International Journal of Mechanical, Industrial and Aerospace Sciences
ISSN: 2517-9950
Vol:7, No:10, 2013

# Estimation of Missing or Incomplete Data in Road Performance Measurement Systems

Kristjan Kuhi, Kati K. Kaare, Ott Koppel

***Abstract*—**Modern management in most fields is performance based; both planning and implementation of maintenance and operational activities are driven by appropriately defined performance indicators. Continuous real-time data collection for management is becoming feasible due to technological advancements. Outdated and insufficient input data may result in incorrect decisions. When using deterministic models the uncertainty of the object state is not visible thus applying the deterministic models are more likely to give false diagnosis. Constructing structured probabilistic models of the performance indicators taking into consideration the surrounding indicator environment enables to estimate the trustworthiness of the indicator values. It also assists to fill gaps in data to improve the quality of the performance analysis and management decisions. In this paper authors discuss the application of probabilistic graphical models in the road performance measurement and propose a high-level conceptual model that enables analyzing and predicting more precisely future pavement deterioration based on road utilization.

***Keywords*—**Probabilistic graphical models, performance indicators, road performance management, data collection.

## I. INTRODUCTION

PERFORMANCEis a term used in everyday life, in engineering, in economics and in many other areas. It can have a general meaning or a specific meaning. For the latter, and particularly for roads, performance must be a measurable entity. Performance measurement techniques represent a key element of road infrastructure asset management systems or pavement management systems. Data collection for these systems is becoming feasible due to innovative technological advancements. This is essential for assessing the current and future state of specific fields and management efficiency in productivity, cost-effectiveness, environmental protection, preservation of investments and other functions.

Measurement of performance and productivity has gained significant interest among both academics and practitioners. Much progress has been made on establishing performance measurement systems (PMeS-s) which include a portfolio of measures aimed to balance the more traditional, single focus view on profitability. The following definitions are used in this paper [1]:

- performance measurement can be defined as the process of quantifying the nature of operation;
- a performance measure can be defined as a metric used to quantify the nature of operation;
- aPeMS can be defined as the set of metrics used to quantify and qualify the nature of operation.

Performance measurement describes the feedback or information on activities with respect to meeting strategic objectives. They are used to measure and improve the efficiency and the quality of the production processes, and identify opportunities for progressive improvements in process performance. Most traditional measures overlook key non-financial performance indicators [2].

According to literature contemporary PMeS should meet the following criteria: support strategic objectives; have an appropriate balance; have a limited number of performance measures; be easily accessible; consist of performance measures that have comprehensible specifications [3]. Other issues that should be considered selecting performance measures that can be used in evaluation includes forecast ability, clarity, usefulness, ability to diagnose problems, temporal effects and relevance [4].

Performance indicators can be used in particular as target criteria in life-cycle analyses within the context of pavement design and/or systematic road maintenance. Uniform performance indicators (PIs) permit an evaluation of the effects of different design and maintenance strategies, but they can also be a basis for predicting road performance and for improving old and developing new prediction models.

Constructing structured probabilistic models of the performance indicators taking into consideration the surrounding indicator environment enables to estimate the trustworthiness of the indicator values. It also assists to fill gaps in data to improve the quality of the performance analysis and management decisions. In this paper authors discuss the application of probabilistic graphical models in the road performance management and propose a high-level conceptual model that enables analyzing and predicting more precisely future pavement deterioration based on road utilization.

PIs are thus an objective tool used in road construction and maintenance at various administrative levels, from the local to the national. PIs for road pavements could be, however, also used as inputs into pavement management systems (PMS)[5]. They can be used, for calculating maintenance needs and thus provide objective arguments for reinvestment in road pavements.

Without maintenance, the pavement network would deteriorate to the stage where major expenditure would be needed and the residual value of the pavement or network would be very low. It is crucial to have continuous knowledge

Kristjan Kuhi, Kati Kõrbe Kaare, and Ott Koppelare are with the Department of Logistics and Transport, Tallinn University of Technology, 19086, 5 Ehitajate Street, Tallinn, Estonia (e-mail: kristjan.kuhi@ericsson.com, kati.korbe@ttu.ee, ott.koppel@ttu.ee).

about the condition of the road network in order to plan maintenance work in an effective manner. PeMS will enable agencies to assign priorities, and to compare maintenance outcomes based on the input of resources.

