# 

# Stochastic prediction of train delays in real-time using Bayesian networks[☆](https://www.sciencedirect.com/science/article/abs/pii/S0968090X18311021?via%3Dihub" \l "aep-article-footnote-id1)

# F. Corman, P. Kecman

## [Transportation Research Part C: Emerging Technologies](https://www.sciencedirect.com/science/journal/0968090X)

[Volume 95](https://www.sciencedirect.com/science/journal/0968090X/95/supp/C), October 2018, Pages 599-615

DOI:[10.1016/J.TRC.2018.08.003](https://doi.org/10.1016/J.TRC.2018.08.003)

## Abstract

In this paper we present a stochastic model for predicting the propagation of train delays based on Bayesian networks. This method can efficiently represent and compute the complex stochastic inference between random variables.

В этой статье мы представляем стохастическую модель для прогнозирования распространения задержек поездов на основе байесовских сетей. Этот метод может эффективно представлять и вычислять сложное стохастическое соотношение (правило) между случайными величинами.

Moreover, it allows updating the probability distributions and reducing the uncertainty of future train delays in real time under the assumption that more information continuously becomes available from the monitoring system. The dynamics of a train delay over time and space is presented as a stochastic process that describes the evolution of the time-dependent random variable. This approach is further extended by modelling the interdependence between trains that share the same infrastructure or have a scheduled passenger transfer. The model is applied on a set of historical traffic realization data from the part of a busy corridor in Sweden. We present the results and analyze the accuracy of predictions as well as the evolution of probability distributions of event delays over time. The presented method is important for making better predictions for train traffic, that are not only based on static, offline collected data, but are able to positively include the dynamic characteristics of the continuously changing delays.

Большинство существующих методов предполагают фиксированными распределения вероятностей для задержек поездов и не учитывают влияние, которое информация в реальном времени о позициях и задержках поездов может оказывать на соответствующие распределения. Чтобы создать реалистичные онлайн-инструменты для управления трафиком в реальном времени, необходимо учитывать динамику неопределенности задержек. Когда становится доступной новая информация о позициях поездов и задержках, неопределенность прогнозирования последующих событий обычно уменьшается. Основная цель данной статьи - изучить влияние, которое горизонт прогнозирования и поступающая информация о движущемся поезде может иметь на предсказуемость последующего времени прибытия и отправления всех поездов. Другими словами, мы пытаемся ответить на вопрос: как меняется распределение вероятностей задержки события во времени?

В этой статье мы сначала описываем метод моделирования рассеяния задержек поездов на основе байесовских сетей. Железнодорожное движение моделируется с помощью вероят-ностной графовой модели, которая использует условную независимость между события-ми, чтобы обеспечить эффективное вычисление их совместного распределения (Коллер и Фридман, 2009). Важным преимуществом этого метода в контексте прогнозирования движения поездов в реальном времени является то, что он позволяет распространяться информации или данным (evidence, признаки) об определенном событии. Другими словами, данные о реализации одного события снижают неопределенность других событий. Следовательно, распределение вероятностей, например, задержки прибытия на станцию изменяется с течением времени дискретными шагами по мере поступления дополнительной информации.

Это может быть использовано диспетчером для оценки вероятности маршрутного конфликта на его участке, вероятности задержки прибытия пересадочного (feeder, питающего) поезда для перевозки пассажиров и т. д. Более того, более точная оценка задержек поездов может быть очень полезной для проверки и оценки современных моделей движения по сети. В частности, этот метод позволяет оценить динамику задержки для перепланирования замкнутого кольца (Corman, Quaglietta, 2015; Caimi et al, 2012), перепланирования онлайн (Gatto et al, 2007; Bauer and Schobel, 2014) и инструментов моделирования (Nash and Huerlimann, 2004; Quaglietta, 2014). Наконец, несмотря на то, что мы сосредотачиваемся на прогнозировании железнодорожного движения в реальном времени, структура и методология моделирования, представленные в этой работе, могут быть расширены для обработки прогнозов в других плановых и ограниченных системах, таких как общественный транспорт, логистические сети и цепочки поставок.

Railway prediction models can be classiﬁed to static (ofﬂine) and dynamic (online), and

deterministic and stochastic, depending on the time span between the time they are run, and

the operations they aim to predict; and on how they tackle uncertainty, respectively.

Модели прогнозирования для железных дорог можно разделить на статические (офлайн) и динамические (онлайн), а также детерминированные и стохастические, в зависимости от промежутка времени между временем их выполнения и операциями, которые они стремятся прогнозировать; и по способу, которому они борются с неопределенностью, соответственно.

Deterministic prediction models assume full knowledge of the future trafﬁc evolution (Dolder

et al, 2009). Some approaches (see for instance (Burdett and Kozan, 2014; Wei et al, 2015)

focus on simulating the trafﬁc based on current state, and determine most likely conﬂicts,

with limited usage of data on past operations.

Even though the more advanced data-driven deterministic models are able to explain a large percentage of process time variability using the values of explanatory variables, a certain degree of uncertainty, especially for dwell times, still remains unresolved (Kecman and Goverde, 2015).

Несмотря на то, что наиболее продвинутые детерминированные модели, основанные на данных, могут объяснить большой процент изменчивости времени процесса путем использования значений независимых переменных, определенная степень неопределен-ности, особенно в отношении времени стоянки, все еще остается неразрешенной (Kecman and Goverde, 2015).

Whereas static prediction models are based on the ofﬂine computed probability distributions and their parameters, dynamic models are updated in real-time as new information becomes available. Most of the stochastic delay propagation models (Buker and Seybold, 2012; Medeossi et al, 2011; Meester and Muns, 2007) were used for ofﬂine analyses of timetables.

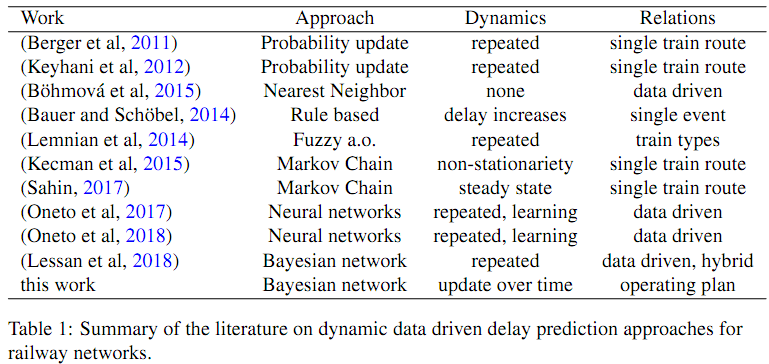
We summarize in Table1 the most related works, which are discussed in what follows. In column 2, the works are categorized according to their approach, i.e. basic algorithm used to determine the conditional probability of future events based on past data and current events.

А learning process can update incrementally the model with some new values of parameters,

as suggested in (Oneto et al, 2017, 2018). Процесс обучения может постепенно обновлять модель с некоторыми новыми значениями параметров, как предложено в (Oneto et al, 2017, 2018).

The last column classiﬁes whether the probabilistic relations between events are inferred from data only, or from some domain knowledge, either at the single event, single route of a train, aggregating together trains of the same type, or considering the entire network operating plan and interactions between trains over the network.

В последнем столбце указывается, выводятся ли вероятностные связи между событиями только на основании данных или на основе некоторого знания предметной области, либо на одном событии, на одном маршруте поезда, либо при объединении поездов одного типа, либо с учетом плана работы всей сети и взаимодействия между поездами на сети.



Regarding the determination of uncertainty and dynamics of delays over time, the usage of past data has been shown in (Sahin, 2017) enough to determine a Markov model of the operations as stochastic processes. Ultimately, running time supplements that would absorb the evolution of the delays in a steady state, as measured in the recorded data, can be computed. In a preliminary work on this topic, the dynamics of uncertainty was included by modelling train delay evolution over time as a stochastic process (Kecman et al, 2015). A train run is represented as a Markov chain with state transitions in discrete moments that represent arrival and departure events from a scheduled stop. After every registered departure or arrival event, the conditional probability distributions of the downstream events are updated with respect to the essential assumption for Markov processes that, given the present, future events do not depend on the past. A train delay evolution is modelled as a non-stationary Markov chain, meaning that the probability of a state change depends on the moment of transition.

