A METHODOLOGY FOR ACCURATE LINK TRAVEL TIME ESTIMATION

E. AGAFONOV, A. BARGIELA

ISM group, School of Computing and Mathematics, The Nottingham Trent University, UK
Email: evgeny.agafonov@ntu.ac.uk

Abstract: Many traffic control strategies implemented into UTMC systems heavily rely on the efficiency and accuracy of travel time estimation. Travel time information is a fundamental parameter in many intelligent traffic and travel information systems and their success highly depends on the accuracy, availability and granularity of travel time estimates. Traditional survey-based travel time estimations are notoriously expensive and inefficient and the use estimates by surveys is restricted only to off-line monitoring and analysis. With the development and installation of Traffic Control Systems of a new generation that operate on the basis of analysis of real-time traffic measurements, it has become possible to carry out travel time estimations using the measurements taken by the control systems. Most of the systems, due to economical reasons, utilise single-loop detectors. It is known, that single loop detectors are incapable of accurate spot speeds estimation due to their physical limitations. Therefore, much effort has been put into investigation and development of methodologies of travel time estimation that rely only on occupancy measurements, as is the case with single-loop detectors. Unfortunately, most of the methods have been developed for highway traffic and strongly rely on some assumption, which make them inapplicable in urban areas. Another common disadvantage of the developed methods is that they require data aggregation, which, as it is well known, leads to the loss of important information. In this paper, we develop a novel methodology, which is free of the above disadvantages and has been shown to be successful in the application to urban links travel time estimation problem.

keywords: data processing, data analysis, data granularity, urban traffic, traffic control, traffic information, travel time, parameter estimation

1. INTRODUCTION

Early attempts to estimate link travel time have used the general traffic flow relationship in order to estimate speed [Wardrop, 1952], [Lighthill and Whitham, 1955], which in turn allows calculation of the travel time since the link length is a constant parameter. The relationship used is

$$speed = \frac{flow}{occupancy \cdot g} \qquad (1)$$

where $1/g$ is the average effective car length – the sum of the car length and the width of the loop detector in the correct units. In [Hall and Persaud, 1989], [Pushkar et al, 1994] the authors investigated the relationship in equation (1) and showed that the accuracy of the equation is a function of many factors including location and weather. They also suggested that the relationship is prone to a systematic bias with respect to occupancy.

Dailey [Dailey, 1997] attempted to improve the speed estimates by taking into account the stochastic nature of measurements of volume and occupancy. He considers measurements from a traffic management system to be realisations from statistical distributions. To address the variability of the observations he employs an extended Kalman filter. Using the filter Dailey derived estimates of the effective car length and speed and then combined them in order to obtain the travel time estimate. The link travel time is obtained as a polynomial of the following form.

$$\tau_{i+1} = 2x\left[\frac{1}{s_{i+1}+s_i} + \frac{(\Delta s_i)^2}{3}\left(\frac{1}{s_{i+1}+s_i}\right)^3 + \frac{(\Delta s_i)^4}{5}\left(\frac{1}{s_{i+1}+s_i}\right)^5 + \ldots\right]$$

where τ_{i+1} is the next estimate of the link travel time, x is the length of the link, s_{i+1} and s_i are $(i+1)^{\text{th}}$ and i^{th} estimates of speed accordingly and $\Delta s_i = s_{i+1} - s_i$.

The disadvantage of the Dailey's filtering approach is that it requires several parameters that must be properly estimated thus reducing its robustness and accuracy.

There have been attempts to improve speed estimation upon single loop detectors [Dailey, 1999], [Coifman, 2001]. Coifman's algorithm is based on explicit identification of sources of errors and suggests the way to reduce their impact.

Despite the efforts in the direction of improvement of speed estimates, all speed estimation based algorithms inherit a set of problems from the underlying traffic model based on the fundamental traffic flow relationship (1).

In [Dailey, 1993], the author presents a method of estimating the mean space speed that is independent of the mean value of the volume. In order to estimate speed of traffic flow he first estimates the mean delay time for propagation between two loops, which is essentially the mean of the link travel time. To estimate the mean of the link travel time Dailey models highway traffic as a continuous flow with some average concentration about which there is some statistical fluctuation. In order to estimate the delay time between two loops the cross-correlation function is used. Considering the property of the time series and the cross-correlation function, Dailey claims that the point of maximum of the cross-correlation function will correspond to the mean delay between the two series, which is the sought travel time estimate. There are several disadvantages of the cross-correlation approach. While cross-correlation can lead to accurate estimation of the mean of travel time, it does not provide the information on other characteristics of travel time as a random variable. Second disadvantage is in the need for aggregation of traffic counts in order to obtain the desired time series, which can lead to the loss of information and worsen the accuracy of the algorithm for certain traffic conditions.

