# TOOL FOR ANALYSIS OF EXTRAORDINARY STATES IN COMPLEX TRAFFIC NETWORKS

**Ing. MARTIN ŠKODÁČEK**
**Prof. Ing. PAVEL PŘIBYL, CSC**
***Department of Transporting Systems***
***Horská 3, Prague 2, Czech Republic***

***Summary:*** *Urban areas are equipped with hundreds of traffic sensors. The basic problem is to transform the tremendous time series of measured traffic parameters to the simple parameter which estimates the quality of services over this network. This happens mostly based on the judgment of dispatchers sitting in control centers. The paper presents a method for evaluating comprehensive transport network based on the methods of artificial intelligence. The collected data are transformed into 3D state space where is being running real-time traffic data analysis instead to provide analysis of time series over a set of measured detectors. State space transform is provided by Principal Component Analysis application. It enables significant data reduction and their representation in new dimension. New data are represented by principal (main) components. These outputs usually have lower volume of data and reduced dimensionality, in comparison with the original input data. In further research step is to be introduced some classification method for recognition of extraordinary traffic situation.*

***Keywords:*** *traffic, quality, reduction, state, space*

## 1. INTRODUCTION

Intelligent traffic system holds a significant role in field of traffic. The quality of traffic management significantly affects our everyday lives. Many densely populated areas are equipped with sufficient amount of strategic and intersections traffic detectors. The system manages traffic streams by traffic lights, variable message signs and by other devices. An absenting part inside of the current telematic systems is inadequate system for evaluation of collected traffic data for global traffic model covering the managed network. Operators in traffic surveillance centers are evaluating situation based on visual survey by CCTV or using simple mechanism for searching locations with too low speed. This paper describes mathematically supported algorithm of traffic quality estimation, based on combination of two methods:

- Current state of traffic situation based on detectors outputs in discrete time slices is conceived as point in a State Space (SS) which shall represent quality of traffic, known as Level of Service (LOS);
- Transformation of each data set in the discrete time into 3D space is carried by Principal Component Analysis (PCA)

Output of the above described system composed of PCA and SS approaches is a state vector in new dimension. Quality of traffic can be classified into categories according position of this vector by couple of methods. One of them uses artificial intelligence methods based on searching relationship and similarities between on-line collected data and historical models of traffic data. A plenty of foreign literature confirms, that by closer evaluation of historical data samples in state space we can better understand characteristics of present traffic states. On this base is possible to provide more effective traffic management system and supply existing system by effective incident detection identification system, queue prediction, etc.Trajectory of LOS viewed in state space gives also very good visual overview about development of quality of traffic. The extraordinary event is possible to recognize directly, feature development of LOS is also possible to see on simple way. State space is generally considered as a new modern approach and it is not used more widely for traffic analysis.

## 2. TRAFFIC DATA COLLECTION

First of all, for considered traffic quality estimation method is necessary to specify data source. Traffic situation can be examined by many approaches from traditional traffic data, floating car data, videodetection or other surveillance method. In our research we work with traditional traffic data collected by traffic detectors. Traffic detectors determined for traffic data collection, in technical terminology (Přibyl, 2005) called strategic detectors and typically provide standard traffic data such as vehicle volume (vehicle count per 3 min), average or unit speed (km/h) and detector loop occupancy (percentage unit) in defined aggregation interval, e.g. each 3 or 5 minutes. For relevant traffic state description is crucial to use more than one traffic data type. Detectors should be placed on "strategically interesting positions", which means on measuring sites with stable conditions.

In our research we use real data from seven measuring profiles based on induction loop detectors installed under concrete surface traffic deck regularly placed in a consecutive cascade on a 10 km highway section near Prague. Traffic detectors covers all traffic lanes on monitored section located each 0.5 to 1 km. Detector profiles location is shown on Fig ***3***.

Traffic data consisting of vehicle volume, detector occupancy and average speed from each detector profile are aggregated into 3 minutes intervals are transferred to highway surveillance center to be saved into the database on real-time basis. For the analysis purposes, data were exported from database and split up into three independent groups, namely volumes, detector occupancies and average speed. During further analysis performed in MATLAB® environment were data sorted according to date and additionally filtered using basic linear regression algorithm. We obtained reference flows, which were used for classification algorithm training. We have also chosen a several flows representing non-standard situations, such as queues or restrictions caused by accidents.