Что касается учета неопределенности и динамики задержек во времени, было показано, что использования прошлых данных достаточно, чтобы определить марковскую модель операций как совокупность случайных процессов (Sahin, 2017). В конечном счете, можно вычислить добавки к времени хода, которые поглотили бы изменение задержек в установившемся состоянии, измеренное в записанных данных.

В предыдущей работе по этой теме динамика неопределенности была включена путем моделирования эволюции задержки поезда во времени как случайного процесса (Kecman et al, 2015). Поездка поезда представлена как цепь Маркова с переходами между состояниями в дискретные моменты, которые представляют события прибытия и отправления с плановой остановки. После каждого зарегистрированного события отправления или прибытия условные распределения вероятностей последующих событий обновляются в соответствии с основным предположением для марковских процессов, что при данном настоящем, будущие события не зависят от прошлого. Эволюция задержки поезда моделируется как нестационарная цепь Маркова, что означает, что вероятность изменения состояния зависит от момента перехода.

In this paper we develop the complementary idea to explicitly include the causal and temporal dependencies of events of other trains in the computation by explicitly modelling the (domain-speciﬁc) interdependence between trains that share the same infrastructure, or have a scheduled passenger transfer, by means of Bayesian networks. Therefore, an observed delay of a train will not only be used to update the probabilities of further events along the route of that train. That is, probability distributions of delay for all events of other trains that may be affected are updated.

В этой статье мы развиваем дополнительную идею явного включения причинно-следственных и временных зависимостей событий, происходящих с другими поездами, в вычисления путем точного моделирования (зависящего от места) взаимозависимости между поездами, которые используют одну и ту же инфраструктуру или имеют плановый пассажирский трансфер, средствами байесовских сетей. Следовательно, наблюдаемая задержка поезда будет использоваться не только для обновления вероятностей дальнейших событий по маршруту этого поезда. То есть, обновляются распределения вероятностей задержки для всех событий других поездов, которые могут быть затронуты.

An illustrative example of the system setup is given in Figure 2. The departure of the ﬁrst train from Station A and its arrival to Station B initiate the procedure to update the probability distri-butions of all other estimated event times (EET) that may be affected by the observed delays. A Bayesian network with a structure that corresponds to a macroscopic trafﬁc model can therefore be used to compute stochastic delay propagation with respect to the capacity constraints as well as the constraints due to passenger, rolling-stock or crew connections (Goverde, 2010).

Иллюстративный пример построения системы приведен на рисунке 2. Отправление первого поезда со станции A и его прибытие на станцию B инициируют процедуру обнов-ления распределений вероятностей всех других расчетных времен события (EET), на кото-рые могут повлиять наблюдаемые задержки. Таким образом, байесовская сеть со струк-турой, которая соответствует макроскопической модели движения, может использоваться для вычисления распространения стохастической задержки с учетом ограничений пропу-скной способности, а также ограничений, связанных с пассажирами, подвижным составом или стыковками бригад (Goverde, 2010).

While the structure of the Bayseian network is ﬁxed by the operating plan (e.g. a timetable), we use historical trafﬁc data to calibrate the resulting Bayesian network with conditional probability distributions and regression coefﬁcients for every two dependent events. Therefore, the inco-ming information from the monitoring system is used to reduce the uncertainty of the future events.

В то время, как структура байесовской сети фиксируется рабочим планом (например, расписанием), мы используем исторические данные о движении для калибровки результирующей байесовской сети условными распределениями вероятностей и коэффициентами регрессии для каждых двух зависимых событий. Следовательно, приходящая информация из системы мониторинга используется для уменьшения неопределенности будущих событий.

Similar approaches based on Bayesian networks have been also used in a recently published paper (Lessan et al, 2018). There, the focus was in studying algorithmic performances of different network training methods (hill climbing, primitive linear), as well as the possibility of using hybrid learning methods to combine pure data-driven approaches with some domain knowledge. The authors found that hybrid structures can satisfactorily model complex relations arising in operations. Their main interest was in proving low prediction error of the models for future events. In comparison, we also base our study on a mix of domain knowledge, exploiting in particular the acyclity arising in graph theoretical models for railway operations, combined with large volumes of data.

Подобный подход, основанный на байесовских сетях, также использовался в недавно опубликованной статье (Lessan et al, 2018). Там основное внимание уделялось изучению алгоритмических характеристик различных методов обучения сети (восхождение в гору, примитивно-линейное обучение), а также возможности использования гибридных методов обучения для объединения подходов, основанных исключительно на данных, с некоторыми знаниями предметной области. Авторы обнаружили, что гибридные структуры могут удовлетворительно моделировать сложные отношения, возникающие в работе. Их основной интерес заключался в доказательстве малой ошибки предсказания будущих событий. Для сравнения, мы также основываем наше исследование на сочетании знаний предметной области, используя, в частности, ацикличность, возникающую в теоретико-графовых моделях для железнодор. операций, и больших объемов данных.

Our biggest contribution and theoretical focus is not per se in prediction accuracy, but in the interesting study of delay dynamics, i.e. how the prediction of a single delay event changes over time as the same event approaches the time *now*.

Наш самый важный вклад и теоретическое направление - не в точности предсказания как таковой, а в интересном исследовании динамики задержки, то есть: как прогноз единичного события задержки изменяется во времени по мере приближения этого события к настоящему моменту.

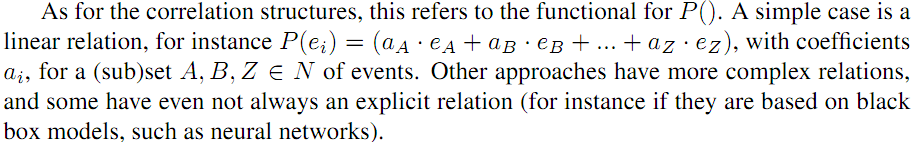
3 Methodological framework 3 Методологический подход

The ultimate goal of this paper is to compute the estimate of (i.e., predict) the probability distribution of the random variables describing an event in the future. To do so, we have to combine the realised value of some random variables, which are somehow connected to the event, with past observations of the random variables which allow us to describe this relation. In general, i.e. event might depend on all events, and maybe also on other exogenous ones. For prediction purposes, only those events in the past are actually usable. In this general framework of stochastic prediction of railway operations, the most important choices are which structure to give to the connections, the form of correlation structures; and the procedure of update of the conditional probabilities as new information is available.

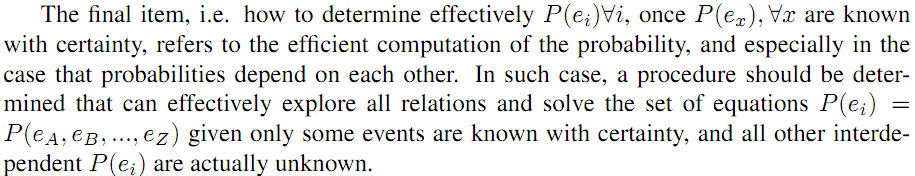
Конечная цель этой статьи - вычислить оценку (т. е., предсказать) распределения вероятностей случайных величин, описывающих событие в будущем. Для этого мы должны объединить фактически реализованные значения некоторых случайных величин, которые так или иначе связаны с событием, с прошлыми наблюдениями случайных величин, которые позволяют нам описать эту связь. В общем,  то есть, событие может зависеть от всех N событий, а может быть, и от других экзогенных. В целях прогнозирования фактически можно использовать только те события 0, 1…, которые произошли в прошлом. В этом общем методе стохастического прогнозирования железнодорожных операций наиболее важным выбором являются вопросы: какую структуру придать стыковкам, форме корреляционных структур; - и процедура обновления условных вероятностей тогда, когда поступает новая информация.

As for the structure, in principle all events might be dependent on all other events, but in practice, models would assume that only some of the connections are more relevant than others. In such case, only some connections are explicitly modelled, the others can be kept only implictly, or completely neglected in the mathematical relations.

Что касается структуры, в принципе все события могут зависеть от всех других событий, но на практике модели предполагают, что только некоторые из связей более актуальны, чем другие. В таком случае явно моделируются только некоторые связи, другие могут сохраняться неявно или полностью игнорироваться в математических соотношениях. Тогда Р(еi) может быть представлено как функция Р (еА, еВ…), то есть только для подмножества событий .



Другие методы имеют более сложные отношения, а некоторые даже не всегда имеют явную связь (например, если они основаны на моделях черного ящика, таких как нейронные сети).