In a more recent work [Petty et al, 1998] the authors developed a more general model of traffic flow, which has certain relationships to the Dailey's cross-correlation approach. Their methodology for estimating travel times between single loop detectors is based on a stochastic model, which suggests practical procedures that lead to accurate travel time estimates.

The authors assume that during a given interval of time the travel times may be regarded as drawn from the same probability distribution. They then estimate that distribution from the cumulative upstream and downstream arrival processes using least squares fitting method. Several difficulties associated with the above methodology have been recognised by the authors. First essential difficulty is the choice of the parameters of the time windows in which the upstream and downstream processes are considered. Having carried out empirical study of the effect of choosing different parameters of the window they found that an adaptive choice is required for generally satisfactory results in both congested and uncongested periods. Second difficulty is the choice of the 'stationary' periods. It is clear that the larger the interval the better precision of the scheme. The authors state that it is possible to detect the transitions in regime and choose the interval according to the available information. They do not provide the methodology but state that the measure of density can be used as an indicator of the transitions in traffic regime. Third problem is associated with the choice of appropriate aggregation level Δ. The authors provide a thorough empirical study of the effects of different levels of aggregation on the results of the scheme.

2. THE URBAN LINK MODEL

Most of the travel time estimation methods have been developed for freeway traffic. Unfortunately, urban links have several distinguishing features that do not allow direct application of the freeway methods without additional modifications. Firstly, the distance between consecutive junctions is relatively short compared to freeway links. On such a link, vehicle's travel time can be as little as 15 seconds, which makes the methods that rely on aggregated traffic counters inherently inaccurate. On urban networks one normally deals with sums of many link travel times in order to obtain the full journey travel time and therefore the estimation of individual link travel times must be as accurate as possible. This renders the aggregation-based algorithms an inappropriate choice. Secondly, an urban link normally has several unobserved (no traffic measurements are taken) side-roads, which allow vehicles that entered the link leave the flow unnoticed (no information on their leaving available) or new vehicles join the flow without being counted on the entrance, thus making the distributions of counters on the entrance and the exit of the link different. The second feature of urban links violates the assumptions of the methods that rely on comparison of the entering and exiting distributions, which therefore cannot be used without modifications that take the unmeasured leaving and joining of vehicles into account.

Taking into consideration all above points, a new methodology for link travel time estimation needs to be developed. In this paper, such new methodology is proposed.

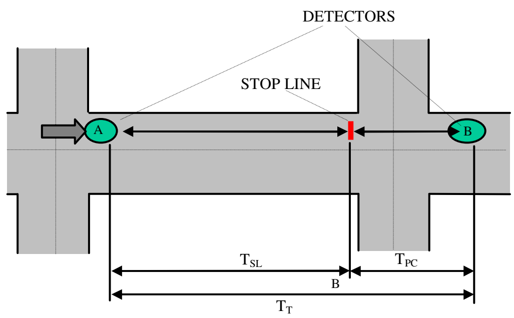

Figure 1. Urban link model and travel time components

A typical urban link is shown in Figure 1. The main feature that distinguishes a typical urban link from freeway links is the presence of a controlled junction in-between the detectors.

It is clear from the figure that total travel time (T_T) on an urban link can be split into two components, which are T_{SL} (SL stands for Stop Line) – travel time from the upstream measurement point up to the stop line and T_{PC} (PC stands for Passage of Crossroad) – travel time from the stop line to the downstream measurement point. The purpose of such division is that the total time may often include the delay induced by the traffic control. Therefore T_{SL} becomes the main measure of travel time in urban links. This measure basically provides the information about the link length and typical time that vehicles spend to traverse the link. It is also worth mentioning that some UTMC systems (for instance, SCOOT) use T_{SL} in order to optimise traffic and produce optimal control. The division of total travel time into two components can also be used for estimation of other important characteristics of traffic such as turning movements. In this paper, only T_T component of the travel time will be considered.

