

ORIGINAL PAPER

Open Access



# Feature selection and extraction in spatiotemporal traffic forecasting: a systematic literature review

Dmitry Pavlyuk

## Abstract

A spatiotemporal approach that simultaneously utilises both spatial and temporal relationships is gaining scientific interest in the field of traffic flow forecasting. Accurate identification of the spatiotemporal structure (dependencies amongst traffic flows in space and time) plays a critical role in modern traffic forecasting methodologies, and recent developments of data-driven feature selection and extraction methods allow the identification of complex relationships. This paper systematically reviews studies that apply feature selection and extraction methods for spatiotemporal traffic forecasting. The reviewed bibliographic database includes 211 publications and covers the period from early 1984 to March 2018. A synthesis of bibliographic sources clarifies the advantages and disadvantages of different feature selection and extraction methods for learning the spatiotemporal structure and discovers trends in their applications. We conclude that there is a clear need for development of comprehensive guidelines for selecting appropriate spatiotemporal feature selection and extraction methods for urban traffic forecasting.

**Keywords:** Feature selection, Feature extraction, Traffic forecasting, Spatiotemporal, Forecasting models, Systematic review

**Subject classification codes:** C33, C45, C51, C53, R41

## 1 Introduction

Spatiotemporal traffic forecasting is based on advanced models that utilise traffic flow information both in spatial and temporal dimensions. Accurate identification of the spatiotemporal structure is an emerging problem of modern forecasting methodologies. Although dependencies between traffic flows at connected road network segments are perfectly supported by the traffic flow theory, their capture for forecasting purposes is a challenging task. Spatiotemporal relationships are not limited by road connectivity but include links between remote (in space and time) points that appear owing to common patterns and interdependence of traffic flows and indirectly connected urban road segments. We consider identification of spatiotemporal dependencies as a special case of the feature selection problem. The objective of

feature selection is to identify a subset of relevant model inputs (features) that simplify the model structure and estimation procedure, yet still provide good forecasting results.

This paper reviews studies that empirically utilise spatiotemporal traffic flow forecasting models, paying special attention to applied feature selection and extraction (FSE) methods. Thus, four main questions for this review are:

- Which FSE methods are applied for spatiotemporal structure identification in empirical traffic forecasting studies? What are the recent trends in this area?
- What is the role of spatiotemporal FSE methods in a methodology of urban traffic forecasting? Is this role acknowledged in existing literature?
- How are spatiotemporal traffic forecasting methodologies empirically covered by different FSE

Correspondence: [Dmitry.Pavlyuk@tsi.lv](mailto:Dmitry.Pavlyuk@tsi.lv)

Mathematical Methods and Modelling Department, Transport and Telecommunication Institute, LV-1019, Lomonosova str. 1, Riga, Latvia



© The Author(s). 2019 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

methods? Are there methodological gaps that should be covered?

- Do the researchers have principles or guidelines for selecting a proper spatiotemporal structure to measure spatial dependencies between traffic links?

Answering these questions, we reveal uncovered methodological areas of spatiotemporal traffic forecasting and suggest directions for future research.

The methodology of the review is based on an intensive literature search and critical analysis. We executed a critical review of a large number of publications to reduce the risk of review bias and missed methodological branches.

This paper is closely linked with several existing reviews but has its own focus and advantages. Firstly, Vlahogianni et al. [1] provided a comprehensive review of 67 papers focussed on traffic forecasting objectives and methods. Although this review is not focussed on spatiotemporal models, it can be used to observe the progress that the scientific community made from 2004. Later, the same authors [2] suggested the identification of spatiotemporal relationships as an important research direction in traffic flow forecasting. Haworth, in another related review [3], evaluated different types of spatiotemporal structures and covered 39 publications. Finally, Ermagun and Levinson [4] presented an extensive review of 130 publications on spatiotemporal traffic forecasting. The methodology of urban traffic forecasting includes analysis and decision making on many critical aspects – forecasting horizon, utilised model and its specification, look-back time interval, temporal resolution of traffic data, measurement of forecasting accuracy, periodic structure of traffic flows, and recurring/abnormal traffic conditions, amongst several others. Each review focussed on its own set of methodological issues, and the novelty of this review also lies in the set of covered topics – we concentrate on spatiotemporal structure identification (via FSE) as a crucial step in spatiotemporal traffic forecasting. Selection of spatiotemporal FSE methods is closely related to the utilised forecasting model, its topology, and the size of an analysed road network, and these characteristics are part of the main focus of this review.

The remainder of this paper is organised as follows. Firstly, we provide a detailed description of the review methodology. Secondly, we present the definition of the spatiotemporal structure and substantiate the problem of spatiotemporal FSE. Thirdly, we classify existing FSE methods and present a review of their use for spatiotemporal traffic forecasting. Fourthly, we present a review of applied methodologies based on utilised FSE methods to discover potential gaps in the literature. Finally, we summarise the current state of

the reviewed area and propose several future research directions.

## 2 Methodology of the review

### 2.1 Search strategy

The literature on FSE in traffic modelling and forecasting is very extensive. The scope of this review is limited to the following dimensions:

- (1) Focus on simultaneous utilisation of spatial and temporal dimensions of traffic flows. Use of the temporal dimension is typical in traffic forecasting, but the spatial dimension (relationships amongst traffic flows at different spatial locations) is ignored in many studies. We included only publications where the spatial dimension is explicitly used in the empirical part of the research (we excluded studies that state a potential utility of spatiotemporal information, but do not use it in practice).
- (2) Focus on empirical applications of spatiotemporal FSE. Thus, we excluded purely theoretical research studies from this review that rarely deal with empirical FSE problems. However, we did include studies that use simulated traffic flow data for analysis of FSE and apply the forecasting methodology.
- (3) Focus on short-term traffic forecasting. We concentrated on studies devoted to short-term traffic forecasting at specified spatial locations; therefore, we excluded studies on a wide range of traffic modelling problems (accident prediction, missing data imputation, travel time prediction, origin-destination matrix estimation, and construction of fundamental diagrams) where spatiotemporal information is also naturally utilised. This exclusion was implemented manually so that we include studies that oriented on another traffic modelling problem (e.g. routing) but solve it via spatiotemporal forecasting.
- (4) Focus on the stochastic nature of spatiotemporal dependencies. We assumed that the spatiotemporal structure of traffic flows is dynamic and stochastic; therefore, it should be estimated on the basis of traffic data. Thus, we excluded studies where spatiotemporal relationships are predefined (e.g. studies based on kinematic wave models).
- (5) Focus on vehicle traffic flows. We excluded studies devoted to bicycles, pedestrians and public transport modelling.

To identify relevant studies, we utilised the following academic search engines: TRID, Scopus, IEEE Xplore, IET Digital Library (search by titles and abstracts),

Google Scholar, and Science Direct (full-text search). The general search pattern was as follows:

*spa\* tempor\* traffic (forecast\* OR predict\*),*

where \* is a wildcard and OR is a logic operator. This pattern covers different references to the spatial dimension (“spatial”, “spatiotemporal”, “space”) and different references to forecasting (“forecast”, “forecasting”).

The search yielded 1186 articles, which were further filtered on the basis of the five criteria specified above. Filtering was performed manually, but we recommend the following set of exclusion keywords that can be used for automatic filtering with a low chance of missing a relevant paper:

*NOT in (“animal\*”, “bus”, “bicyc\*”, “CO2”, “accident\*”, “incident\*”, “generation”, “demand”, “accessi\*”, “household\*”, “freight\*”, “emergenc\*”, “air\*”, “emiss\*”, “wind\*”, “parking\*”, “sharing”)*

The filtered list of publications was complemented by results of forwards and backwards reference snowballing. The resulting bibliographic database includes 211 publications (135 journal articles, 64 conference papers, and 12 theses/scientific reports). Despite the fact that the bibliography appears to be too extensive for a review, we decided to include all publications but limit the discussion regarding FSE methods to groups of studies. A complete list of publications, presented in the Appendix, can be useful for further review of other aspects of spatiotemporal traffic forecasting. Analysed information in every publication includes the following:

- applied spatiotemporal methodology(ies),
- utilised FSE methods, separate for spatial and temporal dimensions,
- topology of the analysed road network segment,
- number of spatial points (sensors or links) in the analysed road network segment,
- alternative non-spatial models,
- data source (country), and
- number of citations.

The last point was included for information purposes only and was not used for publication filtering.

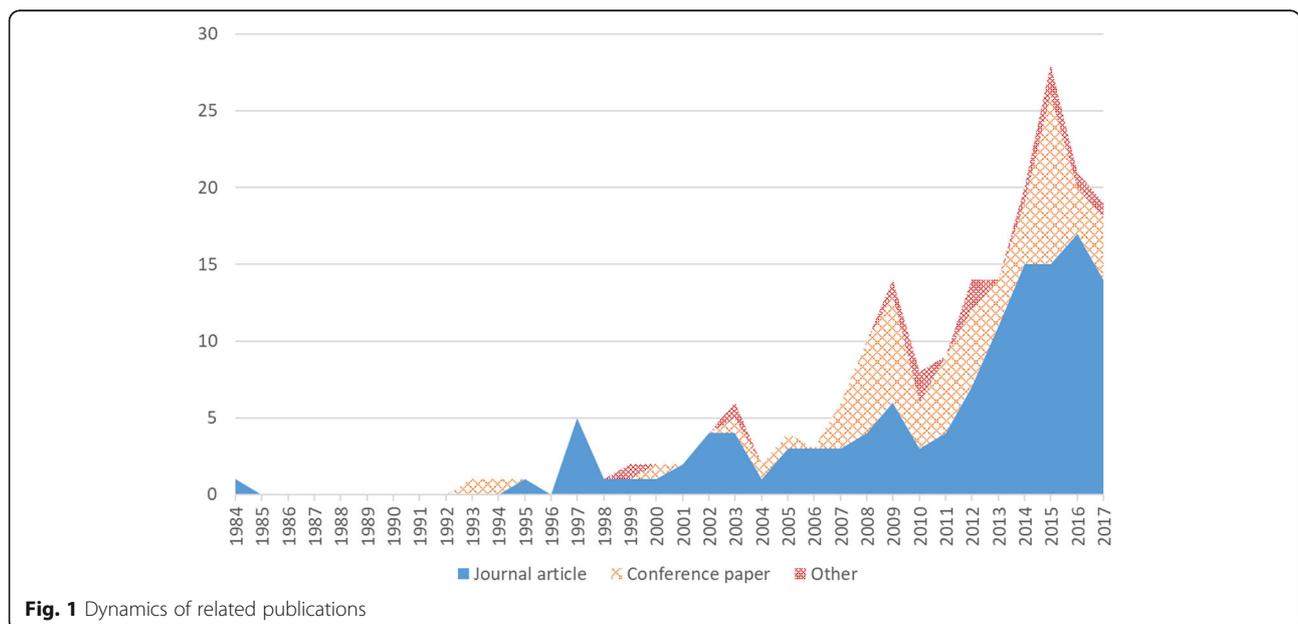
The dynamics of the publication numbers from 1984 to 2017 are presented in Fig. 1 and illustrate the growing interest in spatiotemporal traffic forecasting.

Taking into account the observed trend and number of publications in 2018 (13 publications from January to March 2018), we expect further growth of scientific interest in this field.

Reviewing the publications, we focused on two key elements:

- Applied forecasting methodology (spatiotemporal models and their alternatives)
- Utilised FSE methods

The range of utilised methodologies is fairly large; amongst the most popular we note: feed-forward neural networks (FFNN), k-nearest neighbour (KNN) regression, support vector regression (SVR), Bayesian networks (BN), univariate autoregressive distributed lag (ARDL) model, vector autoregressive (VAR) model, and space-time autoregressive integrated moving average (STARIMA) model.