Последнее: как эффективно определять Р(еi), когда Р(ех) известны определенно, относится к эффективному вычислению вероятности, особенно в случае, когда вероятности зависят друг от друга. В таком случае должна быть определена процедура, которая может эффективно исследовать все отношения и решать систему уравнений Р(еi)= Р (еА, еВ…), при данных только некоторых событиях, которые известны определенно, а все другие взаимозависимые Р(еi) фактически не известны.

Bayesian networks are graphical models for reasoning under uncertainty, where the variables and conditional dependencies between them are represented with a directed acyclic graph G (N,A) (Korb and Nicholson, 2010). Nodes i, j пр N represent random variables. In our model, each *i* random variable is associated to a node, and models directly an event еi , which can be a particular train arriving at a particular station. Each event еi and associated node *i* have associated some attributes: train number, station name and event type (arrival, departure or through). A directed arc (i, j) connects two nodes i and j, and models the dependencies between the events, with the direction of an arc indicating the causality relationship between the variables.

Байесовские сети - это графические модели для доказательства при наличии неопределен-ности, где переменные и условные зависимости между ними представлены в виде ориен-тированного ациклического графа G (N, A) (Korb and Nicholson, 2010). Узлы i, j пр N представляют собой случайные величины. В нашей модели каждая случайная величина i связана с узлом и непосредственно моделирует событие еi, которым может быть конкрет-ный поезд, прибывающий на определенную станцию. С каждым событием еi и соответ-ствующим узлом i связаны некоторые атрибуты: номер поезда, название станции и тип события (прибытие, отправление или проследование). Направленная дуга (i, j) соединяет два узла i и j и моделирует зависимости между событиями, при этом направление дуги указывает причинно-следственную связь между переменными.

Bayesian networks rely on the fact that a random variable typically interacts directly with but a few other random variables to construct a concise representation of reality where only the direct dependencies are encoded in the network (Koller and Friedman, 2009). The structure of the network, i.e., the directed arcs between the nodes that represent the considered events, can either be learned from the data or determined by expert knowledge. The recent trend of implementing sensor technologies and advanced data management systems in many railway networks in Europe allows using the massive databases of historical trafﬁc data for the structure and parameter learning of Bayesian networks.

Байесовские сети основаны на том факте, что случайная величина обычно напрямую взаимодействует с несколькими другими случайными величинами для создания компактного представления реальности, когда в сети закодированы только прямые зависимости (Koller and Friedman, 2009). Структура сети, то есть направленные дуги между узлами, которые представляют рассматриваемые события, может быть получена из данных или определена экспертными знаниями. Современная тенденция внедрения сенсорных технологий и передовых систем управления данными на многих железнодорожных сетях в Европе позволяет использовать огромные базы исторических данных о движении для изучения структуры и параметров байесовских сетей.

In our approach, we resort to a Bayesian network as we can exploit a large body of knowledge on graph theoretical models of railway operations (Hansen and Pachl, 2014). Namely, we know that feasible railway operations plans can be represented as a Direct Acyclic Graph, and use this to efﬁciently determine a network structure, which has relatively few explicit connections (i.e. the degree of connectivity of the network is relatively low), while reaching satisfactory prediction performance. We consider all connections between events as linear. An important property of Bayesian networks is that they explicitly model the quantitative strength of the connections between variables, thus allowing the probabilistic beliefs about them to be updated automatically as new information becomes available. This property enables modelling the dynamic inference between random variables in discrete moments in time (Murphy, 2002).

В нашем методе мы прибегаем к байесовской сети, поскольку мы можем использовать большой объем знаний по теоретическим графовым моделям железнодорожных операций (Hansen and Pachl, 2014). А именно, мы знаем, что возможные планы железнодорожных операций могут быть представлены в виде направленного ациклического графа, и исполь-зуем его для эффективного определения сетевой структуры, которая имеет относительно мало явных соединений (т. е. степень связности сети относительно низка), благодаря чему достигается удовлетворительное прогнозирование. Мы считаем все связи между событи-ями линейными. Важным свойством байесовских сетей является то, что они явно модели-руют количественную силу связей между переменными, что позволяет автоматически обновлять вероятностные представления о них по мере появления новой информации. Это свойство позволяет моделировать динамическую связь между случайными величинами в дискретные моменты времени (Murphy, 2002).

Learning the dependencies from a pure data driven approach (see (Oneto et al, 2017, 2018) for instance) or an hybrid approach (see (Lessan et al, 2018)) would in principle be able to represent all observed correlations and stochastic dependencies between events. In this case Р(еi)= Р (еА, еВ…), where only those events A, B, Z are considered, which have a link (causal, empirical or temporal correlation) with ei which is considered strong enough. The limitations of such an approach are in the difﬁculty to determine causality or a temporal relationship within correlated events, in this way. Moreover, in theory there can be sensible correlations between all possible Nodes (i.e. order of probabilistic relations), and moreover with external exogenous variables (for instance, weather). Thus a lot of effort or even ad-hoc ﬁltering (Lemnian et al, 2014) might help keeping only those relations which are physically and causally meaningful, and remain quick enough when building topology and updating the conditional probabilities.

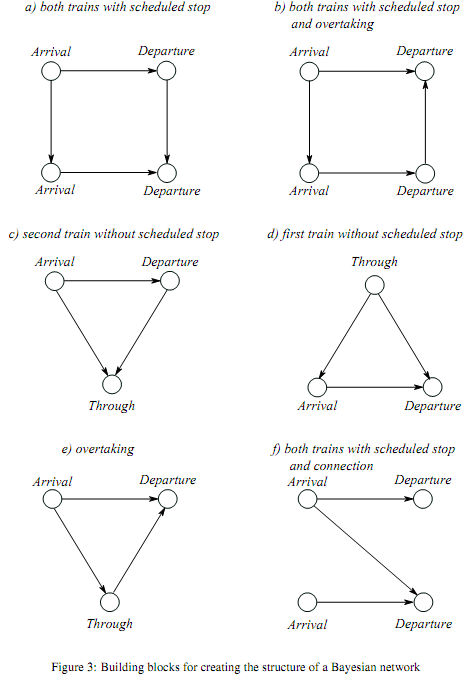
Изучение зависимостей с помощью метода, основанного чисто на данных (см, например, (Oneto et al, 2017, 2018)) или гибридного метода (см. (Lessan et al, 2018)), в принципе, сможет представить все наблюдаемые корреляции и стохастические зависимости между событиями. В этом случае Р(еi)= Р (еА, еВ…), где только те события рассматриваются, связь (причинная, эмпирическая или временная корреляция) с которыми считается доста-точно сильной. Ограниченность такого подхода состоит в том, что таким способом трудно определить причинно-следственную связь или временную взаимосвязь коррелированных событий. Более того, теоретически могут существовать разумные корреляции между всеми возможными узлами (то есть, порядок вероятностных отношений) и, более того, с внешними экзогенными переменными (например, погодой). Таким образом, только большие усилия или даже специальная фильтрация (Lemnian et al, 2014) могут помочь сохранить те связи, которые имеют физическое и причинное значение; и остаются достаточно быстрыми при построении топологии и обновлении условных вероятностей.

We instead resort to domain-dependent knowledge, namely we use as structure of our probabi-listic connections the plan of operation, which would the implicitly model all possible relations between events. In fact, a delay of a train may be a direct predictor of delay of the next event of that train which in turn can be used to estimate the delay its immediate successor and so on. The same principle can be applied to trains that use the same infrastructure and therefore need to be separated by minimum headway times. If two trains use the same part of the infrastructure (block section or a station track) within short time, which is often the case on busy corridors, a delay of the ﬁrst event can be used to predict the delay of the second. Finally, for trains with a scheduled passenger connection in a station, an arrival delay of a feeder train could cause a departure delay of the connecting train. We remark here that those domain-dependent constraints are not consi-dered in the approaches mentioned in Table 1 and in fact are needed to ensure acyclicty of the graph. Experimentally, we ﬁnd that the number of arcs considered in those approaches is relati-vely small, in the same order of magnitude as the amount of nodes.