3. THE METHODOLOGY

Consider Figure 1 that illustrates a typical urban link. Assume that measurements that come from detectors *A* and *B* represent time of arrival of a vehicle at that particular detector. Then the set of measurements forms a time series, or a process, of arrival times. Let us denote by $\{A(t),\ t \geq 0\}$ the series of upstream arrival times and by $\{B(t),\ t \geq 0\}$ the series of downstream arrival times.

It is further assumed, that traffic forms a free flow from *A* to *B*; that is it does not stop at the stop line as shown on the figure. This requirement is necessary since only free flowing traffic carries information about the travel time between the nodes of the link since stopped traffic has a delay induced by the waiting time at the stop line. It is nevertheless possible to separate these times from the whole time of the travel with a simple pre-processing.

Series of arrival times $A(t)$ and $B(t)$ are discrete and countable. Let us enumerate and denote individual arrival times of the above time series by $\{a_i\}$ and $\{b_j\}$. It is intuitively appealing that if a_i is the arrival time of a vehicle at the upstream detector and b_j is the arrival time of the same vehicle at the downstream detector, then subject to the free traffic flow condition, a_i and b_j are related through a relationship of the form

$$b_j = a_i + \varepsilon_k \qquad (2)$$

where ε_k is a stochastic component that represents the delay between arrival of a vehicle at upstream and downstream points and takes into account fluctuation of speed with time. All ε_k, k=1, 2, …, can be considered as realisations of the same random variable ε. It also seems rational that for a short distance, as is the case with urban links, that variable ε be considered to have a symmetrical bell-shaped distribution, such as normal. Then ε can well be characterised by its mean and standard deviation: $\varepsilon = \varepsilon\,(T_T,\ \sigma)$. Then the problem of travel time estimation essentially reduces into the problem of estimation of the parameters of random variable $\varepsilon\,(T_T,\ \sigma)$. Another point to be mentioned is that vehicles in real traffic tend to keep microscopic structure of the flow over significant periods of time; that is the headways between consecutive vehicles change insignificantly. This feature of traffic lies behind the success of the developed methodology.

The methodology is as follows. Arrival times of downstream process are first being shifted into the past by a time τ; that is the following linear transformation is performed.

$$b^*_j = b_j - \tau \qquad (3)$$

The choice of parameter τ will be discussed later. Then the upstream time series is combined with the modified downstream time series in a way that keeps their order. The following Figure 2 illustrates this idea.

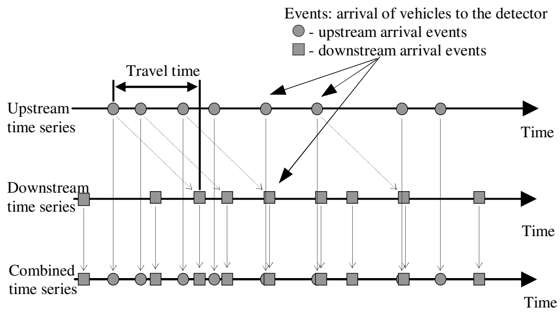

Figure 2. Merging upstream and shifted downstream time series

The dashed arrows on the figure point the related events, i.e. linking the upstream arrival events to the corresponding downstream arrival events. Next step is, using the combined process, to extract pairs of neighbouring arrival times that are from different sources. That is upstream arrival time is associated with the closest in time downstream arrival. The following Figure 3 illustrates the pair extraction procedure.

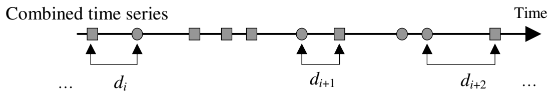

Figure 3. Extracting pairs of neighbouring events that come from different time series

As a result, algorithm produces series of pairs of events $P_k=\{a_i, b_j\}$ where a_i is the arrival time of the upstream event and b_j is the arrival time of the downstream event that form the pair. Several methods of extracting pairs P_k from the combined time series can be suggested. The most effective method would be to extract the pairs in such way, that the sum of corresponding time differences d_k is minimal among all possible choices of pairs. Although this method will yield the result with higher accuracy, it is not efficient. Complexity of the optimal algorithm is $O(N^2)$, where N is the number of elements in the sample of combined time series. A method that chooses pairs using local optimisation has been tested and showed that its accuracy is close to the optimal algorithm but has complexity of $O(N)$, that is it is much more efficient. An algorithm with local optimisation is described below.