**Fig. 1** Dynamics of related publications

The list of applied spatiotemporal FSE methods is also wide, and its analysis requires preliminary classification.

Analysing the topology of the analysed road segment, we classified the studies into three possible network configurations:

- Sequential allocation of spatial points along a freeway,
- Sequential allocation of spatial points along an arterial road,
- Complex network of spatial points

We did not use the conventional traffic engineering road hierarchy for separating freeways and arterial roads; instead, we analysed the frequency of intersections and driveways on the analysed road segment and classified the topology as a freeway if this frequency was relatively low. Any non-sequential placement of spatial points was classified as a network topology.

The dynamics of the analysed topologies are presented in Fig. 2.

We preliminarily conclude that the growing number of studies devoted to spatiotemporal urban traffic forecasting in complex non-sequential spatial settings require specific attention to spatiotemporal structure identification.

**2.2 Definition of the spatiotemporal structure**

Firstly, we provide a formal definition of the spatiotemporal structure to be identified by FSE methods. Assume we have  $n$  spatial locations (sensors, road links, clusters of links) ( $i = 1, \dots, n$ ) that are observed during  $T$  time periods ( $t = 1, \dots, T$ ) (in this paper we consider a discrete representation of the spatiotemporal structure of traffic

flows). Observed data for the target indicator  $y$  (e.g. traffic volume, speed) is presented as an  $n \times T$  matrix,  $=\{y_{i,t}\}$ , that may contain missing values. Thus, the goal of one-step ahead forecasting is estimation of the function  $f$  that maps  $Y$  to values of the target indicator for a time period  $(t + 1)$  for all spatial locations  $i$ :  $\hat{y}_{i,t+1} = f(Y)$ .

Following George and Kim [5], we define the spatiotemporal network (STN) as a dynamic structure of dependencies that includes links between spatial locations at different time periods and may change over time. An STN structure may be represented in the form of a weighted time-expanded graph (Fig. 3).

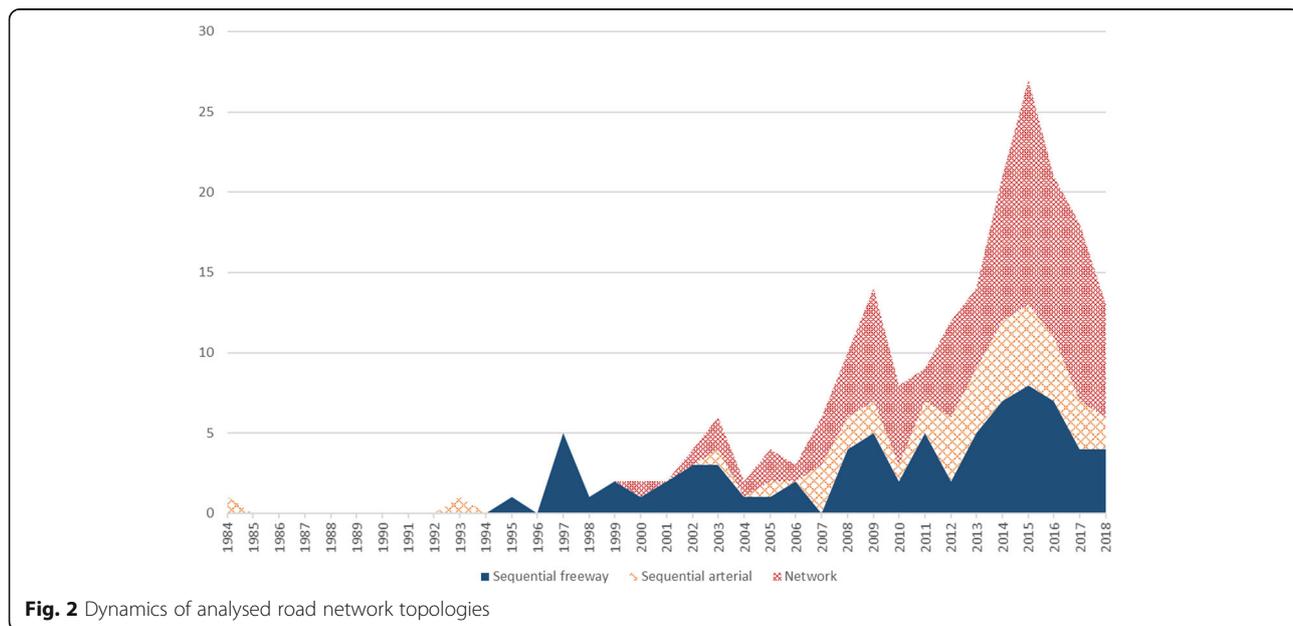
We assume that weights of the time-expanded graph represent the power of the relationship between two graph nodes. Such weights are normally considered as not exogenously provided and their estimation is included in modelling methodologies.

Note that the structure of dependencies in the STN does not necessarily correspond to the physical road network structure, because dependencies generally vary for different levels of time aggregation and may appear even between remote road links.

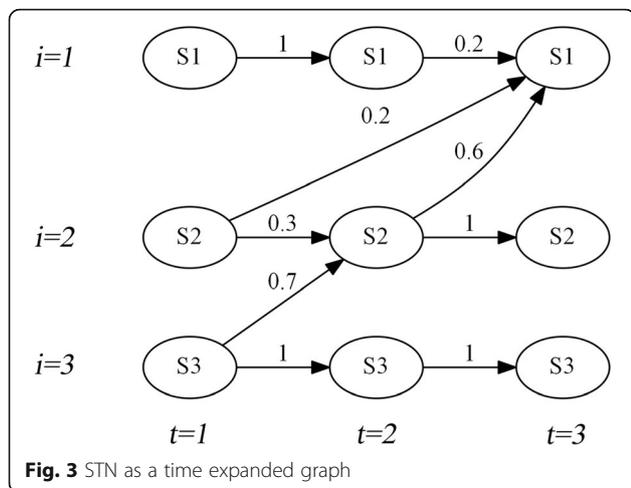
For modelling purposes, the STN is usually presented in matrix form. Let  $\theta$  represent a set of dependencies for the spatial location  $i$  at the time period  $t$  as an STN matrix:

$$\theta_{i,t} = \begin{pmatrix} \theta_{1,1} & \theta_{1,2} & \dots & \theta_{1,t-1} \\ \theta_{2,2} & \theta_{2,2} & \dots & \theta_{2,t-1} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{n,1} & \theta_{n,2} & \dots & \theta_{n,t-1} \end{pmatrix}$$

Coefficients in the STN matrix represent weights in the time-expanded graph and conventionally are set to



**Fig. 2** Dynamics of analysed road network topologies



**Fig. 3** STN as a time expanded graph

zero for absent dependencies (missing edges). We refer to the zero-valued coefficients in STN matrices as STN sparsity. Note that we distinguish between STN matrices and matrices of spatial weights, as is common in empirical research. We use the “spatial weights” term for exogenous information regarding spatiotemporal dependencies as acknowledged in some methodologies (e.g. STARIMA); whereas, STN matrices are estimated by the methodology being applied. Also, note that some methodologies (e.g. spatial panel models) allow spatial dependencies within the same time moment; therefore, the STN matrix  $\theta_{i, t}$  will include one additional column with coefficients for dependencies at time period  $t$ . In this study, we consider traffic forecasting methodologies that usually do not rely on the availability of any information at the time period  $(t + 1)$ ; therefore, we continue with STN matrices as defined above for simpler formulations.

A complete STN structure includes the STN matrices for all spatial locations at all time periods:  $STN = \{\theta_{i, t}\}$ . For example, for the STN structure presented on Fig. 3, the STN matrices are:

$$\theta_{1,2} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}; \theta_{2,2} = \begin{pmatrix} 0 \\ 0.3 \\ 0.7 \end{pmatrix}; \theta_{3,2} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix};$$

$$\theta_{1,3} = \begin{pmatrix} 0 & 0.2 \\ 0.2 & 0.6 \\ 0 & 0 \end{pmatrix}; \theta_{2,3} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}; \theta_{3,3} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$$

We will refer to the STN structure as static if a set of STN matrices does not depend on  $t$ :  $\theta_{i, t} = \theta_i$  for all  $t$ . Otherwise, the STN structure is considered as dynamic.

It should be noted that the total number of parameters in the STN structure is extremely large: the maximum total number of non-zero coefficients for the time moment  $t$  is  $(t - 1) \times n^2$  and for the complete structure is  $(t - 1)! \times n^2$ .

Taking into account that modern intelligent transportation systems (ITS) include several thousand detectors (spatial locations), the total number of coefficients could reach several millions. Dealing with such a large number of parameters is impractical owing to the well-known curse of dimensionality problem, and thus, the problem of selection of the most important features is critical in spatiotemporal traffic flow forecasting.

### 3 Results and discussion

#### 3.1 Review of spatiotemporal FSE methods

The range of utilised FSE methods is extensive. Following the classification of feature selection methods by Chandrashekar and Sahin [6], we conventionally divided FSE methods into the following five classes:

- (1) *Exogenous feature filtering methods* that utilise information regarding dependencies in traffic flows explicitly provided by a researcher.
- (2) *Endogenous feature filtering methods* that select the most informative features using traffic data  $Y$ . Note that both exogenous and endogenous filtering methods select spatiotemporal features before application of forecasting models.
- (3) *Wrapper feature selection methods* that use information about forecasting model performance to determine the optimal set of features.
- (4) *Embedded feature selection methods* that consider feature selection as an internal process of a forecasting methodology.
- (5) *Dimension reduction methods* that reduce the dimensionality of the problem on the basis of clustering or feature extraction techniques. Within the scope of this review, we consider dimension reduction as an alternative technique to learn spatiotemporal relationships that are useful for traffic forecasting.

Note that the presented classification does not correspond to different approaches or data analyses (such as supervised or unsupervised learning) but is based on a point of the forecasting process, where the STN is identified. Exogenous feature filtering is executed before analysis of traffic flow data; endogenous feature filtering and dimension reduction methods use traffic flow data but are applied before construction of a forecasting model; embedded feature selection is executed within a forecasting model; and wrapper feature selection is based on the evaluation results of the forecasting model. FSE methods of the different classes may be applied simultaneously to ensure a maximally sparse STN, but this is rarely utilised in existing studies. Note that spatiotemporal FSE methods may act in two dimensions—spatial and temporal; therefore, we review them separately for all

the classes. A complete list of FSE methods utilised for spatiotemporal traffic forecasting is presented in the Appendix and summarised in Table 1.

The dynamics of different classes of FSE methods in the spatial dimension are presented in Fig. 4.

Exogenous feature filtering is a prevailing class of methods used in 57% of the analysed studies, but its percentage is gradually decreasing (it is less than 50% in the past 5 years). The percentage of other classes that represent the importance of various FSE methods for modern forecasting methodologies is increasing.