Вместо этого мы прибегаем к знаниям, зависящим от места (зоны), а именно, мы исполь-зуем в качестве структуры наших вероятностных связей план действий, который неявно моделирует все возможные связи между событиями. Фактически, задержка поезда может быть прямым предсказателем задержки следующего события этого поезда, которое, в свою очередь, может использоваться для оценки задержки его непосредственного после-дователя и так далее. Тот же принцип может быть применен к поездам, которые исполь-зуют одну и ту же инфраструктуру и поэтому должны быть разделены минимальным интервалом времени. Если два поезда используют одну и ту же часть инфраструктуры (блок-участок или станционный путь) в течение короткого времени, что часто имеет место в загруженных коридорах, задержка первого события может использоваться для прогнози-рования задержки второго. Наконец, для поездов регулярного пассажирского сообщения задержка на станции прибытия первого (питающего) поезда может вызвать задержку отправления стыковочного поезда. Здесь отметим, что эти ограничения, зависящие от места, не учитываются в методах, упомянутых в таблице 1, и фактически необходимы для обеспечения ацикличности графа. Экспериментально мы обнаружили, что количество дуг, рассматриваемых в этих методах, относительно невелико, - того же порядка, что и количества узлов.

All those structured systematic relations between events can be schematized into some template structures, reﬂecting the constraint from railway operations, to the relationship between events. The basic structures that are used to model causal dependencies in most typical situations for one-directional trafﬁc are given in Figure 3.

Все эти структурированные систематические отношения между событиями могут быть схематизированы в некоторые шаблонные структуры, отражающие ограничения от желез-нодорожных операций, для связи между событиями. Основные структуры, которые испо-льзуются для моделирования причинно-следственных зависимостей в наиболее типичных ситуациях для однонаправленного движения, представлены на рисунке 3.



The described principles of causal relationship between delays of different events correspond to deterministic macroscopic trafﬁc models where only train events at stations (departures, arrivals, through runs) are represented as nodes in the graph. This is similar to event activity networks where events еi are of two types, arrival events, or departure events, and activities are describing a particular relation between two events. This model can be easily adapted to a graph, where events are associated to nodes i j, and activities to arcs.

Описанные принципы причинно-следственной связи между задержками различных собы-тий соответствуют детерминированным макроскопическим моделям движения, в которых только события на станциях (отправления, прибытия, проследования) представляются в виде узлов в графе. Это похоже на сети событийной деятельности, где события еi бывают двух типов: события прибытия или события отправления, а действия описывают конкретную связь между двумя событиями. Эта модель может быть легко адаптирована к графу, где события связаны с узлами i j, а действия - с дугами.

In practice each of those constraints can be represented as   where  would represent the time at which event occurs, d(i,j) represent the duration between events i and j , i.e. represented by the arc (i, j). The operators describes the fact that events can only occur when a set of events and durations have respectively occurred and elapsed. In the different cases of running, dwell, headways, connections, planned departures, a suitable choice of the d(i,j) is possible.

На практике каждое из этих ограничений может быть представлено как (), где () представляет момент, в который происходит событие, d (i, j) представляет собой интервал между событиями i и j, то есть представлено дугой (i, j). Операторы описывают тот факт, что события могут происходить только тогда, когда реализовался набор событий и время истекло. В различных случаях хода, остановки, межпоездных интервалов, стыковок, плановых отправлений возможен подходящий выбор d (i, j).

The macroscopic character of the model implies that the detailed infrastructure constraints are neglected. However, they are implicitly represented in precedence constraints between station events. For example, two conﬂicting inbound train routes in a station are represented, in a deterministic macroscopic model, by an arc that indicates that the second train cannot arrive before the ﬁrst train has arrived. In a probabilistic graph used in this paper, the corresponding arc models the probability of delay of arrival event of the second train given the arrival delay of the ﬁrst train (Figure 3, situations (a) and (b)).

Макроскопический характер модели подразумевает, что не учитываются детальные ограничения инфраструктуры. Однако они неявно представлены в ограничениях предшествования между станционными событиями. Например, два конфликтующих маршрута прибывающих поездов на станции представлены в детерминированной макроскопической модели дугой, которая указывает на то, что второй поезд не может прибыть раньше, чем прибыл первый поезд. В вероятностном графе, используемом в этой статье, соответствующая дуга моделирует вероятность задержки прибытия второго поезда с учетом задержки прибытия первого поезда (рисунок 3, ситуации (a) и (b)).

In this paper we assume that the train orders are known. In general, the default plan of operations is the timetable, but every new plan of operations which can be implemented as a reaction to delays (changing order of trains, for instance) should have no two operations depending mutually on each other, to be feasible.

В этой статье мы предполагаем, что порядок поездов известен. В общем, план операций по умолчанию - это расписание, но каждый новый план операций, который может быть реализован как реакция на задержки (например, изменение порядка движения поездов), чтобы быть выполнимым, не должен содержать двух операций, взаимно зависящих друг от друга.

3.2 Parameters learning and inference Изучение и расчет параметров

Once the structure of the network has been determined, probability distributions of delays of

considered events need to be computed. This would mean, computing the joint probabilities Р(еi)= Р (еА, еВ…) of all events that are considered in the structure of the delay relations; i.e. that are associated to an arc in an equivalent DAG representing the operating plan of railway movements. The cumbersome computation of a joint distribution for a large number of random variables is avoided by relying on the network structure that encodes only the direct dependencies between events, according to the Bayesian networks. In this sense, the structures reported in Figure 3, are instrumental in identifying templates of linear relations between events, which are described as joint probability distributions. Joint distribution can thus be represented by a number of local distribution which have only a small number parameters to estimate (Nagarajan et al, 2013), given a network structure, encoding the sensible relations between events.

После определения структуры сети вероятностные распределения задержек необходимо вычислить. Это означает вычисление совместных вероятностей Р (еi) = Р (еА, еВ…) всех событий, которые учитываются в структуре отношений задержки; то есть, это связано с дугой в эквивалентном DAG-представлении (графовом представлении) рабочего плана движения железной дороги. Громоздкое вычисление совместного распределения для большого количества случайных величин можно избежать, полагаясь на структуру сети, которая кодирует только прямые зависимости между событиями, согласно байесовским сетям. В этом смысле структуры, представленные на рисунке 3, служат инструментом для определения шаблонов линейных отношений между событиями, которые описываются как совместные распределения вероятностей. Таким образом, совместное распределение может быть представлено рядом локальных распределений, которые имеют лишь небольшое число параметров для оценки (Nagarajan et al, 2013), учитывая сетевую структуру, кодирующую разумные отношения между событиями.

Koller and Friedman (2009) show that a local distribution can in fact be represented as a linear model in which the parent nodes of the incoming arcs are the explanatory variables. This is in accordance with the earlier research results that demonstrate the accuracy of linear regression for modelling train delay dependencies and justify the selection of normal distribution to model train delays in the Bayesian network model (Bayissa, 2013; Kecman and Goverde, 2015). Distribution parameters for delay of node associated to event are computed with respect to the network structure.

Коллер и Фридман (2009) показывают, что локальное распределение фактически может быть представлено в виде линейной модели, в которой родительские узлы входящих дуг являются независимыми переменными. Это соответствует результатам предыдущих исследований, которые демонстрируют точность линейной регрессии для моделирования зависимостей задержек поездов и оправдывают выбор нормального распределения для моделирования задержек поездов в модели байесовской сети (Bayissa, 2013; Kecman and Goverde, 2015). Параметры распределения для задержки в узле, связанной с событием, вычисляются в соответствии с сетевой структурой.

In particular, referring to the probability  representing the probability that an event would have value (i.e., occur at delay) minutes, we compute the expectation  of the event  as where  represents the set of direct predecessors (parents) of the node associated to event . The coefﬁcients  are obtained by a process of ﬁtting the linear model, while  are observed or estimated delays of parent events. The standard deviation  is computed as a standard deviation of the residuals from the linear model.

В частности, обращаясь к вероятности Р, представляющей вероятность того, что событие будет иметь значение (т. е. происходит с задержкой) в минутах, мы вычисляем математ. ожидание  события как где  представляет собой совокупность прямых предшественников (родителей) узла, связанного с событием . Коэффициенты  получаются путем адаптации линейной модели, в то время как  являются наблюдаемыми или оцениваемыми задержками родительских событий. Стандартное отклонение  рассчитывается как стандартное отклонение невязок (разностей) линейной модели.

