Completion and continuation of nonlinear traffic time series: a probabilistic approach

2003 J. Phys. A: Math. Gen. 36 11369

(http://iopscience.iop.org/0305-4470/36/45/001)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **36** (2003) 11369–11383 PII: S0305-4470(03)64959-0

Completion and continuation of nonlinear traffic time series: a probabilistic approach

D Belomestny[1], **V Jentsch**[1] **and M Schreckenberg**[2]

[1] Institut für Angewandte Mathematik, Interdisziplinäres Zentrum für Komplexe Systeme, Universität Bonn, Meckenheimer Allee 176, 53115 Bonn, Germany
[2] Physics of Transport and Traffic, Universität Duisburg-Essen, Lotharstr. 1, 47048 Duisburg, Germany

Received 18 June 2003, in final form 30 June 2003
Published 29 October 2003
Online at stacks.iop.org/JPhysA/36/11369

Abstract

When dealing with nonlinear time series of car traffic on highways, one of the outstanding problems to be solved is completion and continuation of data in space and time. To this end, the underlying process is decomposed into stochastic and deterministic components. The former is approximated by Gaussian white noise, while the latter refers, apart from always existing trends, to the space- and time-dependent jam propagation process. Jams are modelled in terms of dynamical Bayesian networks with radial basis functions involved. The models developed are used to tackle travel time estimation and prediction. Results are obtained for one of the most crowded traffic areas of Europe, namely the ring-like highway around Cologne.

PACS numbers: 45.70.Vn, 07.05.−t, 05.45.−a, 02.50.−r, 02.60.−x

1. Introduction

To improve freeway traffic conditions in both the long and the short run a variety of control measures can be employed, such as ramp metering, variable speed limits or driver assistance devices. Another promising class of measures to serve this purpose is providing travel time information for travel decision making. The success of travel time information will critically depend on how individuals will respond to it, which in turn depends on the travellers' confidence in the accuracy and reliability of the information and the systems providing it. However, prediction of travel times based on past and current traffic data is not straightforward due to, among others, the high complexity and ill-predictability of the traffic process, faulty or missing observations and different data sources.

An important issue to be mentioned is the difference between travel time estimation and travel time prediction. Travel time estimation encompasses calculating or approximating travel times after the trips have been completed. Estimates of travel times give insight into the actual and past conditions of the road network. Different techniques have been developed

on the basis of time-averaged speeds and volumes (for an overview, see [1]). Usually the data are available only from a limited number of irregularly spaced detectors. In order to ensure high-quality travel time estimations, however, properties of the correlation structure in the underlying spatial–temporal process have to be taken into account. Due to the highly nonlinear characteristics of traffic flow ordinary linear interpolation methods such as spatial–temporal kriging are not appropriate and should thus be avoided.

On the other hand, travel time prediction addresses the problem of calculation or approximation of travel times before trips have actually been made. Since no traffic measurements are available for future periods, travel time prediction requires fundamental knowledge of the underlying processes that will be used to understand, learn and simulate these processes.

For the analysis and prediction of traffic variables data-driven, dynamical and empirical models have been used.

(i) Data-driven models are statistical or inductive models, including (or combining) methods of time series analyses, Kalman filtering, Bayes classification and estimation techniques, as well as connectionist methods (see [5–9]). They require time series of past and present traffic variables, such as speed, flux, density (occupancy) and travel times as input.
(ii) Dynamical models for travel time prediction, such as METANET, SIMRES, STM, DynaMIT (see [1]) or OLSIM (see [2]), are based upon a (macro-, meso- or microscopic) traffic simulation model. Most of these models demand dynamic matrices as input. Individual or aggregate travel times (and other traffic variables) evolve upon repeatedly updating some initial traffic states.
(iii) Empirical models mix results from statistical physics, traffic engineering and observations. We refer to the work of Kerner *et al* [3, 4] who classified jams by a few relevant dynamical features, on the basis of which they were able to trace jams for a specific detector configuration.

Although impressed by the sheer number of models and simulations, we yet feel that the travel time problem is far from being solved. This prompts us to search for another, hitherto not much used approach: our starting point is data and our methods to handle them are drawn from time series analysis. Cleverly mixed with results from empirical models and refined by 'soft computing' methods, we are able to solve the problem with satisfactory precision. Our paper is organized as follows. In section 2, we describe and illustrate the database used in this paper. In section 3, we present the basic formalisms necessary for tackling the problem of completion and continuation. In section 4, we apply the methods of section 3 to complete irregularly sampled traffic data on the Cologne ring. Once this is done, we are able to estimate and forecast travel times. A summary and discussion of our findings is given in section 5.

2. Data

Our traffic data stem from the Cologne ring that embraces the city of Cologne. The ring has a circumference of 52 km and carries an average flow of the order of 130 000 vehicles per day. The larger part is equipped with three lanes, the rest contains two lanes. Speed is limited; limits vary between 80 km h^{-1} and 120 km h^{-1}. The Cologne ring represents a complex network of numerous intersections, junctions and on-off-ramps (see figure 1).

The traffic database is excellent in our case; we have minute aggregates from 50 detectors which are irregularly placed along the ring, with mean distance of about 1 km. Data have been sampled for more than two years, ranging from July 2000 to February 2003. A typical set of measurements from the eastern branch of the ring is shown in figure 2.