### 3.1.1 Class 1: Exogenous feature filtering methods

The most natural explanation of spatiotemporal relationships in traffic flow is based on cars' movement: if a car is observed at a spatial point, it is expected to be observed later at another, downstream point. This fact creates a background for the most popular exogenous feature filtering approach (utilised in 44 studies) – to limit spatiotemporal dependencies to one direct upstream neighbour location. This approach perfectly matches the classic macroscopic traffic flow theory, and its effectiveness for traffic volume prediction is supported by many studies. For other traffic characteristics such as speed or travel time, the direction of this relationship could be different—congestion at a downstream spatial location affects upstream traffic flow; therefore, four studies consider selection of a direct downstream neighbour as a separate alternative specification of spatiotemporal links, and 34 studies simultaneously consider direct upstream and downstream neighbours. Approaches based on direct neighbours work well if two basic conditions are satisfied: 1) a time delay interval (time lag) of phenomena (traffic volume, speed, etc.) between spatial locations is identified correctly, and 2) the analysed road segment is a linear arterial road without traffic signals or ramps. The first issue can be solved within modern forecasting methodologies, but the second one is very limiting for real world urban road networks. A potential workaround is to include the number of intersections (of different types) into the model [7], but the general treatment is to model links between neighbouring spatial locations via independent model parameters. Thus, the Bayesian network, which allows a separate identification of every link, is the most popular modern methodology (10 studies) utilising direct neighbour-based spatiotemporal FSE.

A natural extension of the direct neighbour-based approach is simultaneous utilisation of several upstream locations (13 studies) or a predefined spatial “window” of upstream and downstream locations (8 studies). This approach is more flexible with respect to time lag identification, but in the case of a large interconnected network, it is highly dimensional and requires additional

filtering of features. Convolutional neural networks, a modern deep learning approach applied in four studies [8–11], utilise a predefined spatiotemporal window as an input and implement further FSE by embedded mechanisms.

Many researchers (26 studies) simultaneously utilised data from all available spatial locations, but given that most case studies included only a limited road network segment, this approach can be considered as a special case of the “window” feature selection.

Several researchers (six studies) utilised travel times between locations to reduce the number of spatial links (by excluding locations that are too close and too far to have an explainable influence within a specified time lag). For instance, Min and Wynter [12] utilised this approach to limit the number of coefficients in their vector autoregressive model and found it beneficial for traffic forecasting accuracy. If travel times between spatial locations are assumed as equal, these restrictions could allow use of a higher order neighbourhood (i.e. neighbours of neighbours are included in relationships for the second time lag). Higher order neighbours are typical in STAR-IMA models and were utilised in seven studies, based on this methodology.

An alternative exogenous feature filtering approach, which is not directly based on connections between spatial locations, has been suggested by Ermagun and Levinson [13–15]. The introduced network weight matrix utilises graph characteristics of the road network such as betweenness centrality and vulnerability to discover complementary and competitive spatial links. Network weights can be purely graph-based or enhanced by associated characteristics of traffic flows (e.g. weighted by traffic volume). Associated links are not necessarily connected directly, thus, such spatiotemporal relationships can reach beyond the bounds of the physical road network.

Finally, exogenous filtering of spatiotemporal relationships can be performed on the basis of individual cars' routes. Stathopoulos, Dimitriou, and Tsekeris [16, 17] report application of a micro-simulation procedure for a detailed analysis of spatiotemporal links under different traffic conditions. To the best of our knowledge, there are no studies that utilise real cars' routes for FSE purposes, although the growing availability of probe cars' data creates the possibility of new developments in this direction.

Considering the temporal dimension, the most popular exogenous feature selection method (utilised in 95 studies) is to set a maximum time lag  $T$  and include all lags  $\{1, 2, \dots, T\}$  in the model. The maximum time lag is usually based on the size of an analysed road segment (to allow cars to leave the segment before the specified period). This approach works well for small road

**Table 1** Spatiotemporal FSE methods

Method	Short description	Number of studies	
		Spatial	Temporal
Exogenous filtering methods			
All	All spatial locations within the research road segment	26	–
All upstream	All upstream spatial locations within the research road segment	13	–
Upstream	Only direct upstream neighbour(s)	44	–
Upstream + downstream	Only direct upstream and downstream neighbours	34	–
Downstream	Only direct downstream neighbour(s)	4	–
Higher order	Higher order neighbours (neighbours of neighbours), starting from the second order	7	–
Window	Several upstream and downstream neighbours	8	–
Predefined maximum lag	A set of lags $\{1, 2, \dots, T\}$ , where $T$ is a predefined maximum time lag	–	95
Travel time	With one of the dimensions (spatial or temporal) fixed, the other can be limited by travel time between spatial locations	6	6
Micro-simulation	Estimate spatiotemporal relationships using individual cars' routes	2	–
Network characteristics	Use network characteristic (i.e. betweenness centrality and vulnerability) to discover complementary links	3	–
Endogenous filtering methods			
CCF	Cross-correlation function between traffic at different spatial locations	32	26
Graphical LASSO	Graphical least absolute shrinkage and selection operator	4	1
Granger causality	Granger causality tests, incl. Vector autoregressive model	2	2
LARS	Least-angle regression	3	2
MARS	Multivariate adaptive regression splines	3	3
Custom	Authors' custom formulas (e.g. a combination of physical distance and correlation between spatial locations)	9	2
Wrapper methods			
Empirical	Empirical feature selection based on the forecasting model characteristics (information criterion, RMSE, permutation feature importance, etc.)	12	50
GA	Genetic algorithm with spatiotemporal links in a chromosome and the model performance is based on a fitness function	5	3
PSO	Particle swarm optimisation with spatiotemporal links in a candidate solution	2	2
PSO-GA	Combination of genetic algorithm and particle swarm optimisation	1	1
Embedded methods			
LASSO	Least absolute shrinkage and selection operator (L1-norm loss function) regularisation	7	3
MCP, SCAD	Maximum concave penalty regularisation Smoothly clipped absolute deviation regularisation	1	1
SRM	Structural Risk Minimisation	1	1
Regularised kernel	Regularised kernel function (i.e. Laplacian)	1	1
RBM	Restricted Boltzmann machine, usually as part of a deep learning network	2	2
Sparse AE	Sparse autoencoders, usually as part of a deep learning network	1	1
LSTM	Long short-term memory unit stores temporal information for either long or short time periods	–	5
Internal	Other methodology-specific regularisation	–	2
Dimension reduction			
Spatial clustering/ Temporal aggregation	Clustering of spatial locations using different methods (self-organising maps, empirical grouping, etc.) For temporal dimension – selection of an appropriate temporal aggregation level	12	1
PCA-EVD	Principal component analysis, based on eigenvalue decomposition	9	8

**Table 1** Spatiotemporal FSE methods (Continued)

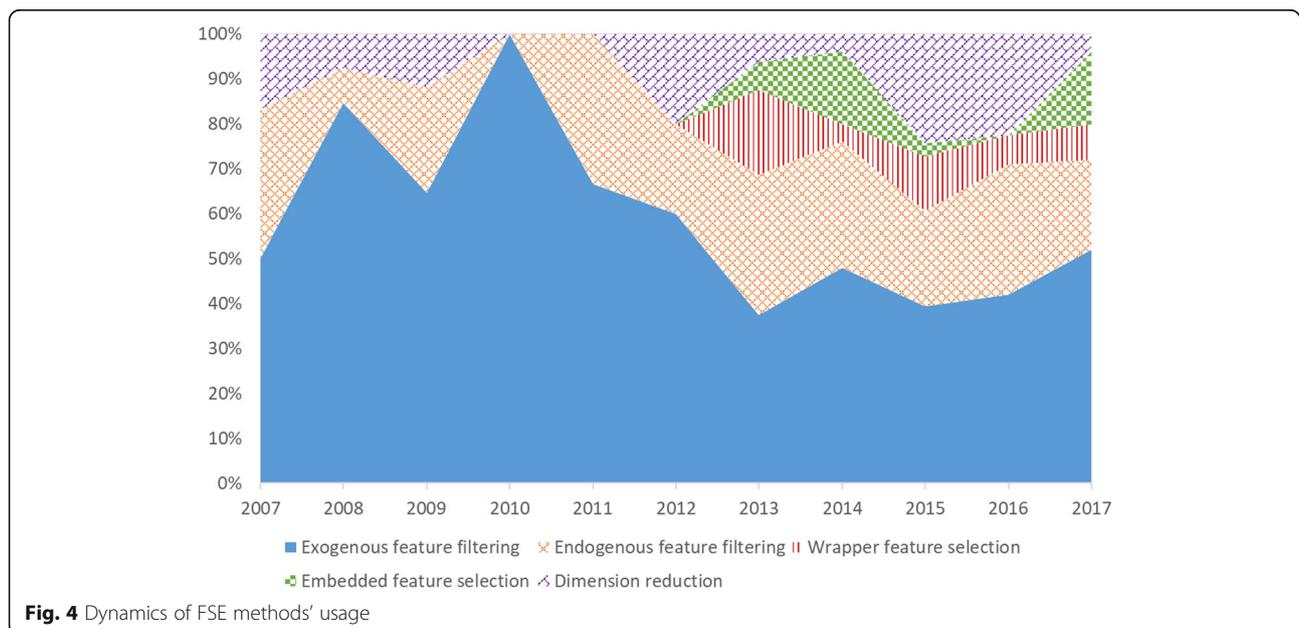
Method	Short description	Number of studies	
		Spatial	Temporal
PCA-SVD	Principal component analysis, based on singular-value decomposition	4	2
LSDA	Local shrunk discriminant analysis	1	1
NMF	Non-negative matrix factorisation	2	1
SSA	Singular spectrum analysis	3	3

segments and regular traffic conditions but is not always suitable for large networks. The effects of congestion in a segment could continue for 2–3 h, and thus, the required maximum time lag for moderately detailed 5-min time spans is quite large. If related spatial locations are predefined, the number of time lags can be limited by the travel time (to exclude excessively fast and slow effects). The former approach is utilised in six analysed studies.

**3.1.2 Class 2: Endogenous feature filtering methods**

In contrast to exogenous feature filtering methods, endogenous methods are based on information regarding traffic flow at different spatial locations. The most widely used statistical technique is based on correlation analysis and the cross-correlation function (CCF). The CCF returns correlation coefficients between traffic flows at different spatial locations with specified time lags and can be used for identification of spatiotemporal relationships. Note that CCF is not based on physical connectivity of the road network and thus it can discover potential relationships between remote spatial locations (e.g. simultaneous traffic flows from different directions to the city centre every morning or to a

stadium on match days). Authors in 21 studies utilise CCF for identification of both spatial and temporal relationships, six studies use it for temporal analysis and 11 studies for spatial dimensions. Application of the CCF function requires definition of a threshold value to exclude insignificant or weak spatial relationships. The formal Student’s test for insignificance of a correlation coefficient is not always appropriate, because this could lead to too many spatiotemporal links. Thus, many authors use a predefined threshold to reach a required level of STN sparsity (e.g. Li et al. [18] used a 0.94 value for the correlation coefficient). The modern graphical least absolute shrinkage and selection operator (LASSO) algorithm allows automatic identification of the most informative spatiotemporal links via estimation of the precision matrix (an inverse of the covariance matrix) based on  $l_1$ -regularisation. The graphical LASSO is applied in four studies [18–21] for filtering spatial relationships, but to the best of our knowledge, only Haworth and Cheng [20] applied it in both spatial and temporal dimensions simultaneously. The results of the graphical LASSO application are promising, but application of other forms of regularisation (i.e. the maximum concave penalty) is also recommended by authors [19].



**Fig. 4** Dynamics of FSE methods' usage

Application of the CCF function for non-stationary time series may lead to a well-known problem of spurious correlations and incorrect conclusions reached regarding significant spatial relationships. To overcome this problem, Hasan and Kim [22] and Pavlyuk [23] applied Granger causality tests for identification of spatiotemporal relationships.