Road authorities collect and retain extensive datasets related to their services and the life-cycle of their infrastructures. It is important to note, however, that proper collection, analysis, refinement and presentation of that data is a prerequisite for using them and for proper reporting to a broader audience. As such, development of appropriate PeMS is required for linking transportation and infrastructure data for road management. Estonian road network as an example is used in this paper.

## II. BACKGROUND

PIs are an essential part of modern road asset management. The basic rationale for having measurable PIs is that limited availability of resources makes it necessary to allocate these resources as effectively as possible among competing alternatives; moreover, that considerations of safety, capacity, serviceability and durability are explicitly recognized.

In developing PIs basic rationale should be considered in the comprehensive approach. A balance in use and reporting, efficiency and effectiveness, a tie to transportation values, objectivity in the measurement used and the stakeholders involved in the development of a framework. PIs should be tied to road administration agency’s policy objectives and to implementation targets or minimum acceptable levels of serviceability [6].

There are different types of PIs and different ways to use them. This understanding is crucial in selecting indicators for an administration [7]. Different indicators should be selected based on whether the intention is to improve [8]:

- the internal efficiency of the road administration;
- the quality of the administration’s products and services;
- the overall performance of the road transport system;
- a particular process of a specific engineering task.

In the process of choosing an effective indicator, it is essential to know:

1) the assumptions regarding the indicator and the rationale for measuring it;
2) the precision and accuracy of any measurements;
3) congruence – many indicators are proxies, so it is essential that the indicator changes in line with the actual behavior;
4) whether a static measure (a value at a point in time) or a vector (a value and direction of change) is more appropriate whether a soft or hard measure is required;
5) if the indicator is going to measure results or behavior;
6) what the likely intended and unintended consequences of the measurement system are [9].

PIs are defined for different types of pavement structures and road categories. In a first step several single PIs describing the characteristic of the road pavement condition are assessed. The next step is the grouping of these single PIs or indices into representative combined performance indices (CPIs) as:

- functional PIs (demands made on road pavements by road users);
- structural PIs (structural demands to be met by the road pavement);
- environmental PIs (demands made on road pavements from an environmental perspective).

Finally, based on the combined PIs and key performance indicators (KPIs) a global (general) PI(Fig. 1) is defined for describing the overall condition of the road pavements, which can be used for general optimization procedures [10].

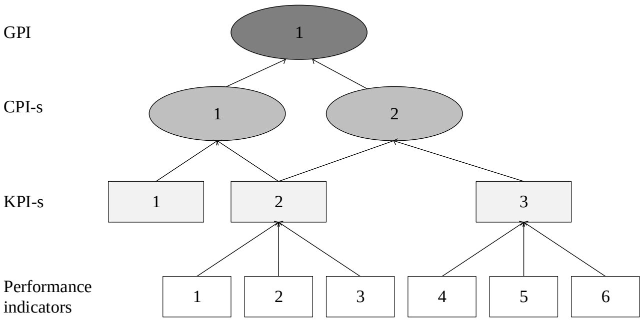

Fig. 1 Conceptual model for road performance indicators [6], [11]

For designing PeMS data should be collected about various structural PIs about the road network and about characteristics having an impact on the deterioration of the road.

## III. DATA COLLECTION TECHNOLOGIES

During the road lifetime, several factors contribute to road deterioration. Among these factors, external influences, such as climate and traffic (both volume and loads), can be measured with various devices. Additionally, road condition can be monitored with several different techniques classified as follows[11], [12]:

1) intrusive, e.g. inductive loops. Systems requiring installation inside the road structure.
2) non-intrusive, e.g. road weather stations. Systems which are placed above road structure.
3) off-roadway, e.g. floating car data. Systems which are not directly connected to road.
4) on-demand, e.g. roughness measurement. This covers analysis performed on-demand basis.