After the upstream and downstream times series have been combined into the resulting time series, a step of the algorithm is given as follows. Assume that current position of an event in the resulting time series is k. Then there are six possible configurations of events following the event at k as given in the below Figure 4.

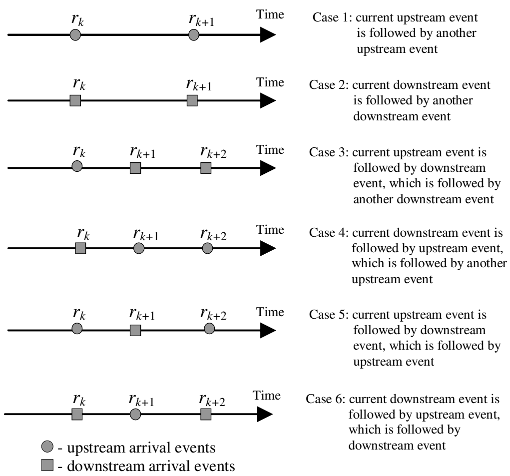

Figure 4. Combinations of events ahead of the current even at k

In the cases 1 and 2 of the figure, the algorithm does not produce pairs since the current even at k cannot be associated with the following event since they have the same origin. If case 1 or case 2 is encountered at step k, the algorithm makes event k+1 its current and repeats the procedure. In cases 3 and 4 there are clear pairs formed by events at k and k+1, since event at k+1 cannot form a pair with event at k+2 because of their same origin. Thus, if case 3 or 4 is encountered, a pair of events at k and k+1 is formed and the algorithm makes its current event for the next cycle the event at k+2. In case 5 and 6 it is not clear what pair is to be taken, the one formed by events at k and k+1 or events at k+1 and k+2. In this case, the algorithm forms a pair of such events, whose difference is smaller. That is if $r_{k+1}-r_k < r_{k+2}-r_{k+1}$ then algorithm produces a pair formed by events k and k+1, and advances to k+2 as its current event for the next cycle and produces a pair formed by k+1 and k+2 otherwise, advancing to k+3 as its current event for the next cycle.

The above algorithm ensures that if an event is associated with a pair, it is associated with that pair only. It also performs a local minimisation of the time differences associated with pairs as shown above for cases 5 and 6. This is also clear that the algorithm's complexity is $O(N)$ where N is the number of events in the resulting (merger) time series.

Once a set of pairs P_k has been obtained, an average of time differences between the events of a pair is calculated. It is calculated for currently applied time shift τ and is consequently considered a function of τ. Let us denote the arrival time of an upstream event that forms part of pair k by a^k and the arrival time of a downstream event that forms part of the same pair k by b^k. Then, consider the following *cost function*

$$F(\tau)=\frac{1}{N_P}\sum_{k=1}^{N_P}\left|a^k-b^k\right| \qquad (4)$$

where N_p is the number of pairs.

All pairs selected by the algorithm can be split into two classes – the ones that are formed by independent upstream and downstream events and those formed by the events that are related through relationship (2). Assume that the set of independent pairs is denoted by Ω_i and the set of dependent pairs is denoted by Ω_d. Then their cardinalities $N_i=|\Omega_i|$ and $N_d=|\Omega_d|$ satisfy $N_p = N_i + N_d$ and (4) can then be written as the following sum

$$\begin{aligned}F(\tau)&=\frac{1}{N_P}\Big(\sum_{k:P_k\in\Omega_i}^{N_i}\left|a^k-b^k\right|+\sum_{l:P_l\in\Omega_d}^{N_d}\left|a^l-b^l\right|\Big)=\\ &\frac{N_i}{N_P}\left(\frac{1}{N_i}\sum_{k:P_k\in\Omega_i}^{N_i}\left|a^k-b^k\right|\right)+\frac{N_d}{N_P}\left(\frac{1}{N_d}\sum_{l:P_l\in\Omega_d}^{N_d}\left|a^l-b^l\right|\right)=\\ &q(\tau)\overline{X}_i(\tau)+(1-q(\tau))\overline{X}_d(\tau);\quad 0\le q(\tau)\le 1 \qquad (5)\end{aligned}$$

where $\overline{X_i}(\tau)$ is the average of time differences $\{d_i\}$ for independent pairs, $\overline{X_d}(\tau)$ is the average of time differences $\{d_j\}$ for dependent pairs and $q(\tau) = \frac{N_i}{N_p}$