Another endogenous feature filtering approach is based on a preliminary application of regularised regression models. Least-angle regression (LARS) is an  $l_1$ -norm-based algorithm that produces a full piecewise linear solution and excludes weak predictors. Recently, it was successfully applied to spatiotemporal FSE by Polson and Sokolov [24] and Yang et al. [25, 26]. Note that although the LARS algorithm and LASSO regularisation share the same principle, we distinguish them within this study based on the point of their application—outside of the model for LARS and within the model for LASSO. Thus, the LASSO approach will be separately discussed with other embedded feature selection methods.

Another technique, multivariate adaptive regression splines (MARS), also was successfully applied by Xu et al. [27, 28] and by Ye et al. [29]. Xu et al. [27] applied MARS to preliminary feature selection for the SVR model, while authors of other studies applied MARS directly to traffic flow forecasting.

Finally, several recent studies discover spatiotemporal relationships on the basis of special methods or indicators, designed by the authors to apply deeper analysis of traffic similarities. Dong et al. [7] constructed an indicator that simultaneously includes adjacency of spatial locations – the shortest distance and number of intersections between them; Zhu et al. [30] utilised similarity of traffic flows at different spatial locations; Cheng et al. [31] weighted the similarity by the distance between links; Deng and Jiang [32] suggested empirical association rules; Pascale and Nicoli [33] utilised a mutual information indicator; Chan et al. [34] applied the Taguchi method; Cai et al. [35] constructed an indicator using the distance and a connective grade of spatial locations and correlations for traffic flows; Wu et al. [36] suggested a custom bi-square function; Chen et al. [37] applied weighted traffic flows as a similarity metric.

### 3.1.3 Class 3: Wrapper feature selection methods

Wrapper feature selection methods are based on multiple evaluations of a forecasting model for selection of an optimal set of features. We consider traffic forecasting as the primary research problem; therefore, the natural key performance indicator is the model's forecasting accuracy. Root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) are the most widely used model performance indicators. All mentioned indicators estimate the in-sample forecasting accuracy and can lead to

incorrect preference of overfitted models with too many spatiotemporal relationships. Thus, many researchers penalise the model's complexity by applying information criteria (Akaike or Bayesian). This approach is applied in most studies based on statistical forecasting models (VAR, STARIMA, etc.). Another option is to apply a cross-validation procedure (e.g. rolling window analysis [38]) to estimate the out-of-sample model performance.

Given the performance indicator of a forecasting model and repeated model evaluations, researchers apply different techniques to find an optimal set of features. The majority of researchers (50 studies) identify an optimal number of time lags empirically (testing the forecasting model for different time lag values), and 12 studies utilised a similar technique for the spatial dimension (e.g. using empirical identification of an optimal number of upstream sensors [38, 39]). In addition, many researchers (24 studies) compared different exogenous and endogenous filtering methods (e.g. network-connectivity versus CCF-based approaches), which can be considered as a special case of empirical wrapper feature selection.

Many forecasting methodologies provide specific metrics to support a decision on feature exclusion. Statistical methodologies apply hypothesis testing routines (i.e. Student's test) for identifying significant features; neural networks allow estimation of elasticities of input components [40]; and random forests include permutation importance heuristics [41]. Using these metrics, researchers can refine the feature set of the forecasting model.

High computational complexity is a well-known problem in wrapper feature selection methods, which is widely solved by application of heuristic algorithms, such as particle swarm optimisation (PSO) and genetic algorithms (GA). Abdulhai et al. [42, 43] suggested application of GA for selection of an optimal number of upstream and downstream spatial locations (as well as for other parameters of their neural network-based forecasting model). Recently GA were applied for spatial [44–46] and temporal [47] feature selection. The PSO approach was applied by Chan et al. [48, 49] and recently combined with GA by Zheng et al. [50].

### 3.1.4 Class 4: Embedded feature selection methods

Embedded methods incorporate feature selection as part of a forecasting model's training process. The LASSO approach is the most widely used in spatiotemporal traffic forecasting (seven studies) and is based on the  $l_1$ -norm of spatiotemporal links:

$$l_1 = \sum_{i=1}^n \sum_{t=1}^{T-1} |\theta_{i,t}|$$

The  $l_1$ -norm in the Lagrangian form is included in the objective function and ensures meaningful feature

selection. Kamarianakis et al. [51] applied LASSO to vector autoregressive models; Piatkowski et al. [52] utilised LASSO and elastic net techniques to construct a graphical (random field) model; Li et al. [53] and Zhou et al. [54] executed preliminary feature selection and constructed LASSO-regularised autoregressive distributed lag models. Haworth and Cheng [55] analysed alternative regularisation techniques, maximum concave penalty (MCP) and smoothly clipped absolute deviation (SCAD), and found MCP beneficial with respect to the estimated STN sparsity.

Long short-term memory (LSTM) units are used in recurrent neural networks for automatic selection of an appropriate temporal memory structure. Such units are widely used for forecasting of time series with unknown duration of time lags between important events and have been effectively applied in several recent studies involving traffic flow [9, 11, 24, 56, 57].

Modern deep learning approaches allow a feature selection mechanism to be embedded into the multi-layer neural network architecture. Huang et al. [58] and Niu et al. [59] applied restricted Boltzmann machines as deep architecture components responsible for feature selection. Alternatively, Lv et al. [60] included sparse autoencoders that enforce encoding of the original set of spatiotemporal links into a smaller set of features (this approach works similar to dimension reduction methods, described below).

### 3.1.5 Class 5: Dimension reduction methods

The feature selection methods described thus far are based on identification of the most important spatiotemporal features (edges in the time-expanded graph). An alternative approach is to apply a dimension reduction technique to limit the number of time periods (layers) or spatial locations (vertices in the time expanded graph). The most widely used technique is spatial clustering and followed by application of a forecasting model to clusters. This technique is applied in 12 studies using different clustering methods. Examples of these methods include neural networks [61, 62], self-organising maps [63], k-means [64, 65], simulated annealing [66], and empirical spatial aggregation [67, 68].

Temporal aggregation is an issue widely addressed in time series forecasting. Although the importance of correct temporal aggregation is widely acknowledged for traffic forecasting [2], it has rarely been directly addressed in publications (recently, Fusco et al. [68] provided empirical evidence of temporal aggregation effects on forecasting accuracy).

Feature selection methods and spatial clustering consider STN identification as a step of forecasting. If STN identification is not a required research result, then standard feature extraction techniques (e.g. principal

component analysis based on eigenvalue decomposition (PCA-EVD)) can be applied to prepare composed inputs for an efficient predictor. Such composed inputs do not represent the STN structure, but do include its most important aspects. PCA-EVD was used in nine studies as a preliminary step for different forecasting models: neural networks [69, 70], support vector regression [71–75], Bayesian networks [76], and random forests [77]. PCA-EVD has also been used as a method for tensor decomposition [78]. In most studies, PCA-EVD was applied for both temporal and spatial dimensions simultaneously.

Amongst other dimension reduction methods, we note applications of PCA based on singular-value decomposition [74, 79–81], non-negative matrix factorisation [82, 83], local shrunk discriminant analysis [84], and singular spectrum analysis [79, 85, 86].

### 3.2 Review of forecasting methodologies and their coverage by FSE methods

The range of utilised spatiotemporal methodologies is large and exceeds 30 methodologies, even after grouping variants of the same methodology. Table 2 summarises the methodologies, their modifications and their coverage by FSE methods. The methodologies are divided into two classes – artificial neural networks (ANN) and statistical models; this classification is conventional and is based on the philosophy and primary goals of modelling (statistical models focus on the structure of relationships amongst inputs and outputs; whereas, ANN are usually used to provide an efficient prediction by learning complex relationships).

A detailed discussion of the presented methodologies, their advantages and shortcomings, lies outside of this review's scope; therefore, we pay limited attention to the dynamics of the different approaches' applications and primarily examine their coverage by FSE methods. The dynamics of utilised spatiotemporal traffic forecasting methodologies are presented in Fig. 5 (data is grouped in two-year periods for better trend representation).

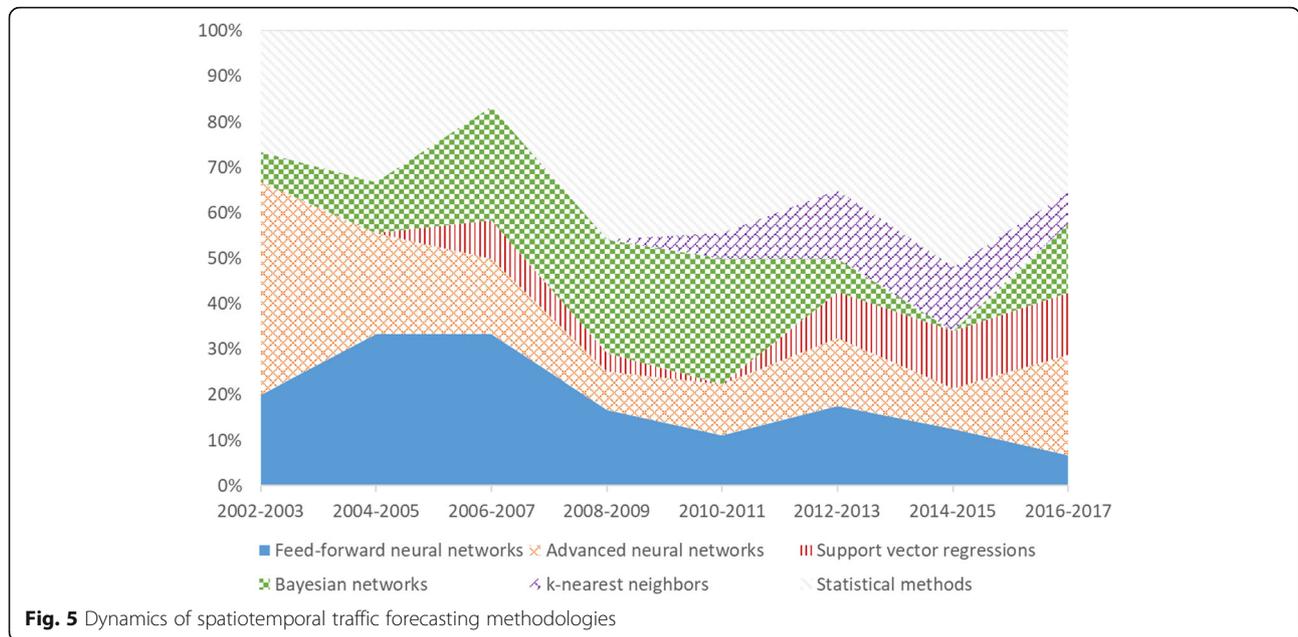
First, we note a considerable reduction of feed-forward neural network (FFNN) applications in the spatiotemporal domain (from more than 30% of studies in 2004–2007 to less than 10% in 2017). This reduction is partly explained by the replacement of FFNN with more advanced neural network architectures (recurrent neural networks, time-delayed neural networks, and, recently, by deep learning techniques). Advances of neural networks widely related to the FSE problem are recurrent, time-delayed, LSTM, and other ANN that include embedded mechanisms for automated feature selection. The fact that such mechanisms directly improve the performance of a pure FFNN (with complicated FSE and the related curse of dimensionality) is supported by the mentioned studies. Second, we note a significant growth

