}_{MLE}$ are respectively:

$$LL(\theta) = \sum_{i=1}^{k} c_i \cdot \log[F(T_{i+1}; \theta) - F(T_i; \theta)]$$

$$\hat{\theta}_{MLE} = \arg\max_{\theta} \sum_{i=1}^{k} c_i \cdot \log[F(T_{i+1}; \theta) - F(T_i; \theta)]$$

Inverse Transform Sampling technique to compute individual speeds according to the distribution parameter and the sample size.

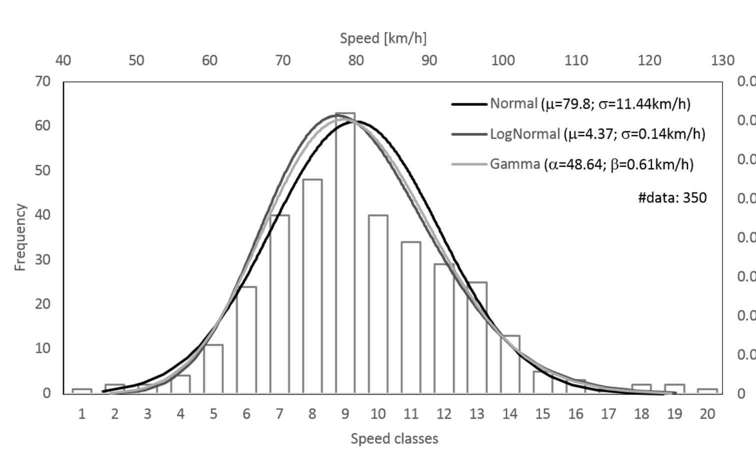
$$V = F^{-1}(u) \qquad \text{where } u = \left(\frac{1}{c_i+1}, \dots, \frac{i}{c_i+1}, \dots, \frac{c_i}{c_i+1}\right)$$

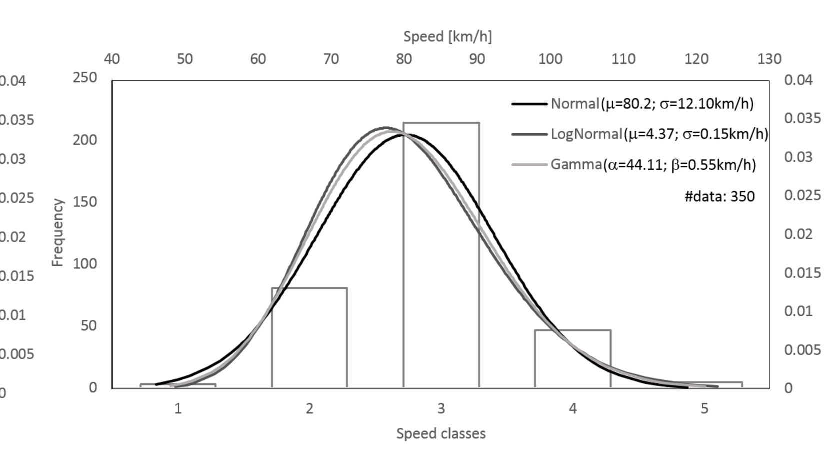
Example. Speed sample of 350 data.

Parameters estimated by Salter and Hounsell (1) methodology

Agg.classes \ Parameters	mean, $\bar{v}$	std.dev, s
20	79.9 km/h	11.5 km/h
5	80.3 km/h	13.4 km/h

Parameters estimated through the proposed algorithm





(1) Salter, J., and N. B. Hounsell. *Highway Traffic Analysis and Design*, 3rd ed. Palgrave Macmillan, Houndmills, 1996. ISBN 0333609034.

ACKNOWLEDGEMENTS

The work included in this paper has been supported by the Compagnia di San Paolo (Italy) and the Politecnico di Torino under the grant "*Bando per il Finanziamento di Progetti di Internazionalizzazione della Ricerca*", approved with the Rectoral Decree n. 208 of the 24th of May, 2013.

The authors wish to acknowledge the Department of Civil and Environmental Engineering at University of Maryland, College Park (US) for hosting Lorenzo Catani as a visiting student. This international cooperation agreement and the support from the Politecnico di Torino and the National Transportation Center at the University of Maryland has made this research possible. Speed data was collected by Alessio Bertola who is gratefully acknowledged for his contribution.