## 3. LITERATURE AND RESEARCH REVIEW

State Space analysis seems to be a convenient method suitable for traffic states description. According to articles focused for traffic data evaluation, state space offer several advantages, mainly:

1. Ability to easily handle multiple inputs and outputs systems;
2. System modelling includes both internal state variables as well as output variable;
3. Directly provides a time-domain solution;
4. Matrix input into vector on the output modelling is highly efficient from computational standpoint;

State space as the representation of all possible states of dynamic system is convenient for a traffic states description. Studies discussing traffic data modelling and prediction show an application of state space for an urban or freeway stretch data (Stathopoulos & Karlaftis, 2002). Authors of other texts described creation of vectors based on traffic stream model data and studied time series chronological, trajectory of vectors in state space (Lan et al. 2007). Different researches were based on implementation into STARIMA models, based on detector location and relationships between sites (Kamarianakis & Prastacos, 2002).

State space can be a base algorithm for Kalman filter traffic state estimator (Wang & Papageorgiou, 2004) or adaptive freeway traffic state estimator (Wang & Papageorgiou, 2008) utilized. Finally, state space has been presented as part of highly sophisticated ITS application of origin-destination estimation (Zhou & Mahmassani, 2007). Authors used state-space model because of its complexity of using all necessary variables, such as historical demand parameters. Investigation of the state space problematic shows, that state space is a modern and convenient method to transform traffic data from more detectors into "state" which will represent a qualitative parameter for further traffic examination. To transform considered discrete slices into a better evaluable form is necessary to use convenient reduction approach.

Principal Component Analysis (PCA), considered as a convenient reduction method, has been discussed in a number of articles. In all papers, main benefits such as formation of new orthogonal dimension, where original data are reduced, non-correlated resulting variables and larger variability of original values covered by the first of principal components of PCA were highlighted. Series of studies were carried out in the last years to find a method for detection of non-standard states from traffic data and its classification (Jin et al. 2008; Foester, 2008). For example in real-time adaptive on-line traffic incident detection article (Xu et al. 1998) were real-time adaptive procedures for obtaining principal components from traffic data, namely speeds and flow densities described. Adaptive means that PCA was formed into neural network structure to create traffic states from the dynamic model. Traffic states were finally evaluated by intelligent classificator. Other texts from the same authors were for huge amount of traffic data compression or missing data supplementing approaches focused (Qu et al. 2007). Some researchers studied day flows reduction into principal components and its behaviour related to cluster analysis. All studies used real historical traffic data and showed high flexibility of proposed method. Negative features like unsuitability for time series analysis because of its variability caused by component score relations were also mentioned. Investigation showed that different data matrices consisted of very similar data series can have different dimension in space, which excludes independent conversion. This brings us to conclusion that we need another method to pre-process data before PCA analysis.

## 4. MATHEMATICAL BACKGROUND

In this section we shortly describe mathematical operation necessary for applying studied method on real traffic data. We start from idea, that traffic state in any city area or freeway stretch can be represented by predefined and ultimate number of states (normal state, slightly crowding traffic, queues, etc.). This description is very close to almost known and often used Level of Service (LoS) categorization, famous from dynamic web maps, variable message signs or radio reports. Main goal is to use the state space for description of traffic situation by states similar to LoS, which will offer more complex macro view on monitored area than offers typically used LoS approaches (Pribyl, 2005) for traffic state estimation. Typical traffic states, which cover all typical traffic situations (normal, medium or dense traffic, congestions), we obtain from historical data. Based on historical data we classify real-time data into clusters, where each cluster represents trained traffic state. Due to the effort to use more than one traffic parameter for traffic quality estimation we assume to use multiple traffic inputs of volumes, speed, and occupancy from a certain number of detectors, we execute traffic data analysis in multiple spaces. To reduce a big amount of data, we use PCA, while preserving its variability, clearing out correlations and creating new orthogonal dimension for better evaluation. After applying PCA reduction on system state, we get a single vector, which belongs to the defined state and represents traffic quality state.