It is difficult to express explicitly the relationship between differences $\{d_i\}$ produced by independent pairs and time shift τ but it can be anticipated that such a relationship will be weaker (or more irregular) than that of the differences produced by dependent events. Using (2) and (3), the relationship between dependent difference d_j and τ can be expressed as

$$d_j(\tau) = b_j^* - a_i = b_j - \tau - a_i = \varepsilon(T_T, \sigma) - \tau \qquad (6)$$

It can be shown that if τ_{m} is defined as

$$\tau_{\mathrm{m}} = \arg\min_{\tau} \mathrm{E}\,|\,\varepsilon(T_T, \sigma) - \tau\,| \qquad (7)$$

then τ_{m} is the median of the random variable $\varepsilon(T_T, \sigma)$. In the case when $\varepsilon(T_T, \sigma)$ has a bell-shaped symmetrical distribution and its first two moments exist and are finite, then the median equals the mean and equals the mode of $\varepsilon(T_T, \sigma)$. Therefore, if the mean of independent differences does not change significantly with τ, then it is reasonable to expect (5) to reach its minimum in the point $\tau = T_T$.

The above arguments can be considered to be a heuristic, as it cannot be proved formally in its general form. Nevertheless, it is strongly appealing and a number of experiments with real traffic data have shown it to be correct. Another point for the correctness of this approach is that when τ approaches T_T, coefficient $(1-q(\tau))$ in (5) tends to grow while $q(\tau)$ itself decreases provided that σ of $\varepsilon(T_T, \sigma)$ is smaller than the average time gap between consecutive arrival events. This, in turn, makes the second term of the sum (5) to have stronger influence on the whole sum and thus minimum in the point $\tau = T_T$ is being stronger emphasised.

This has been the basis of the proposed methodology. Several modifications of the presented algorithms can be suggested, such that extract other useful information from the time series of arrival events. The methodology has clearly several advantages over previously developed approaches. The advantages can be briefly summarised as follows.

1) The methods of the methodology operate at the resolution of data and no aggregation is required thus yielding the most accurate estimations possible
2) The methods are robust towards undercounting and overcounting of vehicles as well as uncounted vehicles that are leaving and joining the link. Such events will be related to the independent pairs or excluded from consideration automatically
3) The presented algorithm is efficient. The efficiency of the algorithm is kN, where k is the number of steps (depends on the method used to generating the steps), and N is the number of events in both the upstream and downstream series of arrival times. It is worth noticing that the correlation-based algorithm is also MN, where M is the number of time lags and N is the number of elements of the aggregated arrival time series (both the downstream and the upstream), but M is normally bigger than k and N does not depend on the number of actual events and in the case if there is small number of arrival events, the methodology of this study will show better performance
4) The set of time differences produced by the pair extraction algorithm in the point of minimum of the cost function can be used to estimate other useful parameters, such as variance, of the travel time as a random variable.

Several methods of choosing the step $\Delta\tau$ of the algorithm can be proposed. The simplest way is to let $\Delta\tau$ be a constant equal to the smallest unit of the time resolution of the data. In this case the optimum solution is guaranteed to be found.

5. EXPERIMENTAL RESULTS

The proposed methodology has been used in an empirical study of a SCOOT controlled region of the town of Mansfield, UK. The following figures present the plots of the cost functions calculated for one of the links of the region.

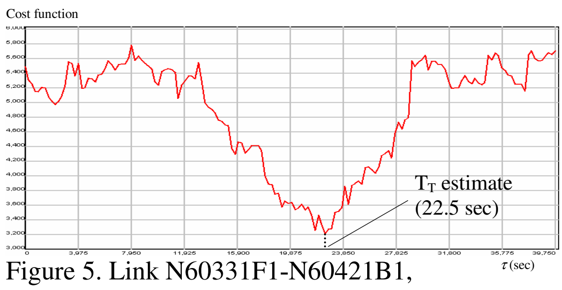

Figure 5. Link N60331F1-N60421B1,
19 February, Tuesday, 1999, 7 pm

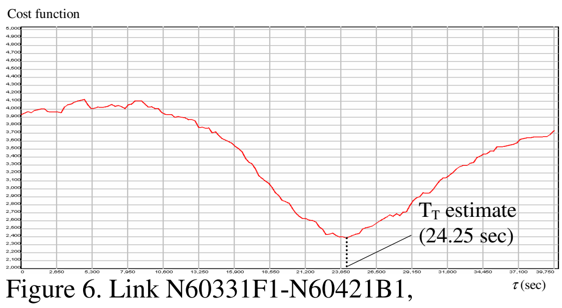

Figure 6. Link N60331F1-N60421B1,
19 February, Tuesday, 1999, all day

Figure 5 shows a plot of the cost function calculated over one hour period of time starting from 9 pm. The point of minimum of the cost function is clearly emphasised and yields the estimate of the mean travel time equal 22.5 seconds. Figure 6 presents a plot of the cost function calculated over all day data. The point of minimum at 23.75 seconds is the estimate of the mean travel time over this period of time.