A local distribution (of an event) is considered to be a univariate normal random variable. This assumption of the Gaussian property of random variables may be seen as a limitation for application in modelling train delays, having in mind the earlier results of statistical analyses of train delays, which typically use Weibull or Gamma distribution to model the delays (Yuan and Hansen, 2007). However, a local distribution in our model describes the delay of an event conditioned on the known delay of its direct predecessors.

Локальное распределение (события) считается одномерной нормальной случайной величиной. Это предположение о гауссовском свойстве случайных величин можно рассматривать как ограничение для применения при моделировании задержек поездов, имея в виду более ранние результаты статистического анализа задержек поездов, которые обычно используют распределение Вейбулла или гамма-распределение для моделирования задержек (Юань и Хансен , 2007). Однако локальное распределение в нашей модели описывает задержку события, обусловленную известной задержкой его прямых предшественников.

The additional information about the realisation time of the previous event can be used to

signiﬁcantly reduce uncertainty and modify the probability distribution of the event time.

For example an arrival delay can be predicted more easily if the departure delay from the

previous station (and/or arrival delay of the previous train to the same station) is known. The entrance delays (delays that do not have any predecessors) can in turn be estimated using

arbitrary ﬁtted distribution (not necessarily normal Gaussian) and given as crisp values as

an input to the model.

Дополнительная информация о времени реализации предыдущего события может использоваться для значительного уменьшения неопределенности и изменения распределения вероятностей времени события. Например, задержку прибытия можно легче предсказать, если известна задержка отправления с предыдущей станции (и / или задержка прибытия предыдущего поезда на ту же станцию). Входные задержки (задержки, не имеющие предшественников), в свою очередь, могут быть оценены с использованием произвольно подобранного распределения (не обязательно гауссова) и представлены в виде четких значений, в качестве входных данных для модели.

Rather than simply describing the dependencies between random variables, Bayesian networks represent a powerful tool for inferential statistics and probabilistic reasoning. When new information about a random variable (evidence) becomes available, it is propagated through the network by updating the posterior probabilities (beliefs) of the relevant nodes. A number of exact (recursive application of Bayes’ theorem) and approximate (Monte Carlo sampling) algorithms exists that can perform this computationally demanding task (Korb and Nicholson, 2010).

Вместо простого описания зависимостей между случайными величинами, байесовские сети представляют собой мощный инструмент для статистических выводов и вероятностных рассмотрений. Когда новая информация о случайной переменной (данные) становится доступной, она распространяется по сети путем обновления апостериорных вероятностей (представлений) соответствующих узлов. Существует ряд точных (рекурсивное применение теоремы Байеса) и приближенных (выборка Монте-Карло) алгоритмов, которые могут выполнять эту ресурсоемкую расчетную задачу (Korb and Nicholson, 2010).

When an evidence about the observed event delay, becomes available, a conditional probability query determines the posterior distribution for each reachable event.

Когда становятся доступными данные о наблюдаемой задержке события, условная вероятность определяет апостериорное распределение для каждого достижимого события.

For every two nodes  and respective events ei and ej, where j is reachable from i, i.e. there is a sequence of directed arcs allowing a path from i to j , a conditional probability query may be used to answer the questions such as: what is the probability that a delay of event (i.e. the realization of node), is larger than minutes, given (assuming) that the delay of event ei (i.e. the realization of i) is y minutes? Moreover, given that the network is calibrated with continuous data and parametrised with linear coefﬁcients, repeated application of linear models gives the most probable outcome of delays for all reachable subsequent events (maximum a posteriori query).

Для каждых двух узлов  и соответствующих событий ei и ej, где j достижимо из i, т. е. существует последовательность направленных дуг, позволяющих пройти от i до j, запрос условной вероятности может использоваться для ответа на такие вопросы, как: какова вероятность того, что задержка события ei (то есть реализация i) есть *y* минут? Более того, учитывая, что сеть откалибрована непрерывными данными и параметризована линейными коэффициентами, повторное применение линейных моделей дает наиболее вероятный результат задержек для всех достижимых последующих событий (максимальный апостериорный запрос).

Exploting once more the acyclicity of the graph representing the plan of operations network and the associated Bayesian network, we can efﬁciently perform the determination of the posterior probabilities of any subsequent event by a visit to all descendants of the event which has just become realised.

Еще раз исследуя ацикличность графа, представляющего план операций сети и связанную с ним байесовскую сеть, мы можем эффективно выполнить определение апостериорных вероятностей любого последующего события, посетив всех потомков только что реализованного события.

The problem here is, that if if is possible that is actually dependent on  , and multiple interrelations between those vehicles are possible. Due to the acyclicity of the network, we can efﬁciently associate a topological order to the Bayesian network, and then just perform the visit of all nodes after the event, i.e. compute  , only when the probabilities of the predecessors of A, B are known or have been already computed.

Проблема здесь в том, что, если, возможно, что фактически зависит от  и возможно множество взаимосвязей между этими транспортными средствами. Благодаря ацикличности сети мы можем эффективно связать топологический порядок с байесовской сетью, а затем просто выполнить посещение всех узлов после события, т. е. вычислить , только если известны или уже рассчитаны вероятности предшественников A, B.

Formally, a topological sorting is a linear ordering of nodes of a graph, such that for any directed arc between nodes and , node appears before node in the ordering. The computational complexity of a topological sorting of an acyclic graph is linear in the amount of nodes and arcs. The greatest advantage of the topological sorting is that computation of the conditional probabilities for any descendants of a node can be done by an efﬁcient linear visit to the ordering (linear in the amount of nodes). This would require at most linear sums, each of which having up to terms, where is the amount of nodes.

Формально: топологическая сортировка - это линейное упорядочение N узлов графа, так что для любой направленной дуги между узлами u и v узел u появляется перед узлом v в порядке. Вычислительная сложность топологической сортировки ациклического графа линейна по количеству узлов и дуг. Самым большим преимуществом топологической сортировки является то, что вычисление условных вероятностей для любых потомков узла может быть получено путем эффективного линейного посещения упорядочения (линейного по количеству узлов). Это потребует не более чем линейных сумм, каждая из которых имеет до N членов, где – N количество узлов.

In our implementation, the time required to sort topologically the graph (in the order of

thousands of nodes and arcs) is about 0.1 second, and the time required to visit all descendants

(events) happening within the next hour is about 0.7 seconds, with about 200 nodes to be

visited. As a comparison, visiting only the next 15 minutes would require about 80 nodes

and 0.6 seconds; visiting up to 2 hours would result in visiting about 300 nodes and 0.8

seconds; visiting the entire network, corresponding to a day of events, would take up to

3.5 seconds. Due to the further stochastic combination and fading phenomena, there is

not much point in computing conditional probabilities for events very far away in time, as delays typically fades away due to buffer times which are inherently considered in the plan

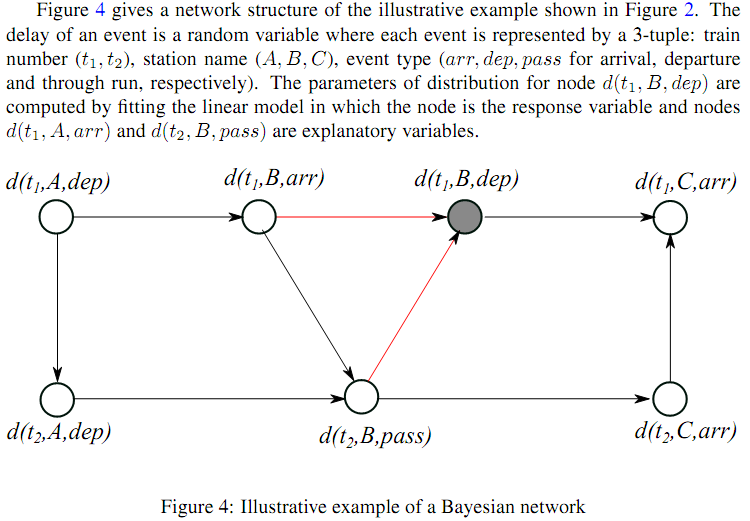
(Andersson et al, 2014), and the variability of the outcomes, and its low accuracy would

make the prediction of very little use. Therefore, the received information typically does

not affect the probability distribution of all events within the prediction horizon (Kecman,

2014).