**Table 2** Coverage of traffic forecasting methodologies by FSE methods

Model	Short description	Total number of applications	Number of applications of FSE methods' class				
			Exogenous feature filtering	Endogenous feature filtering	Wrapper feature selection	Embedded feature selection	Dimension reduction
Artificial neural networks (ANN)							
FFNN	Feedforward ANN	41	34	4	9	–	6
TDNN	Time-delayed ANN	9	13	–	2	–	2
RNN	Recurrent ANN	9	13	2	–	–	2
LSTM	Long short-term memory ANN	6	3	2	–	–	1
SSNN	State-space ANN, incl. Time-delayed state-space ANN (STDNN)	4	4	2	–	–	–
CNN	Convolutional ANN	4	4	–	–	–	–
DBN	Deep belief network, incl. Restricted Boltzmann machine (RBM), stacked autoencoders (SAE), generative adversarial networks (GAN)	4	1	–	–	4	–
NARX	Nonlinear autoregressive exogenous ANN	3	3	–	–	–	–
Other NN architectures	Incl. counter-propagation ANN (CPNN), fuzzy ANN, Graph ANN, general regression ANN, group method of data handling (GMDH)	8	4	1	2	–	1
Statistical models							
BN	Bayesian networks, incl. Conditional random fields	30	15	8	2	1	6
DL/ARDL	Distributed lags /autoregressive distributed lag models, incl. Smoothing models, chaos models	26	19	8	2	2	1
SVR	Support vector regression, incl. Extreme learning machine	23	8	2	4	2	11
KNN	k-nearest neighbour regression	22	15	2	4	0	3
VAR	Vector autoregressive models	21	22	5	–	2	–
STARIMA	Space-time autoregressive integrated moving average, incl. Generalised STARIMA	21	20	13	–	–	1
Kernel	Kernel regressions, incl. Gaussian process regression (GPR)	10	7	4	–	–	–
State-space	State-space models	10	8	–	2	–	–
Tensor models	Tensor completion models, incl. Probabilistic principal component analysis (PPCA)	6	–	1	–	–	6
Decision tree models	Incl. random forest and regression tree	5	4	2	–	1	1
SCTM	Stochastic cell transmission model	4	3	2	–	–	–
MARS	Multivariate adaptive regression splines, incl. Generalised additive model (GAM)	3	1	–	–	2	–
Spatial panel	Spatial panel models	2	1	2	–	–	–

of non-parametric statistical methods (especially k-nearest neighbour regression, support vector regression, and Bayesian networks). Third, multivariate parametric statistical methods (VAR, STARIMA) also exhibit growth in popularity. In our opinion, these trends are at least partly related to advances in FSE methods. Different approaches to FSE, discussed in the previous section, allow application

of modern statistical methodologies to forecasting of traffic flows in large, highly interconnected urban road networks. In combination with the high flexibility of non-parametric approaches, this leads to the observed growth of statistical methodologies' popularity in scientific literature. Note that the observed popularity of methodologies is not directly related to the best forecasting



accuracy. Recently, the requirements for traffic forecasting methodologies have shifted from forecast accuracy to identification of causality. Thus, methodologies that allow easier interpretation of the results and identification of the underlying STN present an advantage in this regard.

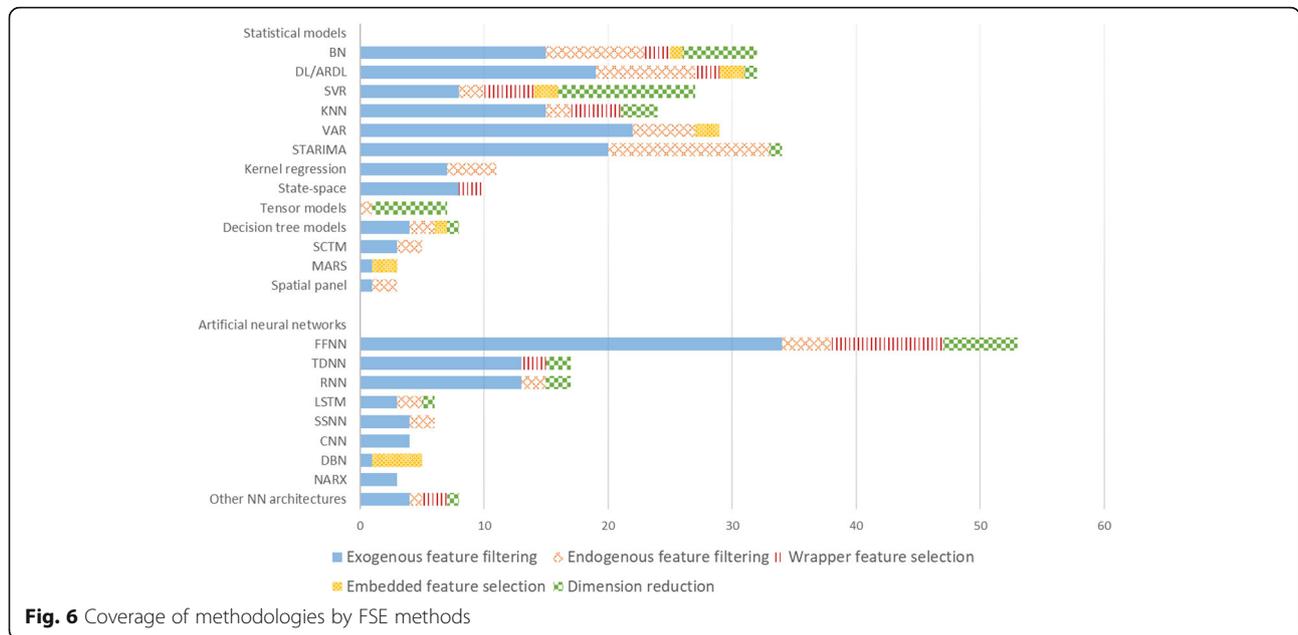
Another trend in the scientific literature is growing attention to the comparison of spatiotemporal methodologies of different classes. Early studies compared spatiotemporal specifications of a selected model with non-spatial baseline models. Vlahogianni et al. [46] were the first to compare the spatiotemporal FFNN with the spatiotemporal statistical (state-space) model. The number of studies with such comparisons was limited to eight studies until 2015, but during the last 3 years, 14 of 67 studies (21%) directly compare spatiotemporal models of different classes. Nevertheless, such comparisons were executed for different case studies (road network segments) and the findings are contradictory. Preferred spatiotemporal FSE is naturally a function of a selected methodology, topology and size of the road network, temporal resolution of traffic data, forecasting horizon, and other methodological issues, and identification of this function in the form of guidelines appears to be impossible based on the limited existing evidence. Development of a framework for the careful comparison of different methodologies (similar to the famous M-competitions [87]) seems extremely important for further methodological development of spatiotemporal traffic flow forecasting.

Coverage of methodologies by different FSE methodologies is not uniform. Figure 6 presents the distribution of different FSE methods over the set of methodologies.

The diagram presents a wide range of uncovered areas that can be considered as potential research directions. Note that not all weakly covered areas make sense or would be considered fruitful for future studies; therefore, we primarily note a lack of general guidelines for selecting spatiotemporal FSE methods.

Exogenous feature filtering is the most widely used approach in almost all forecasting methodologies, except in the SVR, DBN and tensor decomposition models. The use of other FSE methods for DBN and tensor decomposition models is naturally explained by their structure, but the significant number of SVR applications with non-exogenous FSE can be speculatively explained by the significant improvement of empirical results obtained by applying FSE methods from other groups. This conclusion is also supported by the growing total share of non-exogenous FSE, as presented in Fig. 4.

Statistical methodologies are better covered by different FSE methods; whereas, there is a lack of such applications for ANNs. ANNs are, especially, weakly covered by endogenous feature filtering methods (11 studies for ANN versus 49 studies for statistical models). Partly this fact is explained by the “black box” approach that is natural for ANN structures based on an implicit FSE in the ANN training process. This approach has evident shortcomings, especially taking into account that the goal of modern forecasting models is not limited to the forecasted values themselves, but also includes revealing casual relationships. This statement is empirically supported by development of deep learning architectures that explicitly contain FSE mechanisms (e.g. in the form of restricted Boltzmann machines or autoencoders, as described in the previous section).



**Fig. 6** Coverage of methodologies by FSE methods

In contrast, wrapper feature selection methods are more frequently used in ANN than in statistical methodologies. Application of evolutionary algorithms for generating ANN (neuro-evolution) is an emerging methodological trend, but it appears that GA application with statistical methods is an under-researched area in the spatiotemporal traffic forecasting field. In particular, to the best of our knowledge, there are no applications of wrapper feature selection for the popular VAR and STARIMA models.

Applications of dimension reduction methods are distributed more uniformly amongst methodologies, with the only notable exception being SVR. There are several applications where SVR is combined with clustering or PCA-based dimension reduction; therefore, SVR has received the highest coverage by the various feature selection methods generally.

Finally, we note that there is a lack of systematic empirical research on FSE methods in spatiotemporal forecasting models. In summary, 80% (170 studies) consider only one approach to spatiotemporal feature selection, 7% (15 studies) apply several methods within the same class (e.g. different dimension reduction techniques), and 10% (20 studies) compare a pair of selected exogenous and endogenous methods (e.g. CCF versus upstream/downstream connectivity). Amongst the remaining six studies, Hu et al. [63] combined a clustering technique using self-organising maps and their physical connectivity in an FFNN predictor; similarly, Lu et al. [66] consequently applied clustering of spatial locations and CCF-based feature selection; Niu et al. [59] and Tan et al. [79] used CCF for preliminary feature filtering, and RBM and SVD (respectively) for second-stage feature selection; Gebresilassie [72] compared linear regression features with exogenously

selected and PCA-generated features; and Schimbinschi et al. [88] combined road connectivity and CCF-based feature selection with structural risk minimisation regularisation (embedded feature selection). Taking into account a very limited number of studies that compare different FSE methods and potential combinations of methods from different classes, we conclude that this represents an extensive uncovered area for further research.

### 3.3 Spatiotemporal FSE applied in related areas

This literature review is limited to spatiotemporal FSE methods that have already been applied to urban traffic forecasting. However, there are several other areas where spatiotemporal modelling is widely used and where the problem of spatiotemporal FSE is emerging. Namely,

- Energy and electricity systems, e.g. solar and wind energy. Spatiotemporal solar forecasting models use spatially distributed solar radiation power data to enhance forecasting at a given site [89], and wind speed and power forecasting are widely used for wind turbine placement and supply planning [90]. Similar to traffic models, solar and wind power production spatiotemporal data are usually discretised in space and time (obtained in temporarily aggregated form from a discrete number of spatially distributed sensors). Similar data structures lead to similar methodological issues and solutions, including the problem of spatiotemporal FSE. Many of the methodologies discussed in this review have also been applied or could be applied to energy system forecasting [91].

- Image and video processing. Similar to traffic flow, a video stream can be considered as spatiotemporal data (a temporal sequence of two-dimensional frames), and thus, the problem of learning its internal relationships is very similar to spatiotemporal FSE for traffic flow. The problem of forecasting in this case takes the form of video inpainting (reconstructing lost or deteriorated parts of a video stream) or motion detection and prediction (e.g. computer vision). To the best of our knowledge, most popular methods of spatiotemporal FSE for video processing belong to embedded feature selection, as categorised in this review (e.g. LASSO and LARS regularisation) [92]. There are also several specific methods such as sparse dictionary learning [93] applied in video processing that are rarely used for traffic forecasting. Adopting these methods for spatiotemporal traffic forecasting is possibly a promising research direction.

Other application areas where spatiotemporal models play a crucial role are atmospheric and hydrological sciences (e.g. meteorology, climatology and ecology). Dynamic models of flow (e.g. kinematic waves), inherited from atmospheric sciences, are widely adopted for traffic forecasting. Spatiotemporal relationships in these models are presented in the form of partial differential equations and usually are not considered as stochastic. Thus, although the methods are promising, we do not include them within the scope of this review.

