

# M2 Internship position at Inria Sophia Antipolis Méditerranée

## Big Data Analysis and Modeling of Road Traffic

January 22, 2018

### Scientific Context

There exists nowadays a variety of road traffic data sources (magnetic loop detectors, video cameras, floating car data, bluetooth, etc), which can be used for traffic management either for calibration of traditional models [1], or to directly infer traffic forecasts by statistical techniques [2], or even by combining modeling and data analysis techniques. In particular, the heterogeneity of traffic conditions in congested regimes makes it hard to obtain a good matching between simulations and reality, thus preventing from obtaining reliable traffic state predictions beyond short time horizons (30 min). Indeed, traffic models are usually set on a unique choice of the fundamental diagram (flux-density relation), independently of the time of the day and the day of the week. Yet, model parameters may be time dependent, as some traffic conditions and driver behavior may be correlated to specific time ranges. Besides, data analysis techniques may help in opening new perspectives in the interpretation of data and the modeling of road traffic [3].

ACUMES Project-Team has an established experience in (macroscopic) traffic flow models, and disposes of a large set of data coming from 135 loop detectors placed on the freeways in the North of Marseille (A7, A51, A50 and A55). The data-set was provided by the *Direction Interdépartementale des Routes Méditerranée* (DIRMED) and covers 3 months, from September 1st to November 30th, 2015.

On the other side, ZENITH Project-Team masters the analysis of large amount of data, by means of parallel and distributed approaches [4, 5]. Here, analysis covers different goals such as feature correlation detection, clustering or time series similarity queries [6].

### Job Description

In this perspective, we aim to study the traffic data-set mentioned above by time series similarity techniques to extract new knowledge in synergy with modeling and simulation to improve traffic forecast and control.

The study will be based on the following steps:

1. **Data preparation.** Before analysis, data need to be sorted by sensor and direction, and aggregated over time periods (usually 1 or 6 minutes) to obtain averaged quantities measuring density, mean velocity and flow.
2. **Data analytics.** Analysis will rely on time series data mining using the Matrix Profile. The goal is to extract correlations between time series.
3. **Match with macroscopic models.** The information derived by the previous analysis will be gathered and possible impact on modeling techniques will be investigated. If possible, the combination of the two aspects (modeling and data analysis) will be tested on simple configurations.

The intern will be based in Sophia Antipolis, but will be supervised jointly by ACUMES and ZENITH teams.

## Requirements

- Last year of MSc (Master 2) in Mathematics, Engineering or Computer Sciences.
- Background in statistical methods.
- Knowledge of R, Python, Matlab.
- Experience in mathematical modelling is considered an additional plus.

## Salary Range

Approximately 1150 €/month (net).

## Starting date and duration

From April 2018 or later, for 6 months.

## Applications

Please send electronic application including CV, a statement of research interests, and contact details of reference persons to:

Dr. Paola Goatin

E-mail: [paola.goatin@inria.fr](mailto:paola.goatin@inria.fr)

**Contact:** Paola Goatin, PhD HDR  
Inria Sophia Antipolis - Méditerranée  
ACUMES Project-Team  
2004, route des Lucioles - BP 93  
06902 Sophia Antipolis Cedex, FRANCE  
Tel. +33 (0)4 92 38 78 34

**Web Site:** <http://www-sop.inria.fr/members/Paola.Goatin/>

## References

- [1] A. Cabassi and P. Goatin, *Validation of traffic flow models on processed GPS data*, INRIA Research Report no. 8382, September 2013.
- [2] J.-M. Loubes, E. Maza, M. Lavielle and L. Rodriguez, *Road trafficking description and short term travel time forecasting, with a classification method*, *Canad. J. Statist.*, volume 34, pp. 475-491, 2006.
- [3] M. L. Delle Monache, K. Chi, Y. Chen, P. Goatin, K. Han, J.-M. Qiu and B. Piccoli, *Big data and the fundamental diagram: the three phases of three-dimensional traffic*, in preparation.
- [4] S. Salah, R. Akbarinia and F. Masegla, *A highly scalable parallel algorithm for maximally informative k-itemset mining*, *Knowledge and Information Systems Journal*, 2017.
- [5] T. Allard, G. Hbrail, F. Masegla and E. Pacitti, *Chiaroscuro: Transparency and Privacy for Massive Personal Time-Series Clustering*, SIGMOD Conference, 2015.
- [6] D. Edine Yagoubi, R. Akbarinia, F. Masegla and T. Palpanas, *DPiSAX: Massively Distributed Partitioned iSAX*, *International Conference on Data Mining (ICDM)*, 2017.