### 4.1. State space description

System state description appears from physics. States are based on vectors, which contain all the information about the system. State vector is linearly independent set of variables. State space is the set of all possible states of a dynamic system, where each system state corresponds to a unique point in the state space [19], defined generally by two equations

$$\frac{dx(t)}{dt} = A(t).\bar{x}(t) + B(t).\bar{u}(t), \text{ where } B = \begin{vmatrix} q_t^1 & \varkappa_t^1 & r_t^1 \\ q_t^2 & \varkappa_t^2 & r_t^2 \\ q_t^n & \varkappa_t^n & r_t^n \end{vmatrix} \quad (1)$$

Is called state equation and represents change of input matrix. Herein A represent a state matrix, x(t) is input state vector, B is input data matrix and can be defined for example for vehicle volume as it is shown above (1…n is a number of considered detector sites). Finally $\bar{u}_t$ is called control vector and represents state change. Second equation is called output equation and can be written in the following form

$$\bar{y}(t) = C.\bar{x}(t) + D(t).\bar{u}(t) \quad (2)$$

Where, C is output data matrix, D is input weight matrix and can be understand as an internal relationships matrix (e.g. detector position), and y(t) is the output vector. The output vector represents searched traffic quality parameter, which can be presented in many grades, depending on a suitable classification method, for example:

$$\bar{y}(t) = \begin{cases} \text{normal state/incident}\{0,1\} \\ \text{historical model } q, \varkappa, r \ \{\text{např.} \ q_{[5-10]}, \varkappa_{[10-14]}\} \\ \text{level of service } \ y\{A-F\}, y\{1-5\} \end{cases} \quad (3)$$

Generally, each set of values, represented by discrete slice shown on the Fig. 1, represent the state space of a dynamic system. In case of traffic data, these values are represented by real traffic flow parameters. Traffic state can be derived from several parameters simultaneously, as shown on Fig. 1. Future states can be obtained via step-predictor or other prediction methods (Kalman filter, etc.). The number of dynamic system freedom degrees represents the dimension of phase space. In fact, only a certain number of variables are needed to fully describe the system (Terman & Izhikevich, 2010). Dynamic system development corresponds to the trajectory of vectors motion in this space. In state space we observe traffic data at intervals of aggregation as slices, as it is shown on Fig. 2 right. Aggregation slices, visualized as connection between flows, represent system state, which is reduced and examined in next steps for traffic quality monitoring. Reduction method, considered as transformation method, should eliminate less significant and correlated variables and transform a large amount of input variables generated by traffic detectors into a few parameters to create vector for further examination of states a should represent a quality of traffic, known as Level of Service (LOS).