В нашей реализации время, необходимое для топологической сортировки графа (порядка тысяч узлов и дуг) составляет около 0,1 секунды, а время, необходимое для посещения всех потомков (событий), происходящих в течение следующего часа, составляет около 0,7 секунды, при этом необходимо посетить около 200 узлов. Для сравнения, посещение только следующих 15 минут потребует около 80 узлов и 0,6 секунды; посещение до 2 часов приведет к посещению около 300 узлов и 0,8 секунды; посещение всей сети, соответствующее дню событий, займет до 3,5 секунд. Из-за дальнейшего стохастического комбинирования и явления затухания нет особого смысла в вычислении условных вероятностей для событий, находящихся очень далеко во времени, поскольку задержки обычно исчезают из-за времени буферизации, которое по своей сути учитывается в плане (Andersson et al, 2014), а также из-за изменчивости результатов и его низкой точности, поэтому прогноз будет очень мало пригодным. Следовательно, полученная информация обычно не влияет на распределение вероятностей всех событий в пределах горизонта прогноза (Kecman, 2014).



We discuss here brieﬂy the computation complexity required when the topology of the network changes, for instance as a consequence of a dispatching action, and the network must be updated. From a practical perspective, change of orders require some explicit action from dispatchers, and for this reason, the updates to the plan are typically few and very well traceable. Updating the Bayesian network is then matter of updating only the marginal amount of nodes which have changed the incoming/outgoing arcs, i.e. the vast majority of the topology remains the same, while one or two orders of trains would be changed at any time of implementing a dispatching action. The update has thus a local step, i.e. updating the models at the level of the nodes; and a global step, i.e. checking the properties of the entire network.

Мы кратко обсудим здесь сложность вычислений, которая требуется, когда топология сети изменяется, например, в результате действия диспетчеризации, и сеть должна быть обновлена. С практической точки зрения изменение последовательности требует некоторых явных действий со стороны диспетчеров, и по этой причине обновления плана обычно немногочисленны и очень хорошо отслеживаются. В таком случае обновление байесовской сети сводится к обновлению только незначительного количества узлов, которые изменили входящие / исходящие дуги, т.е. подавляющее большинство топологии остается прежней, в то время как одна или две очередности поездов могут быть изменены в любой момент реализации. Таким образом, обновление имеет локальный этап, т.е. обновление моделей на уровне узлов; и глобальный шаг, т.е. проверка свойств всей сети.

For each node, the update of topology can practically be inverting the direction of an arc; or removing one arc; or inserting an additional arc between two nodes. In the worst case, the amount of nodes affected equals all the nodes of the two trains involved (the order changes in a single point, so at least 2 nodes are directly affected; and in the worst case this order has to be propagated to all events of the two trains, for instance if they follow each other along a corridor). For each of those nodes, a linear model has to be estimated, which would be very fast, due to the small amount of nodes involved. Moreover, the most likely delays (for instance inspired by the recorded data) and associated dispatching actions can be precomputed beforehand, and kept in a library of models to be plugged in, anytime the order of trains would changed, similarly to (Van Thielen et al, 2018).

Для каждого узла обновление топологии может практически изменить направление дуги; или снятие одной дуги; или вставку дополнительной дуги между двумя узлами. В худшем случае количество затронутых узлов равно всем узлам двух задействованных поездов (порядок меняется в одной точке, поэтому как минимум 2 узла подвергаются прямому воздействию; а в худшем случае этот порядок должен распространяться на все события двух поездов, например, если они следуют друг за другом по коридору). Для каждого из этих узлов должна быть оценена линейная модель, которая будет очень быстрой из-за небольшого количества задействованных узлов. Более того, наиболее вероятные задержки (например, вызванные записанными данными) и связанные с ними действия диспетчера могут быть предварительно вычислены и сохранены в библиотеке моделей, которые будут подключены в любое время, когда порядок поездов изменится, аналогично (Van Thielen et al., 2018).

The most computationally challenging part of such a procedure is the check that the

entire network remains acyclic at global level, otherwise the entire Bayesian network loses

meaning. In this case, we remark that the very nature of dispatching actions is to avoid any

occurrence of deadlocks. Those deadlocks are in fact cyclic dependences of train orders,

which are equal to a cyclic dependence of events in graph models, which have been used

extensively in railway optimization (Borndorfer et al, 2018). Such a graph theoretical model

can be further deterministic (such as the alternative graph model used widely for dispatching

in (Mascis and Pacciarelli, 2002; D’Ariano et al, 2007; Corman et al, 2014b) among others)

or stochastic (such as the Bayesian network presented here). In this sense, we are certain

that, based on a feasible and conﬂict free dispatching action, the updated topology and

resulting Bayesian network of events would never be able to determine cycles; and we can

safely reduce the computational effort required to estimate the new linear networks for the

updated nodes, if precomputation would not be possible (in which case, it would be even

faster), avoiding a full check of the acyclity of the network. Under those assumptions, the

need to adjust the structure of the network as response to an action of the dispatcher also

does not seem a critical issue from a computational time point, as the time required to ﬁt a

linear model to a node is in the order of the milliseconds, in our case.

Наиболее сложной с вычислительной точки зрения частью такой процедуры является проверка того, что вся сеть остается ацикличной на глобальном уровне, в противном случае вся байесовская сеть теряет смысл. В этом случае мы отмечаем, что сама природа диспетчерских действий заключается в том, чтобы избежать возникновения взаимоблоки-ровок. Эти тупики на самом деле представляют собой циклические зависимости последовательности поездов, которые равны циклической зависимости событий в графовых моделях, которые широко использовались при оптимизации железных дорог (Borndorfer et al, 2018). Такая теоретическая модель графа может быть дополнительно детерминированной (например, модель альтернативного графа, широко используемая для диспетчеризации в (Mascis and Pacciarelli, 2002; D'Ariano et al, 2007; Corman et al, 2014b) среди других) или стохастической (например, представленная здесь байесовская сеть). В этом смысле мы уверены, что на основе выполнимой и бесконфликтной диспетчеризации обновленная топология и результирующая байесовская сеть событий никогда не смогут определять циклы; и мы можем безопасно уменьшить вычислительные усилия, необходимые для оценки новых линейных сетей для обновленных узлов, если предварительные вычисления будут невозможны (в этом случае это будет еще быстрее), избегая полной проверки ацикличности сети. При этих предположениях необходимость корректировки структуры сети в ответ на действия диспетчера также не кажется критической проблемой с точки зрения вычислительного времени, поскольку время, необходимое для адаптации линейной модели к узлу, в нашем случае составляет порядка миллисекунд.

4 Case study 4.1 Description of the data set

The busy corridor between Stockholm and Norrkoping in Sweden. The corridor comprises the 180 km long northern part of the Swedish southern mainline between Stockholm and Malmo. It is a double-track line with mixed trafﬁc. Passenger trafﬁc is dominant with 90% share that comprises both local and intercity trains. The considered corridor has in total 27 stations and junctions, 10 of which accommodate scheduled stops of passenger and freight trains. Approximately 300 trains per day traverse the corridor (fully or partially).

The calibration of the network assumes the computation of parameters for probability distribution of delay for each considered event represented by a node. The training set for computing the parameters of local distributions depends to a great extent on the network structure. For example, an arc representing a delay propagation during a train run between two stations is calibrated based on historical data for the same train, as identiﬁed by the train number, which repeats daily. Similarly, the relevant training set for an arc representing

a headway arc from an event of another train, comprises only the days when the two trains

ran in the same relative order. This means that the recurring phenomena for a particular

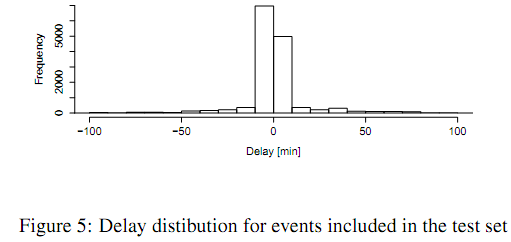
event that may happen due to: (i) variations in travel demand (peak/off-peak hours), (ii) distribution of time reserves in the timetable, are implicitly represented in the model.

Калибровка сети предполагает вычисление параметров распределения вероятностей задержки для каждого рассматриваемого события, представленного узлом. Обучающий набор для вычисления параметров локальных распределений в значительной степени зависит от структуры сети. Например, дуга, представляющая распространение задержки во время движения поезда между двумя станциями, калибруется на основе исторических данных для одного и того же поезда, что определяется номером поезда, который повторяется ежедневно. Точно так же соответствующий обучающий набор для дуги, представляющей дугу движения от события другого поезда, включает только дни, когда два поезда проходили в том же относительном порядке. Это означает, что повторяющиеся явления для конкретного события, которые могут произойти из-за: (i) изменений спроса на поездки (часы пик / непиковый период), (ii) распределения временных резервов в расписании, неявно представлены в модели.