Different techniques are applied to obtain different datasets. For data transfer several wired and wireless technologies are available. Data transfer from field to storage or processing system is instant. For example public mobile wireless technologies offer theoretically up to 5 millisecond latency (one way) of small data packets. Practical results from public non-optimized long-term evolution (LTE) networks show up to 25 millisecond latency rates [13]. Without building special communication access infrastructure majority of decision support applications’ requirements are fulfilled. For real-time control systems the low cost public communication infrastructure needs special optimization, but usually this is not needed.

The amount of data captured and transferred is not low resulting to high speed data streams in some cases. The processing of such amount of data using traditional technologies becomes quickly difficult.

Big Data applies to information that cannot be processed or analyzed using traditional processes or tools [14]. There are techniques available to process and analyze huge amount of unstructured data in parallel for example using Map-reduce paradigm to split up computing jobs in a highly distributed environment.

Users specify a map function that processes a key/value pair to generate a set of intermediate key/value pairs, and a reduce function that merges all intermediate values associated with the same intermediate key. Programs written in this functional style are automatically parallelized and executed on a large cluster of commodity machines. The run-time system takes care of the details of partitioning the input data, scheduling the program's execution across a set of machines, handling machine failures, and managing the required inter-machine communication [15].

The performance index calculation and missing data analytics can be implemented on such paradigm. E.g. Fig. 2proposes an architecture that can be developed that enables analyzing and predicting more precisely future performance. By adding dynamic climate data to the traffic model, prediction of the approximate timeline for probable maintenance needs of pavements will become possible for Estonian conditions. In condition monitoring it is important to gather detailed locations and information about defects, repairs and maintenance history including de-icing frequency and agents used. Several countries including Estonia have outsourced road maintenance from different operators via public tenders. Therefore there needs to be a system supporting unified data collection procedures.

The delivery of a PeMS is affordable, accessible and responsive to the needs of the organization can be achieved only if road sector managers are able to obtain effective feedback provided through a profound approach to data collection and quality. Such feedback provides the information needed for the organization to establish and sustain excellence in program and service delivery to the public.

Proposed PeMS architecture and data collection system is such that information will be provided to consumers dynamically as they request it. Sample usage scenario in Estonia may be a driver starting to drive from Tallinn to Tartu (185km) requesting road condition information from the system – the car will subscribe to the road condition information on the route, receiving it directly and incrementally in real time from the server as the car proceeds on the route [16]. Similarly the police may request information on traffic loads on different roads or locations receiving it from the sensors in the location.

Other potential beneficiaries are:

1) internal and external security forces (police, criminal police, customs, border guard authorities, defense forces);
2) road maintenance and construction authorities;
3) transportation companies;
4) road authorities in developing design and construction guidelines and planning road maintenance[11].

Further on, based on long term performance monitoring results a model can be developed to predict the development of deteriorating effects to pavements and road structures based on climate, traffic and road data.

## IV. INDICES

PeMS is suitable for monitoring the performance of the entire road network in Estonia. The roads managed and supervised by the Estonian Road Administration (ERA) have a better capability to implement the system. The system is designed to collect measurements of the pavement condition, climatic conditions, traffic load and volume and also maintenance records. After modeling, analyzing and evaluating the outcomes of the gathered data it is possible to make suggestions to the design guidelines, construction guidelines and to optimize repair strategies for pavement rehabilitation and maintenance.

Besides valuable information about how to design and construct sustainable roads the system (Fig. 3) enables to develop and successfully implement a proactive approach for prioritizing, preserving, rehabilitating, and maintaining existing pavements.

The process includes technology and implementation planning, software application configuration, database implementation, interfaces to internal applications, testing, training, data conversion, end user and technical documentation, project management, and post-implementation check-up and annual evaluation, development and support.