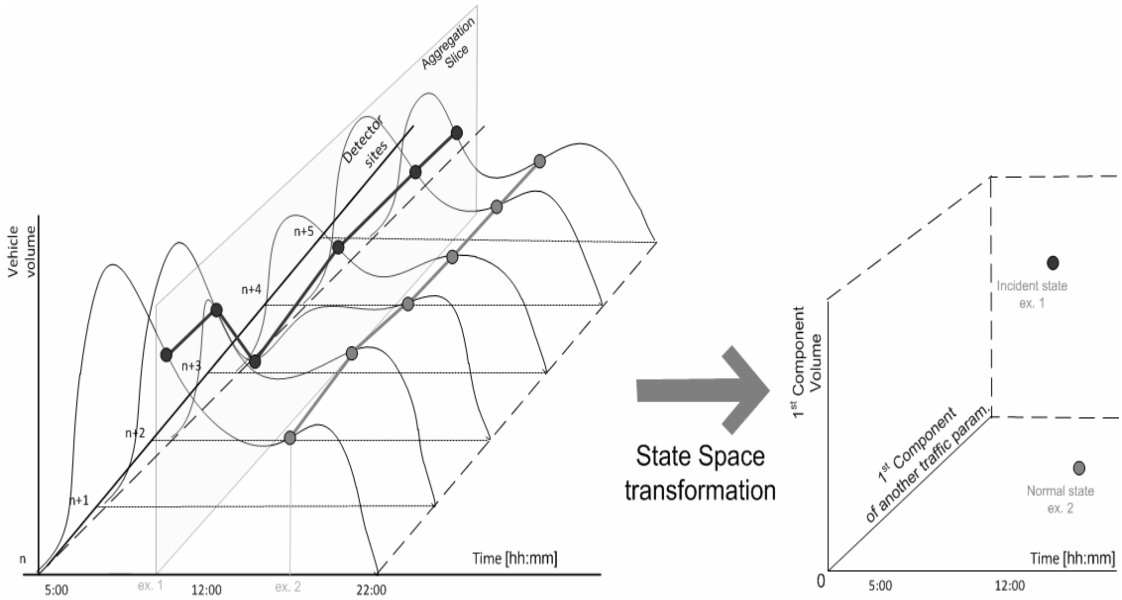


***Figure 1.*** *State Space transformation example. We assume that in each aggregation interval traffic data are produced on each traffic detector. In state space, we analyze all data types (volumes, speeds) from all considered detectors all at once. Data in each aggregation interval (rectangle area) creates a discrete time slice, representing the system state. Our goal is to transform these data using convenient reduction method to get a single vector. One single vector in new dimension allows us to classify global traffic state of observed city area or line stretch*

### 4.2. PCA reduction

Principal Component Analysis was established by Pearson in 1901 and is one of the oldest methods for multiple data analysis. PCA allows transformation of entered data to a smaller spatial dimension, while preserving the original variability. PCA creates new vectors, so that each of the new vectors is a linear combination of the original vectors. Output of this algorithm are primary principal components, secondary outputs are component scores and variances (latent variables). The correlation values of traffic data can be plotted in a new dimension in the

2-D views (main component of intensities, occupancy, or speeds of several detectors), or 3D dimension. The following text explains the PCA analytically.We have a matrix D (m × n), where m is a number of rows and n of columns, representing traffic data set

$$x(t) = [x(1), x(2), \dots x(n)] \tag{4}$$

Where, $x(t) = [x_1(t), x_2(t), \dots x_n(t)]$ and variables are correlated to each other. Variables $x_1$, $x_{2,} \dots$ represent various data traffic data types. Variable T represents a period of monitored flow and t is aggregation time. Because these variables are correlated to each other, another group of uncorrelated vectors exists and is given by equation

$$y(t) = [y(1), y(2), \dots y(n)] \tag{5}$$

Where, $y(t) = [y_1(t), y_2(t), \dots y_m(t)]^T$ and $y_k$ (t) are linear combination of x(t) given by

$$Y = P.X \tag{6}$$

Where Y represents a new dimension generated by multiplication of P and X matrices, where P is projection matrix , where and X represents original data matrix. Using PCA, we obtain new dimension Y consisting of principal components (PC) from original data matrix X. First few principal components cover the largest variability part of original data. Since there are for reconstruction typically used only first 2-3 principal components, information loss results in creation of error matrix (Meloun & Militký, 2006). There are several ways how to implement PCA and calculate principal components, most frequently used singular value decomposition (SVD) or conversion through covariance matrix (Shlens, 2009). The number of used principal components leads to standard PCA mathematical model.

**4.3. QoT Mathematical model**

Based on knowledge of state space transformation and PCA reduction we designed a mathematical model for traffic quality (QoT) estimation, based on combination of these methods. As an input we use traffic data collected by traffic detectors located in monitored area namely vehiclevolume, detector occupancy  and average speed . For each discrete time slice we transform these data into a matrix structure

$$s(t) = \begin{pmatrix} q_t^1 & \varkappa_t^1 & r_t^1 \\ q_t^2 & \varkappa_t^2 & r_t^2 \\ q_t^n & \varkappa_t^n & r_t^n \end{pmatrix} \tag{7}$$

Where t represents timestamp.Each of these matrices representing traffic state is consequently reduced to several principal components by PCA with minimum information loss. However, correlations among measuring profiles are effectively removed.

$$s(t) \xrightarrow{PCA\ reduction} PC(t) \tag{8}$$

The reduced states represented by the first principal components of each reduced data type are stored in a database. For the currently analysed data, the database is named RS, historical database containing data collected for a long period is named Θ. Finally, traffic quality level is determined by comparing currently analysed reduced states against model states, usingthe following formula.

$$\{RS(t), \Theta(t)\} \Rightarrow QoT_{Level}(t) \tag{9}$$

The output levels can be defined either heuristically or based on long term experience.