To the best of our knowledge, there are no published literature reviews on spatiotemporal FSE involving multiple areas/disciplines. Merging of methodologies and experience from different applied areas is an important but extensive research direction.

### 3.4 Selecting an approach to spatiotemporal structure identification

The choice of an appropriate method for identification and weighting of spatial and spatiotemporal relationships is a critical requirement for urban traffic forecasting. To the best of our knowledge, there are no methodologies or guidelines for solving this problem. A list of bibliographic sources, covered by this review, contains a very limited number of research studies where different approaches to identify spatiotemporal relationships were compared and proper conclusions regarding their applicability were made. Thus, development of guidelines for spatiotemporal FSE is an important advantage that could not be properly accomplished on the basis of our literature review. The best result is noting the actual method choice made by researchers, and assuming that this choice is well-grounded and optimal for the analysed spatial settings (which in general may not be true).

To discover clues for preferred spatiotemporal FSE methods, we clustered all bibliographic sources on the basis of three variables – utilised spatiotemporal model, analysed road topology (sequential freeway, sequential arterial road or network), and size of the selected road network fragment (number of spatial links). Results of the clustering are presented in Table 3 and illustrated in Fig. 7.

We clustered application evidence of different spatiotemporal models; therefore, if a bibliographic source contains results for several models, we consider them as separate observations (393 spatiotemporal models in 211 sources). The number of clusters (three) was selected on the basis of the average silhouette width, and clustering was performed by the conventional k-means algorithm with Gower's distance-based similarity. The overall internal clustering quality is good (average silhouette width = 0.517) and formed clusters could be conventionally referred to as:

- Cluster 1: Statistical models for a complex network topology of medium size
- Cluster 2: ANN for freeways with small number of links
- Cluster 3: Various models for arterial roads with small number of links

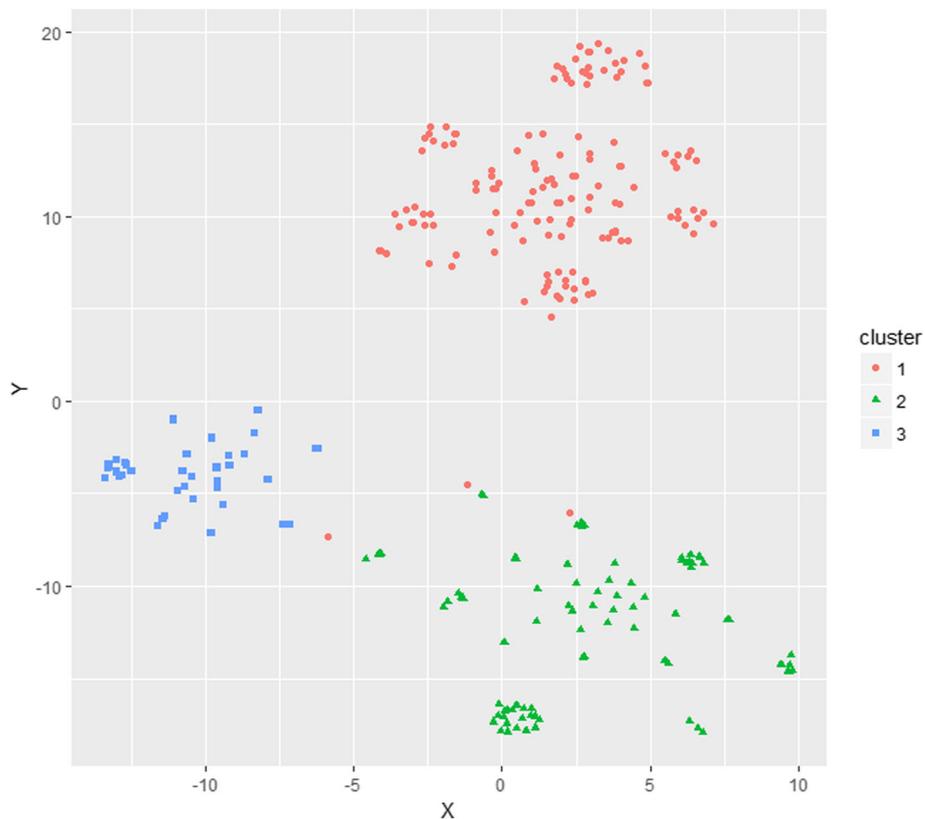
Research studies in Cluster 1 utilise endogenous spatiotemporal FSE more often – the most popular approach is based on cross-correlation functions. In addition, dimension reduction methods are widely used in this cluster (PCA is the most popular). Cluster 2 and Cluster 3 are homogeneous in terms of selected spatiotemporal FSE methods and are mainly based on exogenous filtering (e.g. inclusion of directly connected upstream points). Taking into account that Cluster 1 studies are newer (median year of publication is 2015), we can conclude that conventional exogenous spatiotemporal FSE worked well for sequential spatial settings (freeways and arterial roads) with a small number of analysed locations. Recently the focus of spatiotemporal traffic forecasting has shifted to complex road networks, where endogenous and other spatiotemporal FSE methods are more beneficial.

In addition to observations from clustering analysis, we attempted to apply a classifier (decision tree-based) to discover principles or rules for selecting spatiotemporal FSE methods. The estimated accuracy of classification was extremely low, which lead us to the conclusion regarding the absence of straightforward principles available from the literature.

Summarising the analysis above, we conclude that there is a lack of attention to determining the proper choice of spatiotemporal FSE methods in literature on

**Table 3** Results of bibliographic source clustering

Cluster	1	2	3	
Average silhouette width	0,517	0,512	0,530	
Cluster size	176	143	74	
Clustering variables				
Spatiotemporal model	Top 1	STARIMA	FFNN	FFNN
	Top 2	BN	VAR	KNN
	Top 3	SVR	TDNN	DL/ARDL
	Top 4	VAR	RNN	SVR
Topological structure	Top 1	Network	Sequential freeway	Sequential arterial
Number of links	Median	26	7	4
Conventionally referred as		Statistical models for a network topology of medium size	ANN for freeways with small number of links	Various models for arterial roads with small number of links
Target variables				
Year	Median	2015	2011	2014
Spatial FSE	Top 1	CCF	Up/downstream	Up/downstream
	Top 2	Upstream	Upstream	All
	Top 3	PCA	All	Upstream
Temporal FSE	Top 1	Predefined	Predefined	Predefined
	Top 2	CCF	Empirical	Empirical
	Top 3	Empirical	CCF	CCF



**Fig. 7** Illustration of bibliographic source clustering

urban traffic forecasting, which highlights the necessity for empirical studies in this direction to develop comprehensive guidelines for selecting the appropriate spatiotemporal FSE method(s).

#### 4 Conclusions

Spatiotemporal traffic forecasting is an emerging field in the scientific literature, and correct identification of the spatiotemporal structure plays an important role in this research area. Feature selection and extraction methods allow revealing of spatiotemporal relationships and improving the forecasting accuracy and robustness of modern forecasting methodologies. The present paper systematically reviews a broad range of traffic flow forecasting literature (211 publications) regarding utilised spatiotemporal methodologies and applied feature selection and extraction methods. The key findings and conclusions of the review are as follows:

- (1) Spatiotemporal approaches that utilise both spatial and temporal relationships are gaining scientific interest in the field of traffic flow forecasting. The annual number of related publications has doubled during the past decade and is expected to continue to grow.
- (2) Definition of the spatiotemporal structure of traffic flow should not be limited to physical road network connectivity, but should also include relationships that are distant in space and time. Thus, the role of data-driven feature selection and extraction methods becomes more important in empirical studies.
- (3) Feature selection and extraction methods can be conventionally divided into five classes (exogenous and endogenous feature filtering, wrapper feature selection and embedded feature selection methods and dimension reduction methods). We analysed the dynamics of method applications from different classes in the field of spatiotemporal traffic forecasting and concluded that the general trend has recently shifted from exogenous feature filtering to a variety of data-driven feature selection methods.
- (4) During the past 15 years, the trend of applied spatiotemporal methodologies has gradually shifted from ANN to multivariate parametric and non-parametric statistical methods. We believe that this shift is partly related to development of advanced feature selection and extraction methods, which improve statistical model estimation for large data sets. At the same time, we note a growing number of deep learning applications in 2017–2018 that use embedded mechanisms for feature extraction.
- (5) Another trend in the empirical literature is a growing focus on comparing spatiotemporal methodologies of different classes (ANN, parametric and non-parametric statistical methods). This type of comparison was rarely performed in earlier studies; whereas, over the last three years, 21% of studies directly compare spatiotemporal models of different classes.
- (6) The effectiveness of spatiotemporal forecasting methodologies is difficult to compare on the basis of the existing literature. Most studies are based on a selected case study (a small road network segment) and results involving executed methodology comparisons remain study-specific (and are often contradictory). Development of a framework for comparison of different methodologies (similar to the famous M-competitions) is highly recommended for further methodological development of spatiotemporal traffic flow forecasting.
- (7) Coverage of forecasting methodologies by feature selection methods is not uniform. Several methodologies (i.e. SVR) have been intensively tested with different feature selection approaches; whereas, several others (i.e. VAR) have not been widely analysed. In addition, the majority of publications are limited to the application of a single approach for feature selection and there is a lack of studies based on combining different feature selection methods. These findings point to a broad direction for future research.
- (8) Insufficient attention has been paid to a proper choice of spatiotemporal FSE methods in literature on urban traffic forecasting. We conclude that there is a need for additional empirical studies in this direction to develop comprehensive guidelines for selecting appropriate spatiotemporal FSE methods.

The added value of this review includes the trends discovered in the methodology of spatiotemporal traffic forecasting and empirical insights into applied feature selection methods. The list of 211 studies, classified by the applied methodology and spatial and temporal feature selection and extraction methods is a self-contained contribution to assist further literature analyses in this field. Systematically reviewing the scientific literature, we discovered several important methodological and empirical gaps and have suggested directions for future research.

#### 5 Additional files

**Additional file 1:** Summary of bibliographic sources. (DOCX 129 kb)

**Additional file 2:** Bibliographic database. (XLSX 24 kb)

**Additional file 3:** R script of the executed analyses. (R 4 kb)

### Acknowledgements

Not applicable.

### Funding

The author was financially supported by the specific support objective activity 1.1.1.2. "Post-doctoral Research Aid" (Project id. N. 1.1.1.2/16/1/001) of the Republic of Latvia, funded by the European Regional Development Fund. Dmitry Pavlyuk's research project No. 1.1.1.2/VIAA/1/16/112 "Spatiotemporal urban traffic modelling using big data".

### Availability of data and materials

The bibliographic database that was generated and analysed during the current study is available in the Zotero repository, <http://bit.ly/spatiotemporalFSE>, and in the Additional file 1. In addition, the same bibliographic database is provided in spreadsheet format (Additional file 2) and supplemented with an R script of the executed analyses (Additional file 3).

### Author's contribution

DP executed analysis of bibliographic sources and prepared the final manuscript. The author read and approved the final manuscript.