The primary objectives of developing a PeMS are the following [18]:

1) create the ability to provide a secure method to assure data integrity and access to all vested and interested parties which will lead to improved communication and help optimize the investment in maintenance activities;
2) allow for importing of data from other databases to create a single source for pavement data to improve the network and project decision making process;
3) provide a single source of reports for internal and external stakeholders to improve the decision making process;
4) provide timely and accurate information executive management to improve the daily decision making process;
5) create the ability to prioritize pavement projects based on objective distress and condition indices leading to optimized investment and long-term cost avoidance;
6) monitor pavement segments, sections over time to see the change in the pavement condition to optimize the timing of maintenance decisions leading to long-term cost avoidance;
7) create predictive models for assisting in optimizing the planning and programming of future projects leading to long-term sustainability and cost avoidance;
8) provide the ability to centralize and standardize the tracking of history through integration of key data sources to facilitate optimization of the investment in maintenance and rehabilitation activities;
9) provide pavement performance and history of activities at

any project location within the network to facilitate optimization;

10) provide the ability to determine a level that will support a defined pavement performance level focused on extending the service life by applying preservation treatments early instead of having to rehabilitate the pavement later leading to long-term cost avoidance.

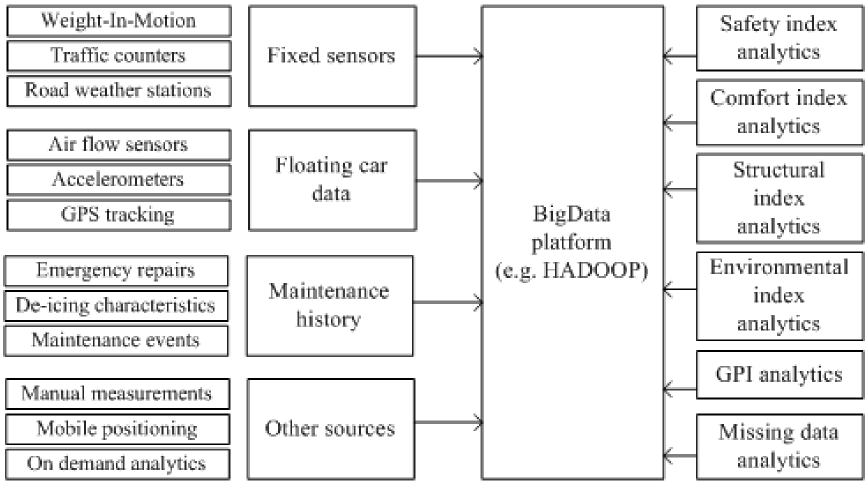

Fig. 2 PeMS configuration for road network [11], [17]

Common way of combining different parameters is defining indicator as weighted sum of different measured values. These weights must be calibrated after according to road network conditions, giving higher weights to more descriptive items.

Technical parameter for cracking describes the severity and spread of pavement cracks. As different crack types are described differently, e.g. longitudinal cracking is measured as length, while alligator cracking is measured as area, formulas (1) are to distinguish and combine collected values based on their spatial properties [5], [11].

$$TP_{cr} = Min\left(100;\ Min\left(100;\ \frac{1}{A_{ref}} \cdot \sum_m \left[W_m \cdot \sum_i \left(S_{cr,a,i} \cdot A_i\right)\right] \cdot 100\right) + Min\left(100;\ \frac{1}{A_{ref}} \cdot \sum_n \left[W_n \cdot I_{width,l} \cdot \sum_j \left(S_{cr,l,j} \cdot L_j\right)\right] \cdot 100\right) + Min\left(100;\ \frac{1}{A_{ref}} \cdot \sum_o \left[W_o \cdot I_{area,k} \cdot \sum_k \left(S_{cr,E,k} \cdot E_k\right)\right] \cdot 100\right)\right) \quad (1)$$

where

$TP_{cr}$ is technical parameter cracking;

$A_{ref}$ is reference-area;

$E_{ref}$ is total number of referred elements (e.g. number of concrete slabs);

$I_{area,k}$is standard area of elements with cracks (e.g. area of concrete slab);

$I_{width,l}$ is standard influence width of linear cracks (e.g. 0,5 m);

$S_{cr,a,I}$is severity of crack type i;

$A_i$ is cracked area of crack type i;

$L_j$ is cracking length of crack type j;

$S_{cr,l,j}$is severity of crack type j;

$E_k$ is number of elements with cracks of type k;

$S_{cr,E,k}$is severity of cracks on an element of crack type k;

$W_m$ is weight of cracked areas;

$W_n$ is weight of cracked length;

$W_o$ is weight of cracked elements.