## 5. DATA EXAMINATION AND TRAFFIC QUALITY ESTIMATION

Research oriented to estimation of quality of traffic above a set of detectors described in this paper is based on working with real data collected on D1 highway near Prague, as was noted in chapter 3 – Traffic data collection. From both sources, model and incidental, a two hour flows around incident detection time were cut out. Thus for further analysis we used six data matrices. All matrices were rotated by ninety degrees, to get discrete slices. In next step we linked model and incident matrices into one matrix for each data type separately, which represent dimension unification necessary for next step. Finally we transformed discrete slices using PCA method into a new dimension. Each time aggregation formed by discrete slice is in new PCA dimension represented by simple value called state vector instead of array given by amount of traffic detectors. This state vector is the primary output of algorithm described above and represents traffic quality parameter. Output state vector is three dimensional and could be formed by: first three main components of chosen traffic data type (Fig. 4), 1st component of volume, speed and occupancy data (Fig. 5) or by two traffic flow parameters (e. g. volume and speed) and time, when traffic data were recorded. This quality of traffic output vector is ready to classify into predefined states based on knowledge of all states from historical data of trained model using methods of artificial intelligence and training recognition structures. Recognition algorithms can be based on using carefully created historical models. The goal is to recognize position of real-time state vector in new 2-D or 3-D space according to model vector from historical database indicated in the same day time using classification algorithms similar to Euclidean distance measurement or highly professional methods like classification three, neural networks or pattern recognition.

Analysis output examples for non-incident and incident traffic situation are shown on following figure 4. We see two clusters, which represents typical states given by trained model and incident data. Vectors matching incidental situation are situated out of model vectors cluster. Few outliers between two clusters represent transition states when traffic situation in incident flow was changing from normal to incident state and back.

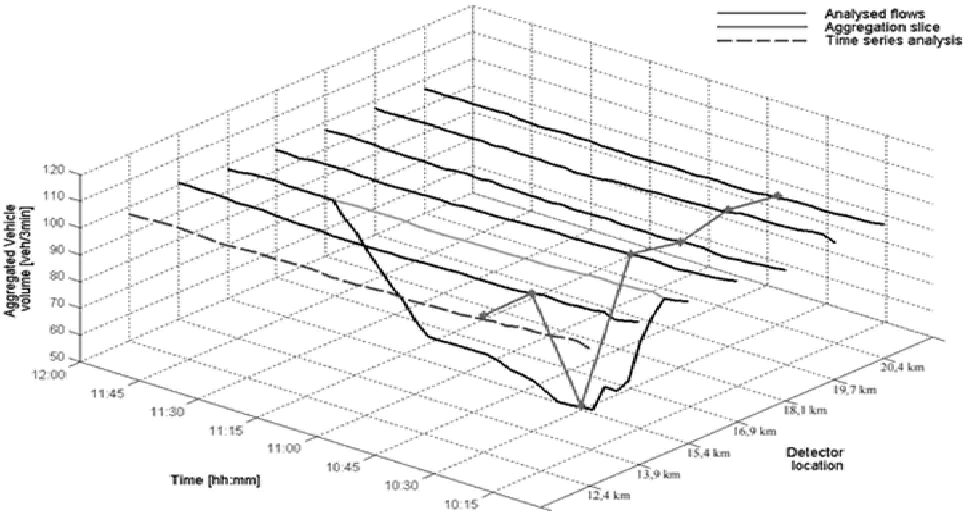


***Fig 2.*** *Time-space vs. State-space analysis. Figure shows traffic volume flows measured on seven detectors. Time series analysis means that each detector is analyzed separately (dashed flow means one detector). State space analysis means that all detectors are analyzed together, where data sliced in aggregation interval (grey slice connecting all detectors) are transformed by particular method into new space. Decrease of traffic volume on one of detector means that incident occurs. Average (historically typical) flow is showed by light grey. Analogical situation are valid both for occupancies and average speed flows (not shown in this paper)*