4.2 Experimental setup

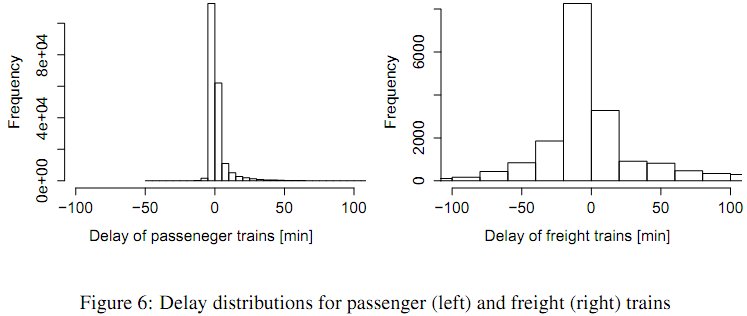
The dynamic component of the experimental environment comprises the actual process plans for all trains within the prediction horizon and actual train event times. The actual route for each train is given on the level of station events, i.e., ordered list of stations and event types. This is crucial for building the probabilistic graph model. As the prediction horizon moves, new trains are added to the model.

Динамический компонент экспериментальной среды включает фактические планы процессов для всех поездов в пределах горизонта прогноза и фактическое время поездных событий. Фактический маршрут для каждого поезда указан на уровне событий на станциях, то есть, задан упорядоченный список станций и типов событий. Это очень важно для построения вероятностной графовой модели. По мере перемещения горизонта прогноза в модель добавляются новые поезда.



4.3 Modelling of passenger and freight train delays

The passenger trains operate strictly according to a timetable. Their deviation from the scheduled paths is relatively small and therefore they frequently follow the scheduled orders, thus providing sufﬁcient amount of data in the training set to build the robust estimates of the distribution parameters. On the other hand, freight trains often signiﬁcantly deviate from their schedule and they are often running on the ad hoc created paths that do not recur frequently.



The interaction between ad hoc freight train paths and successive and preceding trains

could be modelled using the dependence between the planned headway time between two

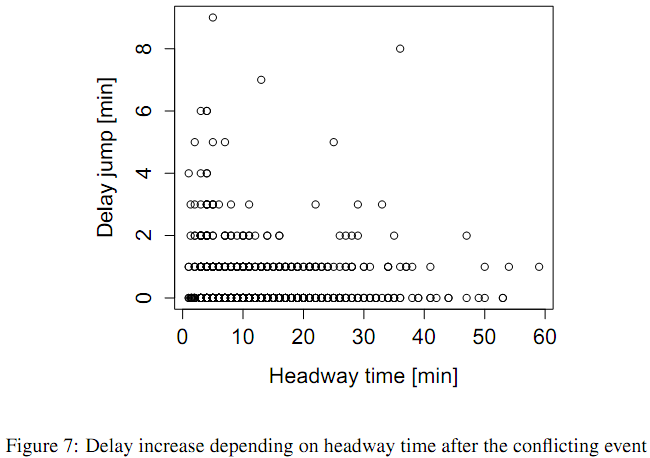
events and the resulting delay increase for the second train. A delay increase is computed as

a difference in delay between two successive events of a single train. Delay increase for all

trains which may be delayed due to a hindrance from a freight train were extracted from the

data and compared with the observed headway time between the two trains. The results are

presented in Figure 7.



As anticipated there is a general trend that delay increase is higher and more likely to happen for short headway time after a freight train. However, regression analysis showed no signiﬁcant impact of headway time on delay increase with the value of indicating that only 8% of variation of a delay increase can be explained by the short headway time after the preceding train. In fact the analysis showed that mean and median of headway times after an inserted freight train path are 15.85 and 12.43 minutes, respectively.

Как и ожидалось, существует общая тенденция, согласно которой увеличение задержки больше и задержка произойдет с большей вероятностью при малом интервале движения после грузового поезда. Однако регрессионный анализ не показал значительного влияния интервала движения на увеличение задержки, при этом только 8% роста вариаций задержки можно объяснить малым интервалом движения после предыдущего поезда. По факту, анализ показал, что среднее и медианное значение интервала после введенного грузового поезда составляют 15,85 и 12,43 минут соответственно.

This could be explained by the fact that trafﬁc controllers carefully choose the path for ad

hoc freight trains so that the scheduled trafﬁc is affected as little as possible. For that reason

we exclude the headway arcs between the events of out-of-schedule freight trains and the

preceding and succeeding passenger trains from the model. Similar analysis is required for

model applications on mixed-trafﬁc corridors with a higher share of ad-hoc freight trains.

Such trains should be explicitly included in the network with the corresponding headway

arcs constructed between them and other trains.

Это можно объяснить тем фактом, что диспетчеры тщательно выбирают маршрут для специальных грузовых поездов, чтобы как можно меньше повлиять на график движения. По этой причине мы исключаем из модели дуги движения между событиями неграфиковых грузовых поездов, предшествующих и последующих пассажирских поездов. Аналогичный анализ требуется для применения моделей в коридорах со смешанным движением с более высокой долей специальных грузовых поездов. Такие поезда должны быть явно включены в сеть с построением соответствующих интервальных дуг между ними и другими поездами.

5 Results

5.1 Model structure, size and parameters

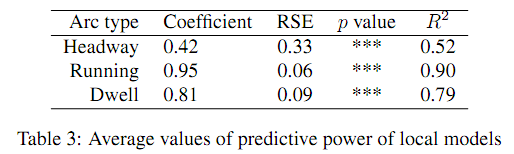
An update about a realised event time is therefore propagated one hour in the future which is the longest prediction horizon we consider. The average network size is 137.12 nodes and 262.43 arcs. The average number of children per node is therefore 1.92 which indicates that approximately two explanatory variables are needed to compute the distribution parameters for each node.

Table 3 shows the predictive power of local linear models averaged over the 2194 events

considered in the test set. The importance of each explanatory variable is considered sepa-

rately in order to analyse the correlation of delay increase during different processes. The

table shows correlation coefficients, residual standard error (RSE - остаточная стандартная ошибка), value as an indicator of variable importance (\*\*\* means p<0,01 which indicates very strong importance) and R2 which presents the percentage of variance explained by the predictor. The strongest correlation is captured between the delays of events associated with a running process of a train.



Arrival delay of a train is also a good predictor of its departure delay from the same station

(dwell process). However, lower delay correlation between an arrival and the subsequent

departure event reﬂects the fact that dwell times are more difﬁcult to predict and may act as a source of delay as well as a delay buffer (Kecman and Goverde, 2015). Тем не менее, более низкая корреляция задержки между прибытием и последующим событием отправления отражает тот факт, что время стоянки труднее предсказать, и оно может действовать как источник задержки, а также буфер задержки.

A delay of the preceding conﬂicting event (headway process) is also an important predictor although with a weaker predictive power (lower value of R2). The average multiple R2 with complete local models explains 92% of delay variation.

5.2 Prediction model performance

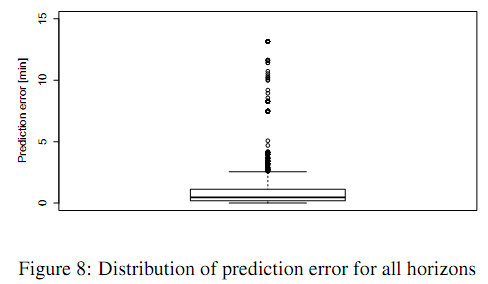
This section reports the prediction accuracy of the model when applied on the peak hours

(6:30-9:00 and 16:30-19:00) of the test day. After the observation of each train event in

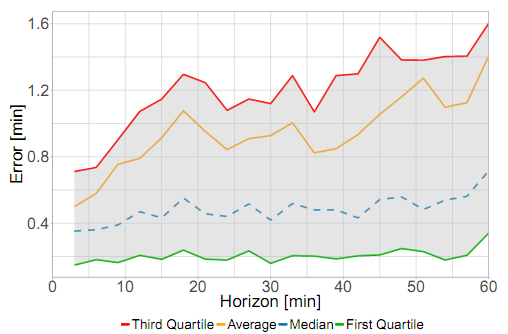
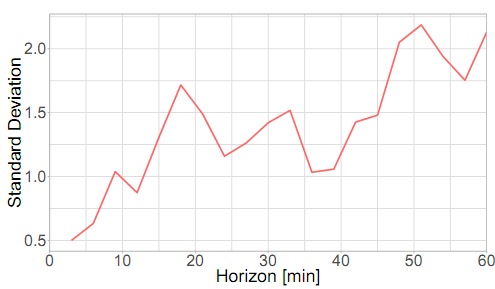
the speciﬁed period, the algorithm predicts the future trafﬁc evolution in the next hour. The predicted values are compared against the realised event times and the distribution of prediction error for all predictions is given in Figure 8.