Alternative 1 considers the mean value of the weighted single PIs other than the maximum weighted single PI influenced by a factor $p$ (2).

$$CPI_i = min\left[5;\ I_1 + \frac{p}{100} \cdot \overline{(I_2, I_3, \ldots, I_n)}\right] \quad (2)$$

where$I_1 \geq I_2 \geq I_3 \geq \ldots \geq I_n$ and $I_1 = W_1 \cdot PI_1$; $I_2 = W_2 \cdot PI_2$;...; $I_n = W_n \cdot PI_n$.

Alternative 2 considers the second largest weighted single PI influenced by a factor $p$. All other PIs which are less than the second largest weighted single PI are not taken into consideration (3).

$$CPI_i = min\left[5;\ I_1 + \frac{p}{100} \cdot I_2\right] \quad (3)$$

where$I_1 \geq I_2 \geq I_3 \geq \ldots \geq I_n$ and $I_1 = W_1 \cdot PI_1$; $I_2 = W_2 \cdot PI_2$;...; $I_n = W_n \cdot PI_n$.

Alternative 1 is the preferred combination procedure for the calculation of Combined Performance Indices (CPIs) because it takes all relevant input values into consideration. However, alternative 2 can be useful for specific applications.

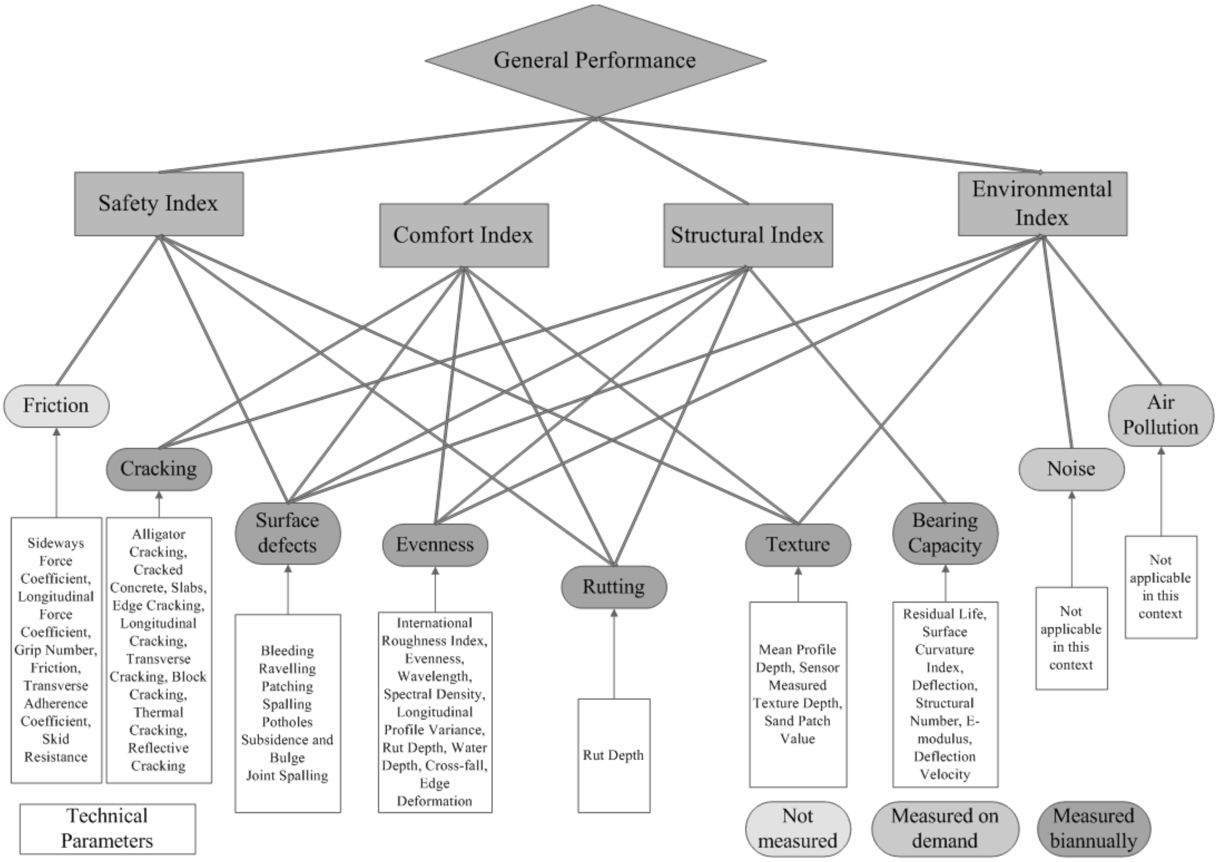

Fig. 3 Performance indices as proposed by COST354 [5], [11], and EVITA [19]

As the sensor network covers mostly main roads, initial indicators should also cover main road network, which carries over 50% of traffic volume on Estonian state roads while constituting only 10% of total length of state roads [20]. With sensor network expanding, those indicators can be extended to whole road network. Possible performance indicators are as follows.

Freeze-thaw degradation: Combining freeze-thaw cycle data from road weather stations with traffic volumes from thawing periods shows possible damage caused by vehicles while pavement is more susceptible to damage than usual. In longer term, these values can be calibrated against actual measured degradation values.

Average speed compared to speed limit. In Estonia, it is common to keep traveling speed close to posted speed limit. If average speed is significantly lower than speed limit, it could indicate problems with road itself.

Weather effect on speed: Combining average speed data with weather conditions gives insight to whether vehicles are travelling slower than usually in similar conditions, indicating problems with road construction, for example drainage.