Figure 7 illustrates the cost function calculated over a broader range of time shifts (from –4 min to 4 min). It is clear from the plot that there is only one point of the "best match" between the upstream and downstream series which is located at the point of 24.25 seconds. The cost function behaves as expected: its bigger values correspond to the areas of time shift τ where most of the events are uncorrelated and smaller values correspond to the neighbourhood of 24.25 seconds where some events are correlated.

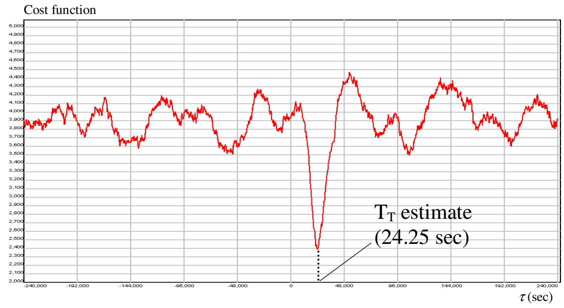

Figure 7. The cost function calculated over a broader range of time shifts.

6. CONCLUSION

We have presented a novel approach to the travel time estimation problem which has several advantages over existing methods. Although the methodology has been developed in the context of urban traffic and travel time estimation problem it can be used in other areas where a problem can be reduced to finding a delay between two point processes.

REFERENCE

Coifman, B. 2001, "Improved Velocity Estimation Using Single Loop Detectors". Transportation Research: Part A, vol 35, no 10, pp. 863-880

Dailey, D.J. 1993, "Travel time estimation using cross-correlation techniques". Transportation Research, Part B, 27B (2), 97-107

Dailey, D.J. 1997, "Travel time estimates using a series of single loop volume and occupancy measurements". Transportation Research Board, 76th Annual Meeting, Washington

Dailey, D.J. 1999, "A statistical algorithm for estimating speed from single loop volume and occupancy measurements". Transportation Research B. Vol. 33, pp 313-322.

Hall, F.L., Persaud, B.N. 1989, "Evaluation of speed estimates made with single-detector data from freeways traffic managemtnt systems". Transportation Research Record, (1232):9-16

Lighthill, M., Whitham, G. 1955, "On kinematic waves, a theory of traffic flow on long crowded roads". Proc. Royal Society, London, 229, 317.

Petty, K., Bickel, P., Jiang, J., Ostland, M., Rice, J., Ritov, Y., Schoenberg, F. 1998, "Accurate estimation of travel times from single-loop detectors". Transportation Research, Part B: Methodological, 32(1):1-17

Pushkar, A., Hall, F.L., Acha-Daza, J.A. 1994, "Estimation of speeds from single-loop freeway travel times in real time from flow measurements". Journal of Transportation Engineering 122, pp. 185-191

Wardrop, J. 1952, "Some Theoretical Aspects of Road Traffic Research". Proc. Inst. Civil Engineers, Part 2, pp. 325-378

Evgeny Agafonov received the BSc degree in Mathematics and Information Science from the Chelyabinsk State Technical University, Russia, in 1997. From 1997 to 1999, he worked as a Unix Systems Administrator and Software Developer at the Scientific Library of ChSTU and in 1999 he joined the Intelligent Simulation and Modelling (former RTTS) group at the Nottingham Trent University as a PhD student.

Mr Agafonov's research interests are in data mining, information theory, probability theory, stochastic processes, systems modelling and simulation

Andrzej Bargiela is Professor of Simulation Modelling at the Nottingham Trent University and is a leader of the Intelligent Simulation and Modelling group. He is the Research Coordinator in the School of Computing and represents the Nottingham Trent University on the U.K. Engineering Professors Council. His research interests are focused on fundamental issues of measurement and information in the context of modelling and simulation of various physical and engineering systems. His research has been published widely and he has been invited to lecture on simulation and modelling at a number of universities.