The box-plot indicates the median (line in the middle of the box), the 1st and the 3rd quartiles (upper and lower bound of the box) and data maximum and minimum (ends of the upper and lower whisker). Despite the outliers in prediction errors, which are not excluded from the analysis, the plot shows a high prediction accuracy of the Bayesian network model.

Коробчатая (ящичковая) диаграмма указывает медиану (линия в середине прямоугольника), 1-й и 3-й квартили (верхняя и нижняя границы прямоугольника), а также максимум и минимум данных (концы верхнего и нижнего элементов). Несмотря на выбросы в ошибках предсказания, которые не исключены из анализа, график показывает высокую точность предсказания модели байесовской сети.



Mean absolute error (MAE) is obtained in each interval by computing the mean value of all corresponding absolute prediction errors. Figures 9 and 10 respectively show the MAE (as average, median, and ﬁrst and third quartile), and standard deviation, for each considered prediction horizon. As expected, both MAE and standard deviation decrease as the smaller prediction horizon is considered. While average and mean increase for longer time horizons, also the gap (i.e. the area of the band in Figure 9) between the ﬁrst and the third quartile increases, (and similarly the standard deviation ins Figure 10) reporting the lower degree of conﬁdence on the values predicted, and not only a higher error, but also a higher variability of the error.

5.3 Comparison of different prediction approaches

We also compare different algorithms from the state of the art of dynamic prediction, in

Figure11. Namely, we compare the approach proposed using the Bayesian network to

encode the structure of the dependencies (named Bayesian in Figure), with a variety of other

approaches. One is to consider only dependencies between the events of the same train, but

deterministically. In other terms, this approach (named Propagation in Figure) assumes a

train in the future would keep the current delay. А именно, мы сравниваем подход, предложенный с использованием байесовской сети для кодирования структуры зависимостей (названный байесовским на рисунке), с множеством других подходов. Один рассматривает только зависимости между событиями одного и того же поезда, но детерминистически. Другими словами, этот подход (названный на рисунке «Распространение») предполагает, что поезд в будущем сохранит текущую задержку.

We also report two different approaches, which perform a nearest neighbor search in the space of the recorded data, to deliver the prediction (based on the same ideas as (Bohmova et al, 2015), to ﬁnd events happened in the past, which showed similar characteristics of the current moment. This works as follows. Мы также сообщаем о двух разных подходах, которые выполняют поиск ближайшего соседа в пространстве записанных данных, чтобы сделать прогноз (на основе тех же идей, что и (Бомова и др., 2015)), чтобы найти события, произошедшие в прошлом, которые показали аналогичные характеристики текущего момента. Это работает следующим образом.

We assume a future event ei described by a train number t, at a station, and event type, is to be predicted, given a current event ej. The predicted delay of ei is computed as the average of the delay of all events ek recorded in the past, which have either 1) the same train number at the same station (named *NN-Train* in Figure); or 2) the same train number at the same station, while the delay associated to ek is the same (or the most similar) as the delay associated to event ej, this latter corresponding to the most recent observation of the run of the train of both events ei and ej. This latter Nearest Neighbor structure is named *NN-Train-delay* in Figure. We also tried other Nearest Neighbor structures based on time intervals, which did not perform better than those presented.

Мы предполагаем, что будущее событие ei, описываемое номером поезда t на станции и типом события, должно быть предсказано по текущему событию ej. Предсказанная задержка ei вычисляется как среднее значение задержки всех событий ek, зарегистрированных в прошлом, которые имеют либо 1) одинаковый номер поезда на той же станции (обозначенной *NN-Train* на рисунке); или 2) тот же номер поезда на той же станции, а задержка, связанная с ek, такая же (или наиболее близкая), что и задержка, связанная с событием ej, причем последнее соответствует самому последнему наблюдению за движением поезда по обоим событиям ei и ej. Эта последняя структура ближайшего соседа на рисунке называется *NN-Train-delay*. Мы также попробовали другие структуры «ближайших соседей», основанные на временных интервалах, которые не показали лучших результатов, чем представленные.

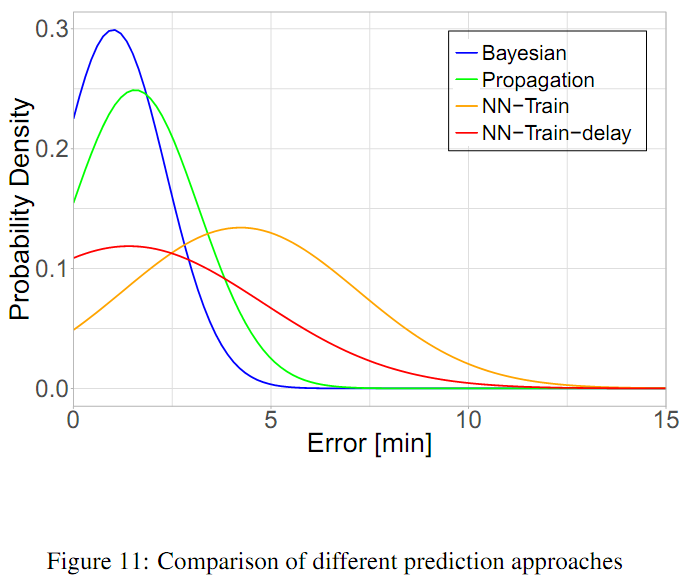


Figure 11 reports the kernel estimate of the probability distributions of the error, for all

events, such that the time horizon of prediction between events and is between 35 and 40 minutes. A curve more on the left side of the Figure represents smaller errors, in prob-

abilistic terms. From this point of view, the advantage of the Bayesian network against the

other approaches is evident, delivering a higher probability of smaller errors.

It is interesting to see how the actual information on the delay (in deterministic terms, the propagation approach); in statistic terms, the difference between NN-train and NN-train-delay is actually very important. This underlines the strength and potential of dynamic approaches which are able to combine past information on delays, with current (online) information on delay, such as the Bayesian approach, which outperforms all other approaches presented here. Интересно посмотреть, как актуальная информация о задержке (в детерминистских терминах, метод распространения); с точки зрения статистики разница между *NN-train* и *NN-train-delay* на самом деле очень важна. Это подчеркивает силу и потенциал динамических методов, которые могут сочетать прошлую информацию о задержках с текущей (онлайн) информацией о задержках, таких как байесовский подход, который превосходит все другие представленные здесь методы.

5.4 Dynamic updates of probability distributions

Finally, we analyse the dynamics of probability distribution of an event over time as more

information about the event becomes available. Bayesian inference is used to compute the

posterior distributions (Korb and Nicholson, 2010). Figure 12 shows an example of how

the distribution of arrival time of a train to the ﬁnal station evolves over time in six discrete

steps. As the event becomes closer (horizon H decreases), the tendency is that the standard

deviation becomes smaller thus achieving sharp distributions that converge toward a 1-point

distribution at the moment when the event is realised. On the other hand, the expected value of the distribution ﬂuctuates a bit over its ﬁnal value.

Наконец, мы анализируем динамику распределения вероятности события во времени по мере того, как становится доступной дополнительная информация о событии. Байесовский вывод используется для вычисления апостериорных распределений (Korb and Nicholson, 2010). На рис. 12 показан пример того, как распределение времени прибытия поезда на конечную станцию изменяется во времени в виде шести дискретных шагов. По мере приближения события (уменьшение горизонта H) наблюдается тенденция к тому, что стандартное отклонение становится меньше, что приводит к получению узких распределений, которые сходятся к 1-точечному распределению в момент реализации события. С другой стороны, ожидаемое значение распределения немного отклоняется по сравнению с его окончательным значением.