## V. Models

Missing data can be estimated by multitude of different methods. Interpolation, time series modeling or probabilistic modeling is the most common [21]. In this paper we explore Bayesian Belief Networks as base method for giving meaningful estimates of missing data. The state of the pavement of a road section is often uncertain because the observations about it are partial.

Only some aspects of the causes of the pavement deterioration process are observed or the observations are noisy given the exact time moment. When using deterministic models the uncertainty of the pavement state is not visible thus applying the deterministic models are more likely to give false diagnosis [11].

A Bayesian network is a directed acyclic graph representing the joint probability distribution of all variables in a domain. The topology of the network conveys direct information about the dependency between the variables. In particular, it represents which variables are conditionally independent given another variable [21].

Given the knowledge represented as a Bayesian network, it can be used to reason about the consequences of specific input data, by what is called probabilistic reasoning. This consists of assigning a value to the input variables, and propagating their effect through the network to update the probability of the hypothesis variables. The updating of the certainty measures is consistent with probability theory, based on the application of Bayesian calculus and the dependencies represented in the network. Several algorithms have been proposed for this probability propagation. Bayesian networks can use historical data to acquire knowledge but may additionally assimilate domain experts' input [21]-[22].

Dynamic Bayesian Networks (DBN) is an attempt to add temporal dimension into the BN model. Often a DBN incorporates two models; an initial net B0 learned using information at time 0, and the transition net B1 learned with the rest of the data. Together B0 and B1 confirm the DBN.

Authors suppose a time series dataset. Fig. 4 illustrates the probabilistic model. Variable X represents the variable to be estimated, variables Y and Z represent pieces of Bayesian network corresponding to all the related variables to X. Xt+1 represents the value of variable X at the time t+1, and Xt-1 represents the value of variable X at the time t-1.