CASE STUDY

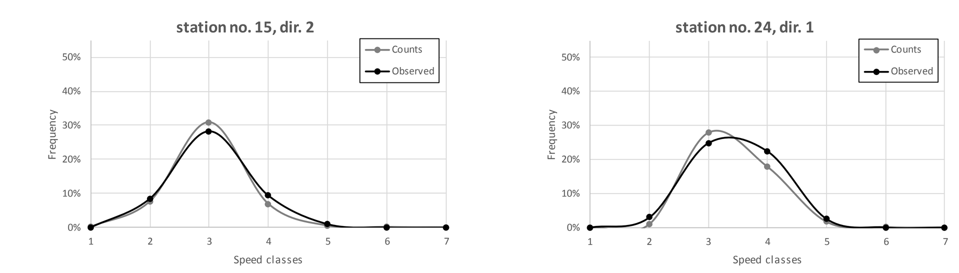
Speed surveys were performed in six stations on rural roads in the Province of Turin (North-West of Italy). Data we collected at the same time both with loop detectors (counts) and cross-registration (individual), during off-peak hours to avoid traffic congestion.

Road name	Station ID	Inductive loop stations (1 hour)		Video speed survey (15 minutes)			
		Time	sample size	Time	sample size	V_{min}	V_{max}
SP6	1	12 - 1 pm	791	12.27 – 12.42 pm	202	33.5	116.5
			733		208	29.7	131.2
SP177	12	12 - 1 pm	205	12.37 – 12. 52 pm	48	30.9	92.7
			330		81	27.6	93.3
SP176	14	11 am - 12 pm	199	11.44 – 11. 59 am	38	47.4	91.4
			150		51	42.6	106.6
SP267	15	9 - 10 am	386	9.45 – 10 am	103	16.5	95.7
			365		95	34.7	105.4
SP220	16	10 - 11 am	248	10.37 – 10.52 am	64	44.5	89.1
			231		51	43.0	99.4
SP183	24	1 - 2 pm	384	1.29 – 1.44 pm	106	43.8	107.5
			363		90	36.9	134.8
		Total	4385	Total	1137		

Double-loop detectors: No. of classes: 7
thresholds: 30, 50, 70, 90, 110, and 130 km/h
aggregation period: 1 hour