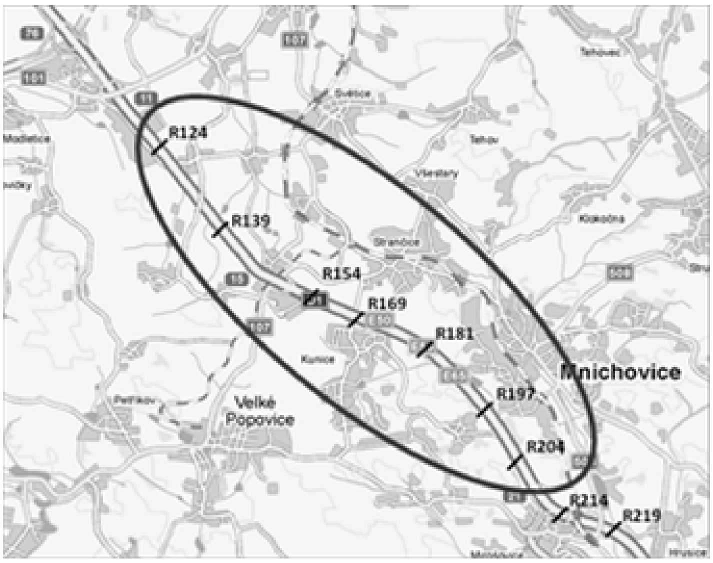


***Fig 3.*** *Detector location on D1 highway near Prague*

Vector distribution of model and incident data is equal, which is given by linked model and incident matrices. If the PCA transformation was performed separately for the model and for incident data, dimensions would not be identical. New orthogonal dimension is given by component scores which are different for each data set. PCA reduction of joined input matrix consisting of discrete slices will create both common dimension for model and incident. This step is absolutely necessary for using examination techniques based on knowledge of historical model or predefined system states.

Recognition of traffic state or both traffic quality parameter using only one traffic parameter, as it is shown on fig. 4, does not leads to reliable results as it is referred in text (Přibyl et al. 2003). Using only one data type in original or reduced dimension leads to miss important events, which may occur in the selected parameter later or due to the monitored area structure may not occur at all (Přibyl & Mach, 2003). Using first component of volume and occupancy, volume and speed, occupancy and speed or all three parameters together will lead to more accurate traffic quality estimation outputs than using only one traffic parameter.

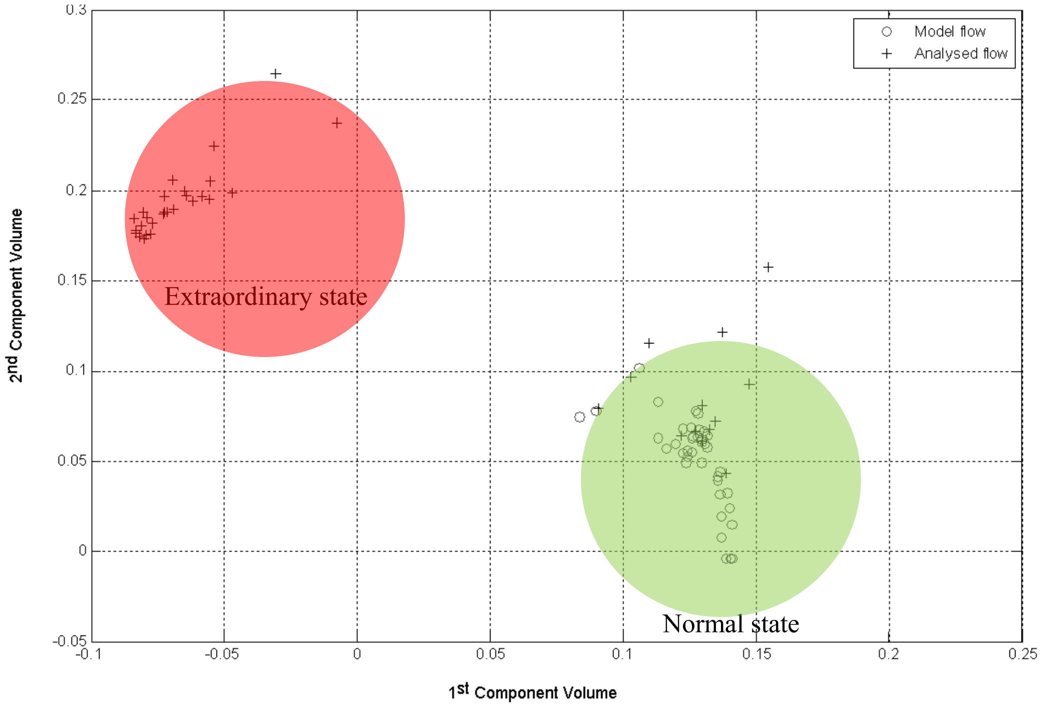


***Fig 4.*** *First and second PC coordinates of normal vs. extraordinary traffic situation based only on volume data projected into new dimension*

In reduced dimension, first principal component covers the biggest part of variables from the original dimension. In case of vehicle volume data analysed in example show on Fig. 2, further research focused for each principal component showed, that first three components covered overall 95% variability of input values. The first PC of volume data corresponds to approx. 80%, of original vehicle volume variance, the second PC after all correlation with the first PC corresponds to approx. 10% and third PC to 5% of original volume variance. First principal components represent for each discrete time slice one reduced state of traffic quality and as well as one record inabove mentioned $RS$ or $\Theta$ databases.