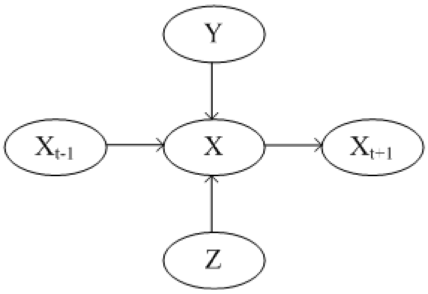

Fig. 4 Dynamic probabilistic model for data estimation [21]

This model represents a dynamic model which provides accurate information for estimating the variable in two senses: firstly, using related information identified by automatic learning algorithms or experts in the domain, or both; secondly, using information of the previous and incoming values. This information includes the change rate of the variable according to the history of the signal.

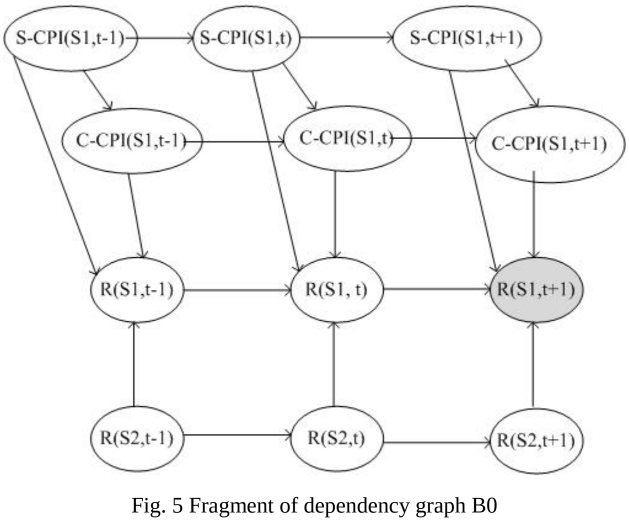

Fig. 5 Fragment of dependency graph B0

Combining Figs. 3 and 4, plus adding road section historical data, we can build an initial B0 dependency graph, which fragment is represented on Fig. 5, of PIs ja CPIs to use for missing data estimation. R represents rutting PI, C-CPI is comfort index, S-CPI is safety index. S1 and S2 are similar road sections. In similar way all the dependencies are captured on a graph to build initial models.

The observations in Estonian Smart Road Datacenter case are done in very different time intervals varying from minutes to years [23]. To find probable state of the pavement of a road section we need to consider also the time passed from the last given observation. This requires the dynamic nature of the probability modeling (e.g. using Hidden Markov Models and DBN).

Some of the currently missing values can be derived from other measurements and similar historical cases. Some can be approximated and still used in the performance calculus giving indication to the end user of the probability of correctness [7].

## VI. CONCLUSIONS

In this paper author's proposed using probabilistic graphical modeling of road KPIs to eliminate gaps in road performance data as well as estimating the trustworthiness of the PI values. Accurate prediction of pavement PIs is important to pavement management agencies.

Reliable and accurate predictions of pavement performance can save significant amounts of public resources through better planning, maintenance, and rehabilitation activities. High-level conceptual model that enables analyzing and predicting more precisely future pavement deterioration based on road utilization was proposed.

## REFERENCES

[1] A. Neely, M. Gregory, and K. Platts, "Performance Measurement System Design. A Literature Review and Research Agenda," *Int. J. Operations & Production Management*, Vol. 15, no. 4, pp. 80-116, 1995.

[2] T. Wegelius-Lehtonen, "Performance measurement in construction logistics," *Int. J. Production Economics*, no. 69, pp. 107-116, 2001.

[3] S. Tangen, "Performance management: from philosophy to practice," *Int. J. Productivity and Performance Management*, Vol. 53, No. 8, pp. 726-737, 2004.

[4] *Performance Measures for Road Networks: A Survey of Canadian Use*. Transportation Association of Canada, 2006. 67 p.

[5] J. Litzka, B. Leben, F. La Torre, A. Weninger-Vycudil, M. de Lurdes Antunes, D. Kokot, G. Mladenovic, S. Brittain, and H. Viner, *The way forward for pavement performance indicators across Europe.* Vienna: Austrian Transportation Research Association, 2008. 68 p.