Cross-registrations: recording period: 15 minutes

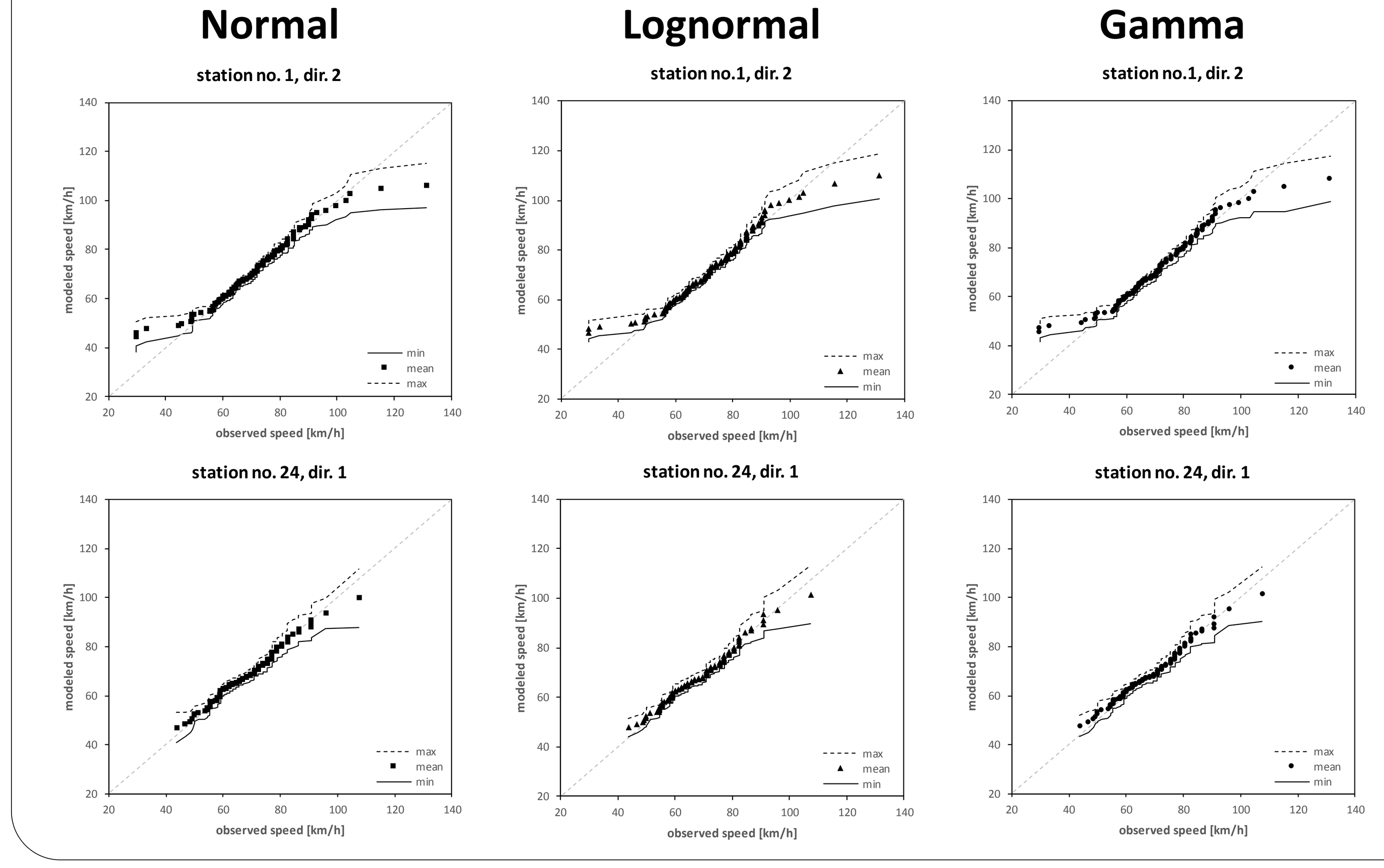
…but comparable observation frequencies inside each speed interval (difference lower than 10%)



Collected data were also analyzed to estimate the sample estimates ($\bar{v}$, s or α, β) and to define if they are prone to follow a specific distribution by using the Kolmogorov-Smirnov test.

- estimated parameters are consistent (max difference equal to 12%) ➡ **STATIONARITY** of flow
- no evidence to suggest that one function distribution (Normal, Lognormal or Gamma) is superior to the others

Algorithm application. Distribution means are similar to those computed in the preliminary analysis; standard deviations are influenced by data aggregation (only seven classes) and by the sample composition process (inverse transform sampling).
Due to different observation periods, to compare observed and modeled speeds, reduced samples were created through a random extraction from the outcome vector, with the actual number of values inside each interval.
The algorithm accuracy has been evaluated by evaluating the range in variability of the modeled speeds. For each statistical distribution, the algorithm was iterated a hundred times to define the limits of the estimates (solid line).



➡ The range in variability is small in the central part of the distribution, while it tends to increase towards the tails of the functions, leading to a greater variability in modeled speeds

➡ Although the maximum and minimum values can deviate from the equality line, the data close to the average are very close to it. As expected, the algorithm works reasonably well and sometimes very well for the speed located across the central speed classes.

Some measures of accuracy (MPB, MAD, SDE, MPE) were used to provide information on the goodness-of-fit between observed and modeled data.
These indexes confirm the results of statistical analysis performed on observed samples, or rather there is no one distribution that fits better with respect to the others.

CONCLUSIONS

- the modeled distribution parameters are similar to those of original data; the mean is generally not affected by aggregation, while some discrepancies are noticed in the standard deviation;
- the choice of the statistical distribution which best interprets the field observations may be made on the basis of accuracy measures such as those considered in this manuscript;
- the proposed algorithm facilitates the construction of databases of individual data that are essential for conducting a variety of investigations which deal with speed analysis.

Download the poster here!

Session 288 – Statistical Methods In Transportation
Convention Center, Hall E – B131
Monday, January 9, 2017 (10:15 am - 12:00 pm)

96th Annual Meeting - Transportation Research Board
January 8-12, 2017 ▪ Washington, D.C.

Paper code: 17-01738

mail to: ccirillo@umd.edu