Considering the knowledge discussed in paragraph above, for relevant traffic quality estimation will be more convenient to use first principal component of two or three traffic flow data types, such as volume, speed and occupancy. It is both possible to combine two traffic flow data types with time axe for well-arranged visualization. The final reduction of three data types of model and incident flow mentioned above where traffic quality can be potentially estimated from first component of each reduced data type is shown on fig. 5. On figure is also noticeable benefit of variables reduction. Two hours flow of seven detector features from 280x3 records, when three data types (vehicle volume, speed, occupancy) are collected. In new space are these data reduced to just 40 records, which represents vector composed of three various first components.

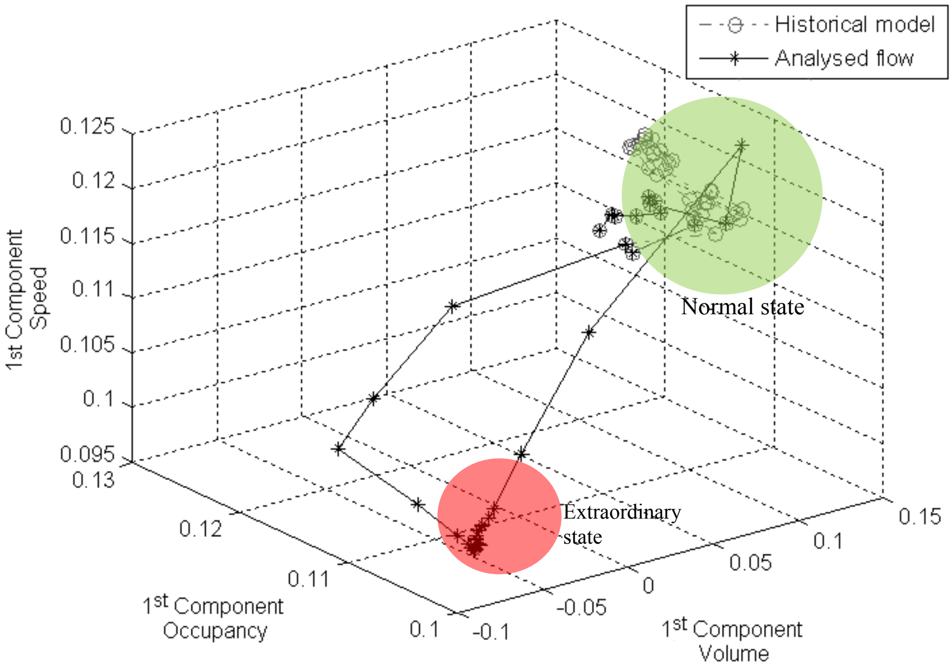


***Fig 5.*** *Traffic state vectors in the new reduced dimension. The figure compares two-hour flows from the example shown above, but with common component dimension. Grey dashed plots represent historical model from equivalent day and time period. Black plots represent flow with detected incident. We see that incident values lies out of model cluster and can be identified based on distance from the model state in discrete time*

## 6. CONCLUSION AND FURTHER RESEARCH

Traffic data analysis for the purpose of designing ITS system for traffic quality estimation brings undisputed advantages, such as quick incident and queue detection, and creates the head stone for reliable traffic management system. We know from experience, those operators from the number of surveillance video displays on the video wall are not able to recognize bad traffic situation immediately and respond appropriately. When this process is automated, effectivity of data evaluation increases and operators will only verify and confirm traffic states and take correct action to minimize all external and unwanted influences. That is the reason for historical traffic data investigation, and searching for new approaches to its evaluation and archivation. State space using principal component analysis for transformation, supplemented by appropriate classification algorithm provides optimal basement for this research.

The importance of this research lies in three main facts. First, good traffic quality evaluation system will help with integration with variable message signs, mobile navigation devices, RDS-TMC or internet to drivers to select an optimum route. Secondly, the above mentioned on-line system could provide an accurate knowledge of existing traffic conditions in order to guide traffic control more effectively. Lastly, efficient dispatching of emergency services is enhanced.