[6] M. A. Ismail, R. Sadiq, H. R. Soleymani, and S. Tesfamariam, "Developing aroad performance index using a Bayesian belief network model," *J. Franklin Institute*, no. 348, pp. 2539-2555, 2011.

[7] *Performance Indicators for the Road Sector. Summary of the field tests*. Paris: OECD Publications, 2001. 85 p.

[8] Z. J. Radnor, and D. Barnes, "Historical analysis of performance measurement and management in operations management," *Int. J. Productivity and Performance Management*, no. 56(5/6), pp. 384-396, 2007.

[9] R. Haas, G. Felio, Z. Lounis, and L. C. Falls, "Measurable Performance Indicators for Roads: Canadian and International Practice," in *Proc. Annual Conf. Transportation Association of Canada,* Vancouver, 2009.

[10] D. Osborne, and T. Gaebler, *Reinventing Government*. Boston: Addison-Wesley, 1992. 427 p.

[11] K. Kõrbe Kaare, *Performance Measurement of a Road Network: A Conceptual and Technological Approach for Estonia*. Tallinn: TUT Press, 2013. 162 p.

[12] K. K. Kaare, and O. Koppel, "Performance measurement data as an input in transportation policy,“ in *Conf. papers of the XXVIII Int. Baltic Road Conf.*, Vilnius, 2013.

[13] J. Weimer, Y. Xu, C. Fischione, K. H. Johansson, P. Ljungberg, C. Donovan, A. Sutor, and L. E. Fahlén, "A Virtual Laboratory for Micro-

Grid Information and Communication Infrastructures," in $3^{rd}$*IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*, Berlin, 2012.

[14] P. C. Zikopoulos, C. Eaton, D. deRoos, T. Deutsch, and G. Lapis, *Understanding Big Data. Analytics for Enterprise Class Hadoop and Streaming Data*. McGraw-Hill, 2012. 141 p.

[15] J. Dean, and S. Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," *Communications of the ACM*, Vol. 51, no. 1, pp. 107-113, 2008.

[16] K. Kõrbe Kaare, K. Kuhi, and O. Koppel, "Tire and pavement wear interaction monitoring for road performance indicators," *Estonian J. Eng.,* no. 18(4), pp. 324-335, 2012.

[17] K. Kõrbe Kaare, K. Kuhi, and O. Koppel, "Developing Road Performance Measurement System with Evaluation Instrument," *World Academy of Science, Engineering and Technology*, no. 72, pp. 90-96, 2012.

[18] *Pavement management and performance*. California Department of Transportation, 2012.http://www.dot.ca.gov/hq/maint/Pavement/Offices/Pavement_Management/ (accessed 15.07.2013).

[19] M. de Lurdes Antunes (Ed.), *Framework for implementation of Environment Key Performance Indicators. EVITA (Environmental Indicators for the Total Road Infrastructure Assets) Deliverable D4.1*. InstitutFrançais des Sciences et des Technologies desTransports, de l'Aménagement et des Réseaux and PMS-Consult, 2011. 19 p.

[20] *Annual Report 2012*. Tallinn: Estonian Road Administration, 2013. 74 p.

[21] P. H. Ibargüengoytia, U. A. Garcıa, J. Herrera-Vega, P. Hernandez-Leal, E. F. Morales, L. E. Sucar, and F. Orihuela-Espina, "On the Estimation of Missing Data in Incomplete Databases: Autoregressive Bayesian Networks," in *ICONS 2013,* Seville, 2013, pp. 111-116.

[22] D. Smith, G. Timms, P. De Souza, and C.D'Este, "A Bayesian Framework for the Automated Online Assessment of Sensor Data Quality," *Sensors (Basel)*, 12(7), pp. 9476–9501, 2012.

[23] T. Jairus, "Estonian Road Information System Tark tee (Smart road), in *Conf. papers of the XXVIII Int. Baltic Road Conf.*, Vilnius, 2013.