### Competing interests

The author declares that he has no competing interests.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 29 August 2018 Accepted: 9 January 2019

Published online: 24 January 2019

### References

- Vlahogianni, E. I., Golias, J. C., & Karlaftis, M. G. (2004). Short-term traffic forecasting: Overview of objectives and methods. *Transport Reviews*, 24, 533–557 <https://doi.org/10.1080/0144164042000195072>.
- Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, 43, 3–19 <https://doi.org/10.1016/j.trc.2014.01.005>.
- Haworth, J. (2014). *Spatio-temporal forecasting of network data*. London: PhD diss., University College London.
- Ermagun, A., & Levinson, D. (2018). Spatiotemporal traffic forecasting: Review and proposed directions. *Transport Reviews*, 1–29. <https://doi.org/10.1080/01441647.2018.1442887>.
- George, B., & Kim, S. (2013). *Spatio-temporal networks*. New York: Springer New York.
- Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, 40, 16–28 <https://doi.org/10.1016/j.compeleceng.2013.11.024>.
- Dong, C., Shao, C., & Li, X. (2009). Short-Term Traffic Flow Forecasting of Road Network Based on Spatial-Temporal Characteristics of Traffic Flow. In: *Proceedings of the 2009 WRI World Congress on Computer Science and Information Engineering* (pp. 645–650). Los Angeles: IEEE.
- Cao, Q., Ren, G., & Li, D. (2018). Multiple Spatio-temporal scales traffic forecasting based on deep learning approach. In *Compendium of papers of the Transportation Research Board 97th annual meeting* (p. 18). Washington: Transportation Research Board.
- Du, S., Li, T., Gong, X., et al. (2017). Traffic flow forecasting based on hybrid deep learning framework. In *Proceedings of the 12th international conference on intelligent systems and knowledge engineering (ISKE)* (p. 6). Shanghai: IEEE.
- Ma, X., Dai, Z., He, Z., et al. (2017). Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction. *Sensors*, 17, 818 <https://doi.org/10.3390/s17040818>.
- Yu, H., Wu, Z., Wang, S., et al. (2017). Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. *Sensors*, 17, 1501 <https://doi.org/10.3390/s17071501>.
- Min, W., & Wynter, L. (2011). Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C: Emerging Technologies*, 19, 606–616 <https://doi.org/10.1016/j.trc.2010.10.002>.
- Ermagun, A. (2016). *Network Econometrics and Traffic Flow Analysis*. Minneapolis and Saint Paul, Minnesota: PhD diss., University of Minnesota.
- Ermagun, A., & Levinson, D. (2018). Spatio-temporal short-term traffic forecasting using the network weight matrix and systematic Detrending. In *Compendium of papers of Transportation Research Board 97th annual meeting* (p. 14). Washington: Transportation Research Board.
- Ermagun, A., & Levinson, D. M. (2018). Development and application of the network weight matrix to predict traffic flow for congested and uncongested conditions. *Environment and Planning B: Urban Analytics and City Science*, 239980831876336 <https://doi.org/10.1177/2399808318763368>.
- Dimitriou, L., Tsekeris, T., & Stathopoulos, A. (2008). Adaptive hybrid fuzzy rule-based system approach for modeling and predicting urban traffic flow. *Transportation Research Part C: Emerging Technologies*, 16, 554–573 <https://doi.org/10.1016/j.trc.2007.11.003>.
- Stathopoulos, A., Dimitriou, L., & Tsekeris, T. (2008). Fuzzy modeling approach for combined forecasting of urban traffic flow. *Computer-Aided Civil and Infrastructure Engineering*, 23, 521–535 <https://doi.org/10.1111/j.1467-8667.2008.00558.x>.
- Li, Z., Jiang, S., Li, L., & Li, Y. (2017). Building sparse models for traffic flow prediction: An empirical comparison between statistical heuristics and geometric heuristics for Bayesian network approaches. *Transportmetrica B: Transport Dynamics*, 1–17 <https://doi.org/10.1080/21680566.2017.1354737>.
- Hara, Y., Suzuki, J., & Kuwahara, M. (2018). Network-wide traffic state estimation using a mixture Gaussian graphical model and graphical lasso. *Transportation Research Part C: Emerging Technologies*, 86, 622–638 <https://doi.org/10.1016/j.trc.2017.12.007>.
- Haworth, J., & Cheng, T. (2014). Graphical LASSO for local spatio-temporal neighbourhood selection. In *Proceedings the GIS research UK 22nd annual conference* (pp. 425–433). Glasgow: University of Glasgow.
- Sun, S., Huang, R., & Gao, Y. (2012). Network-scale traffic modeling and forecasting with graphical lasso and neural networks. *Journal of Transportation Engineering*, 138, 1358–1367 [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000435](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000435).
- Hasan, M. M., & Kim, J. (2016). Analysing functional connectivity and causal dependence in road traffic networks with granger causality. In *Australasian transport research forum 2016 Proceedings* (p. 19). Melbourne: Australasian Transport Research Forum Incorporated.
- Pavlyuk, D. (2018). On Application of Regime-Switching Models for Short-Term Traffic Flow Forecasting. In W. Zamojski, J. Mazurkiewicz, J. Sugier, et al. (Eds.), *Proceedings of the Twelfth International Conference on Dependability and Complex Systems DepCoS-RELCOMEX* (pp. 340–349). Brunow: Springer International Publishing.
- Polson, N. G., & Sokolov, V. O. (2017). Deep learning for short-term traffic flow prediction. *Transportation Research Part C: Emerging Technologies*, 79, 1–17.
- Yang, S., Shi, S., Hu, X., & Wang, M. (2015). Discovering spatial contexts for traffic flow prediction with sparse representation based variable selection. In *Proceedings of Intl Conf on ubiquitous intelligence and computing and 12th Intl Conf on autonomic and trusted computing and 15th Intl Conf on scalable computing and communications and its associated workshops (UIC-ATC-ScalCom)* (pp. 364–367). Beijing: IEEE.
- Yang, S., Shi, S., Hu, X., & Wang, M. (2015). Spatiotemporal context awareness for urban traffic modeling and prediction: Sparse representation based variable selection. *PLoS One*, 10(22) <https://doi.org/10.1371/journal.pone.0141223>.
- Xu, Y., Chen, H., Kong, Q.-J., et al. (2016). Urban traffic flow prediction: A spatio-temporal variable selection-based approach. *Journal of Advanced Transportation*, 50, 489–506 <https://doi.org/10.1002/atr.1356>.
- Xu, Y., Kong, Q.-J., & Liu, Y. (2013). A spatio-temporal multivariate adaptive regression splines approach for short-term freeway traffic volume prediction. In *Proceedings of the 2013 16th International IEEE Conference on Intelligent Transportation Systems - (ITSC)* (pp. 217–222). The Hague: IEEE.
- Ye, S., He, Y., Hu, J., & Zhang, Z. (2008). Short-term traffic flow forecasting based on MARS. In *Proceedings of the Fifth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)* (pp. 669–675). Shandong: IEEE.
- Zhu, T., Kong, X., & Lv, W. (2009). Large-scale travel time prediction for urban arterial roads based on Kalman filter. In *Proceedings of the International Conference on Computational Intelligence and Software Engineering (CISE)* (pp. 1–5). Wuhan: IEEE.