The mathematical model described in this study may represent the core of such intelligent traffic system. Modern mathematical approaches lead to use of artificial intelligence elements, where system is trained of all possible states from historical data. Based on the memorized states it is able to evaluate the level of quality of traffic on roads and inform drivers about conditions. This concept represents an alternative to a simple roads expanding.

**References**


[1]. Qu, L. HU J., Zhang Yi., 2007. A Flow Volumes Data Compression Approach for Traffic Network Based on Principal Component Analysis., Proceedings of the IEEE, pp. 125 - 130.

[2]. QU Li, Li Li, Hu J., Zhang Yi., 2009. PPCA-Based Missing Data Imputation for Traffic Flow Volume: A Systematical Approach. IEEE Transactions on ITS, pp. 512 – 522.

[3]. Foester G., 2008. Traffic state estimation using hierarchical clustering and principal components analysis. Fraunhofer Institute for Transportation and Infrastructure Systems.

[4]. Xu, H., Kwan, C. M., Hayness, L., Pryor, J. D., 1998. Real-time adaptive on-line traffic incident detection. www.sciencedirect.com, pp. 173 – 183.

[5]. Jin, X., Zhang, Yi, LI Li, HU J., 2008. Robust PCA-Based Abnormal Traffic Flow Pattern Isolation and Loop Detector Fault Detection., www.sciencedirect.com, pp. 829 – 835.

[6]. Kamarianakis Y., Prastacos P., 2002. Space-Time modeling of traffic flow. http://www-sre.wu-wien.ac.at/.

[7]. Lawrence, L.W., Sheu, J.B., Huang, Y., 2007. Investigation of temporal freeway traffic patterns in reconstructed state space.Transportation research part, Elsevier, pp.116 - 136.

[8]. Meloun M., Militký J., 2006. Kompendium statistického zpracování dat., 2nd Edition, Academia, ISBN: 80-200-1396-2, pp. 237-272. (in Czech).

[9]. Meloun M., Militký J., Hill M., 2005. Statistické zpracování dat v příkladech., Academia, ISBN 80-200-1335-0, pp. 59-93. (in Czech).

[10]. Pavelka, M., 2007. Analýza dopravní situace na trase Šterboholy-Barrandov (Prague, Jizni Spojka)., Research report for SafeTun project (in Czech)

[11]. Přibyl P., 2005. Inteligentní dopravní systémy a dopravní telematika., CTU mimeographed, CTU Publishing, 2005, ISBN 80-01-02811-9. (in Czech)

[12]. Přibyl P., Mach R., 2003 Řídicí systémy silniční dopravy., FTS CTU, CTU Publishing, ISBN 80-01-02811-9, pp. 161-175. (in Czech)

[13]. Přibyl, P., Svítek, 2001. M., Inteligentní dopravní systémy. BEN Publishing. (in Czech)

[14]. Shlens J., 2009. A Tutorial on Principal Component Analysis., Center for Neural Science, New York University, http://www.snl.salk.edu.

[15]. Stathopoulos, A., Karlaftis, M.G., 2003. A multivariate state space approach for urban traffic flow modeling and prediction. Transportation research part, Elsevier, pp. 121 – 135.

[16]. Škodáček, M., 2011 Traffic models and traffic quality on the roadways. dissertation study, CTU Faculty for Transportation Sciences (in Czech)

[17]. Terman D., Izhikevich E., 2010. State space. http://www.scholarpedia.org/.

[18]. Tsekeris T., Stathopoulos A., 2006. Measuring Variability in Urban Traffic Flow by Use of Principal Component Analysis., Journal of Transportation and Statistics V9, N1.

[19]. Wang Y., Papageorgiou M., 2004. Real-time freeway traffic state estimation based on extended Kalman filter. Transportation research part, Elsevier, pp. 141 – 167.

[20]. Wang Y., Papageorgiou M., Messmer A., Copolla P., Tzimitsi A. Nuzollo A., 2009 An Adaptive freeway traffic state estimator., Automatica 45, Elsevier, pp. 10-24.

[21]. Zhou X., Mahmassani H., 2007. A structural state space model for real-time traffic origin-destination demand estimation and prediction in a day-to-day learning framework. Transportation research part, Elsevier, pp. 823 – 840♦