31. Cheng, T., Wang, J., Haworth, J., et al. (2011). Modelling dynamic space-time autocorrelations of urban transport network. In *Proceedings of the 11th international conference on Geocomputation 2011* (pp. 215–220). London: University College London
32. Deng, R., & Jiang, L. (2011). Traffic state forecast of road network based on spatial-temporal data mining. In *Proceedings of the Third International Conference on Transportation Engineering* (pp. 734–739). Chengdu: American Society of Civil Engineers
33. Pascale, A., & Nicoli, M. (2011). Adaptive Bayesian network for traffic flow prediction. In *Proceedings of the 2011 IEEE Statistical Signal Processing Workshop (SSP)* (pp. 177–180). Nice: IEEE.
34. Chan, K. Y., Khadem, S., Dillon, T. S., et al. (2012). Selection of significant on-road sensor data for short-term traffic flow forecasting using the Taguchi method. *IEEE Transactions on Industrial Informatics*, 8, 255–266 <https://doi.org/10.1109/TII.2011.2179052>.
35. Cai, P., Wang, Y., Lu, G., et al. (2016). A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting. *Transportation Research Part C: Emerging Technologies*, 62, 21–34 <https://doi.org/10.1016/j.trc.2015.11.002>.
36. Wu, Y.-J., Chen, F., Lu, C.-T., & Yang, S. (2016). Urban traffic flow prediction using a Spatio-temporal random effects model. *Journal of Intelligent Transportation Systems*, 20, 282–293 <https://doi.org/10.1080/15472450.2015.1072050>.
37. Chen, J., Li, D., Zhang, G., & Zhang, X. (2018). Localized space-time autoregressive parameters estimation for traffic flow prediction in urban road networks. *Applied Sciences*, 8, 277 <https://doi.org/10.3390/app8020277>.
38. Schimbinschi, F., Nguyen, X. V., Bailey, J., et al. (2015). Traffic forecasting in complex urban networks: Leveraging big data and machine learning. In *Proceedings of the 2015 IEEE International Conference on Big Data* (pp. 1019–1024). Santa Clara: IEEE.
39. Ratrou, N. T. (2014). Short-term traffic flow prediction using group method data handling (GMDH)-based additive networks. *Arabian Journal for Science and Engineering*, 39, 631–646 <https://doi.org/10.1007/s13369-013-0681-3>.
40. Dougherty, M. S., & Cobbett, M. R. (1997). Short-term inter-urban traffic forecasts using neural networks. *International journal of forecasting*, 13, 21–31.
41. Ou, J., Xia, J., Wu, Y.-J., & Rao, W. (2017). Short-term traffic flow forecasting for urban roads using data-driven feature selection strategy and Bias-corrected random forests. *Transportation Research Record: Journal of the Transportation Research Board*, 2645, 157–167 <https://doi.org/10.3141/2645-17>.
42. Abdulhai, B., Porwal, H., & Recker, W. (1999). *Short term freeway traffic flow prediction using genetically-optimized time-delay-based neural networks*. Berkeley: University of California.
43. Abdulhai, B., Porwal, H., & Recker, W. (2002). Short-term traffic flow prediction using neuro-genetic algorithms. *ITS Journal-Intelligent Transportation Systems Journal*, 7, 3–41 <https://doi.org/10.1080/1713930748>.
44. Basyoni, Y., Abbas, H. M., Talaat, H., & El Dimeery, I. (2017). Speed prediction from mobile sensors using cellular phone-based traffic data. *IET Intelligent Transport Systems*, 11, 387–396 <https://doi.org/10.1049/iet-its.2016.0279>.
45. Chen, X., Wei, Z., Liu, X., et al. (2017). Spatiotemporal variable and parameter selection using sparse hybrid genetic algorithm for traffic flow forecasting. *International Journal of Distributed Sensor Networks*, 13(14) <https://doi.org/10.1177/155014771713376>.
46. Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. (2005). Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach. *Transportation Research Part C: Emerging Technologies*, 13, 211–234 <https://doi.org/10.1016/j.trc.2005.04.007>.
47. Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. (2007). Spatio-temporal short-term urban traffic volume forecasting using genetically optimized modular networks. *Computer-Aided Civil and Infrastructure Engineering*, 22, 317–325 <https://doi.org/10.1111/j.1467-8667.2007.00488.x>.
48. Chan, K. Y., Dillon, T., Chang, E., & Singh, J. (2013). Prediction of short-term traffic variables using intelligent swarm-based neural networks. *IEEE Transactions on Control Systems Technology*, 21, 263–274 <https://doi.org/10.1109/TCST.2011.2180386>.
49. Chan, K. Y., Dillon, T. S., & Chang, E. (2013). An intelligent particle swarm optimization for short-term traffic flow forecasting using on-road sensor systems. *IEEE Transactions on Industrial Electronics*, 60, 4714–4725 <https://doi.org/10.1109/TIE.2012.2213556>.
50. Zheng, L., Zhu, C., Zhu, N., et al. (2018). A feature selection based approach for urban short-term travel speed prediction. *IET Intelligent Transport Systems*, 16 <https://doi.org/10.1049/iet-its.2017.0059>.
51. Kamarianakis, Y., Shen, W., & Wynter, L. (2012). Real-time road traffic forecasting using regime-switching space-time models and adaptive LASSO. *Applied Stochastic Models in Business and Industry*, 28, 297–315 <https://doi.org/10.1002/asmb.1937>.
52. Piatkowski, N., Lee, S., & Morik, K. (2013). Spatio-temporal random fields: Compressible representation and distributed estimation. *Machine Learning*, 93, 115–139 <https://doi.org/10.1007/s10994-013-5399-7>.
53. Li, L., Su, X., Wang, Y., et al. (2015). Robust causal dependence mining in big data network and its application to traffic flow predictions. *Transportation Research Part C: Emerging Technologies*, 58, 292–307 <https://doi.org/10.1016/j.trc.2015.03.003>.
54. Zhou, X., Hong, H., Xing, X., et al. (2017). Discovering spatio-temporal dependencies based on time-lag in intelligent transportation data. *Neurocomputing*, 259, 76–84 <https://doi.org/10.1016/j.neucom.2016.06.084>.
55. Haworth, J., & Cheng, T. (2014). A comparison of Neighbourhood selection techniques in Spatio-temporal forecasting models. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences XL-2*, 7–12 <https://doi.org/10.5194/isprsarchives-XL-2-7-2014>.
56. Liang, Y., Cui, Z., Tian, Y., et al. (2018). A deep generative adversarial architecture for network-wide spatial-temporal traffic state estimation. In *Compendium of papers of Transportation Research Board 97th annual meeting* (p. 22). Washington: Transportation Research Board
57. Zhao, Z., Chen, W., Wu, X., et al. (2017). LSTM network: A deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems*, 11, 68–75 <https://doi.org/10.1049/iet-its.2016.0208>.
58. Huang, W., Song, G., Hong, H., & Xie, K. (2014). Deep architecture for traffic flow prediction: Deep belief networks with multitask learning. *Intelligent Transportation Systems*, 15, 2191–2201 <https://doi.org/10.1109/TITS.2014.2311123>.
59. Niu, X., Zhu, Y., & Zhang, X. (2014). DeepSense: A novel learning mechanism for traffic prediction with taxi GPS traces. In *Proceedings of the 2014 IEEE Global Communications Conference (GLOBECOM)* (pp. 2745–2750). Austin: IEEE.
60. Lv, Y., Duan, Y., Kang, W., et al. (2015). Traffic flow prediction with big data: A deep learning approach. *IEEE Trans Intelligent Transportation Systems*, 16, 865–873 <https://doi.org/10.1109/TITS.2014.2345663>.
61. Srinivasan, D., Wai Chan, C., & Balaji, P. G. (2009). Computational intelligence-based congestion prediction for a dynamic urban street network. *Neurocomputing*, 72, 2710–2716 <https://doi.org/10.1016/j.neucom.2009.01.005>.
62. Yin, H., Wong, S. C., Xu, J., & Wong, C. K. (2002). Urban traffic flow prediction using a fuzzy-neural approach. *Transportation Research Part C: Emerging Technologies*, 10, 85–98 [https://doi.org/10.1016/S0968-090X\(01\)00004-3](https://doi.org/10.1016/S0968-090X(01)00004-3).
63. Hu, C., Xie, K., Song, G., & Wu, T. (2008). Hybrid process neural network based on spatio-temporal similarities for short-term traffic flow prediction. In *Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems (ITSC)* (pp. 253–258). Beijing: IEEE.
64. Hu, J., Song, J., Yu, G., & Zhang, Y. (2003). A novel networked traffic parameter forecasting method based on Markov chain model. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics* (pp. 3595–3600). Washington: IEEE.
65. Liu, L., Khalilia, M., Tan, H., & Zhuang, P. (2009). Traffic pattern forecasting using time series analysis between spatially adjacent sensor clusters. In *Proceedings of 2009 international conference on machine learning and cybernetics* (pp. 3155–3160). Hebei: IEEE.
66. Lu, H., Sun, Z., & Qu, W. (2015). Big data-driven based real-time traffic flow state identification and prediction. *Discrete Dynamics in Nature and Society*, 2015, 1–11 <https://doi.org/10.1155/2015/284906>.
67. Ahn, J., Ko, E., & Kim, E. Y. (2016). Highway traffic flow prediction using support vector regression and Bayesian classifier. In *Proceedings of the 2016 International Conference on Big Data and Smart Computing (BigComp)* (pp. 239–244). Hong Kong: IEEE.
68. Fusco, G., Colombaroni, C., & Isaenko, N. (2016). Short-term speed predictions exploiting big data on large urban road networks. *Transportation Research Part C: Emerging Technologies*, 73, 183–201 <https://doi.org/10.1016/j.trc.2016.10.019>.
69. Ishak, S., & Alecsandru, C. (2004). Optimizing traffic prediction performance of neural networks under various topological, input, and traffic condition settings. *Journal of Transportation Engineering*, 130, 452–465.
70. Ishak, S., Kotha, P., & Alecsandru, C. (2003). Optimization of dynamic neural network performance for short-term traffic prediction. *Transportation Research Record: Journal of the Transportation Research Board*, 1836, 45–56.

71. Agafonov, A., & Myasnikov, V. (2015). Traffic Flow Forecasting Algorithm Based on Combination of Adaptive Elementary Predictors. In M. Y. Khachay, N. Konstantinova, A. Panchenko, et al. (Eds.), *Revised selected papers of the 4th International Conference on Analysis of Images, Social Networks and Texts* (pp. 163–174). Yekaterinburg: Springer International Publishing.
72. Gebresilassie, M. A. (2017). *Spatio-temporal traffic flow prediction*. Stockholm: MSc thesis, Royal Institute of Technology.
73. Jin, X., Zhang, Y., & Yao, D. (2007). Simultaneously prediction of network traffic flow based on PCA-SVR. In D. Liu, S. Fei, Z. Hou, et al. (Eds.), *Advances in neural networks – ISNN 2007* (pp. 1022–1031). Nanjing: Springer Berlin Heidelberg.
74. Mitrovic, N., Asif, M. T., Dauwels, J., & Jaillet, P. (2015). Low-dimensional models for compressed sensing and prediction of large-scale traffic data. *IEEE Transactions on Intelligent Transportation Systems*, 16, 2949–2954 <https://doi.org/10.1109/TITS.2015.2411675>.
75. Xing, X., Zhou, X., Hong, H., et al. (2015). Traffic flow decomposition and prediction based on robust principal component analysis. In *Proceedings on the 2015 IEEE 18th International Conference on Intelligent Transportation Systems (ITSC)* (pp. 2219–2224). Las Palmas: IEEE.
76. Sun, S., Zhang, C., & Yu, G. (2006). A Bayesian network approach to traffic flow forecasting. *IEEE Transactions on Intelligent Transportation Systems*, 7, 124–132 <https://doi.org/10.1109/TITS.2006.869623>.
77. Salamanis, A., Kehagias, D. D., Filelis-Papadopoulos, C. K., et al. (2016). Managing spatial graph dependencies in large volumes of traffic data for travel-time prediction. *IEEE Transactions on Intelligent Transportation Systems*, 17, 1678–1687 <https://doi.org/10.1109/TITS.2015.2488593>.
78. Tan, H., Wu, Y., Shen, B., et al. (2016). Short-term traffic prediction based on dynamic tensor completion. *IEEE Transactions on Intelligent Transportation Systems*, 17, 2123–2133.
79. Tan, H., Song, L., Cheng, Y., et al. (2014). A tensor completion-based traffic state estimation model. In *Proceedings of the 14th COTA international conference of transportation professionals* (pp. 298–309). Changsha: American Society of Civil Engineers
80. Wu, Y., Tan, H., Peter, J., et al. (2015). Short-term traffic flow prediction based on multilinear analysis and k-nearest neighbor regression. In *Proceedings of the 15th COTA international conference of transportation professionals* (pp. 556–569). Beijing: American Society of Civil Engineers
81. Zhao, J., Gao, Y., Tang, J., et al. (2018). Highway travel time prediction using sparse tensor completion tactics and K nearest neighbor pattern matching method. *Journal of Advanced Transportation*, 2018, 16.
82. Han, Y., & Moutarde, F. (2013). Statistical traffic state analysis in large-scale transportation networks using locality-preserving non-negative matrix factorisation. *IET Intelligent Transport Systems*, 7, 283–295 <https://doi.org/10.1049/iet-its.2011.0157>.
83. Han, Y., & Moutarde, F. (2012). Analysis of large-scale traffic dynamics using non-negative tensor factorization. In *Proceedings of the 19th ITS world congress* (p. 12). Vienna: AustriaTech
84. Xu, L., Wang, Y., Yu, H., & Li, H. (2015). Feature extraction of urban traffic network data based on locally sensitive discriminant analysis algorithm. In *Proceedings of the 15th COTA international conference of transportation professionals* (pp. 2192–2203). Beijing: American Society of Civil Engineers
85. Guo, F., Krishnan, R., & Polak, J. W. (2012). Short-term traffic prediction under normal and incident conditions using singular spectrum analysis and the k-nearest neighbour method. In *Proceedings of the IET and ITS conference on road transport information and control (RTIC 2012)* (pp. 11–17). London: Institution of Engineering and Technology.
86. Shang, Q., Lin, C., Yang, Z., et al. (2016). A hybrid short-term traffic flow prediction model based on singular Spectrum analysis and kernel extreme learning machine. *PLoS One*, 11, 25 <https://doi.org/10.1371/journal.pone.0161259>.
87. Makridakis, S., & Hibon, M. (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting*, 16, 451–476 [https://doi.org/10.1016/S0169-2070\(00\)00057-1](https://doi.org/10.1016/S0169-2070(00)00057-1).
88. Schimbinschi, F., Moreira-Matias, L., Nguyen, V. X., & Bailey, J. (2017). Topology-regularized universal vector autoregression for traffic forecasting in large urban areas. *Expert Systems with Applications*, 82, 301–316 <https://doi.org/10.1016/j.eswa.2017.04.015>.
89. Amaro e Silva, R., & Brito, M. (2018). Impact of network layout and time resolution on spatio-temporal solar forecasting. *Solar Energy*, 163, 329–337 <https://doi.org/10.1016/j.solener.2018.01.095>.
90. Jung, J., & Broadwater, R. P. (2014). Current status and future advances for wind speed and power forecasting. *Renewable and Sustainable Energy Reviews*, 31, 762–777 <https://doi.org/10.1016/j.rser.2013.12.054>.
91. Ringkjøb, H.-K., Haugan, P. M., & Solbrekke, I. M. (2018). A review of modelling tools for energy and electricity systems with large shares of variable renewables. *Renewable and Sustainable Energy Reviews*, 96, 440–459 <https://doi.org/10.1016/j.rser.2018.08.002>.
92. Mairal, J. (2014). Sparse modeling for image and vision processing. *Foundations and Trends® in Computer Graphics and Vision*, 8, 85–283 <https://doi.org/10.1561/06000000058>.
93. Lee, P. Y., Loh, W. P., & Chin, J. F. (2017). Feature selection in multimedia: The state-of-the-art review. *Image and Vision Computing*, 67, 29–42 <https://doi.org/10.1016/j.imavis.2017.09.004>.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)