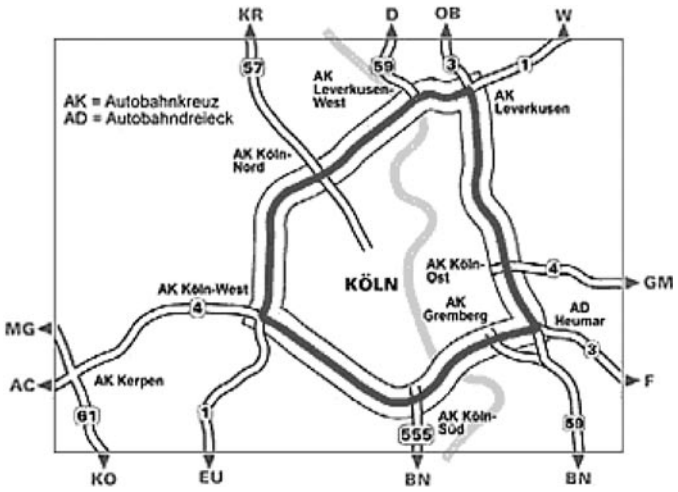

Figure 1. Geometry of the Cologne ring. Shown are the ring along with the major branchings.

Each panel represents a specific location (indicated by kilometres) and the aggregated velocity records for several hours and for all lanes. One clearly observes free flow, where maximal speeds of 120 km h^{-1} are attained, as distinct from congested flow, where velocities drop to 50 km h^{-1} or below. We note that congested flow on the ring is largely governed by external factors, such as inflow and outflow of cars that fold into the internal traffic dynamics, thus causing predominantly 'stop and go' oscillations with quasi-synchronized lane velocities. This can be seen in figure 2, where congestion events are characterized by an increase of velocity variation for each lane but a decrease of velocity variations (synchronization) between each other. 'Real' jams with zero flux are the absolute exceptions on the Cologne ring.

3. The model

Another important feature can clearly be seen in figure 2. Traffic time series are nonlinear and nonstationary near the transition zone, where free flow tends to become congested flow and vice versa. Unfortunately, methods that cope with nonlinearities are still sparse [10] and often impractical. Inventing new methods, however, is a hazardous endeavour. So we have decided to combine known mathematical methods to generate problem-oriented methods. For this to be successful, it is necessary first to analyse the specific nature of processes that underlie the time series.

3.1. The physical base

The physical mechanisms consist of both random and deterministic components. Determinism is introduced by the road geometry, traffic rules and various external factors, giving rise to well-known trends, such as daily and weekly cycles. Superimposed on this are fluctuations, reflecting random elements in the driver's behaviour who seeks to maximize his own advantage by braking, accelerating and lane changing. However, as can be seen from figure 2, the most prominent features in traffic are jams or jam-like structures, characterized by a break-down in velocity. Such events may be extended in time and space, but may as well be short lived and much localized. Jam events have long been known, the principles of which have been revealed

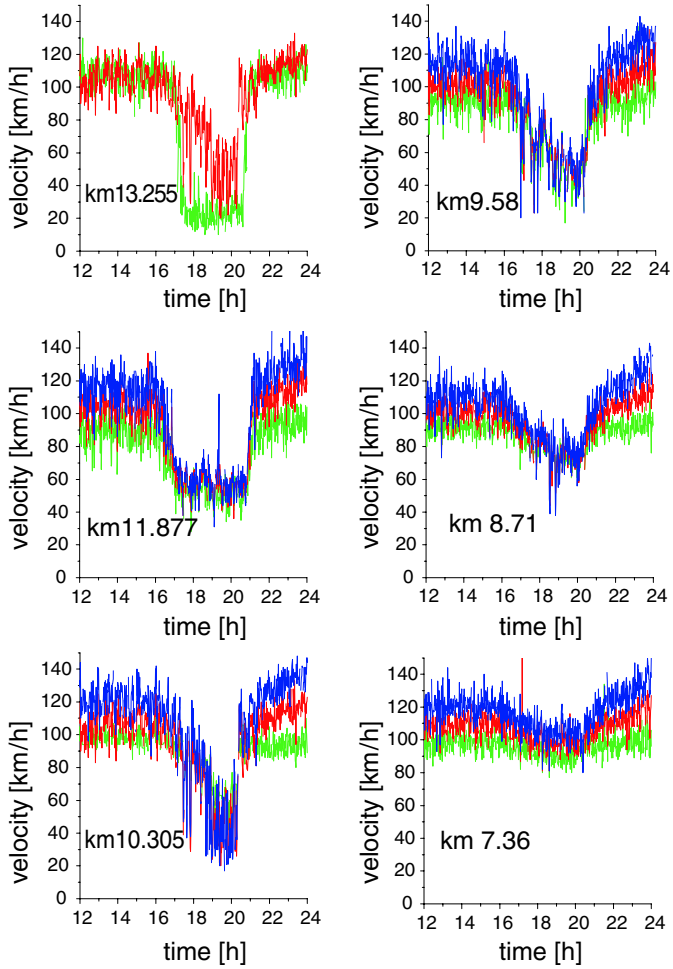

Figure 2. Velocity time series for six adjacent locations on the A3 near Heumar with driving direction towards the south. Blue, red and green correspond to the left, middle and right lanes, respectively.

by macroscopic theory [11]. In free-flow conditions, information flows in the same direction as traffic does, while in congested conditions, information flows in the opposite direction. So-called shock waves occur, as traffic from upstream is forced to slow down due to slower traffic downstream. If the difference between speed and speed variations of the two colliding regimes is large enough, the resulting jam will move in the upstream direction. However, it is only recently that jams have been treated in a more systematic fashion (see [4]). According to these authors moving jams are governed by a few parameters independent of their size, origin and preceding traffic state. Exploiting, in addition, spatial correlations, jams seem to be predictable and traceable. This will be taken up and applied to the Cologne ring traffic. In a way, we formalize the empirical results presented in [3], without differentiating, however, subtleties of jams as mentioned in [4]. For our purposes, this is not really relevant. We use a Bayesian type of network to trace the evolution of jam-like structures. Combining this with the methods of historical profiles, we are able to complete and continue the traffic time series.

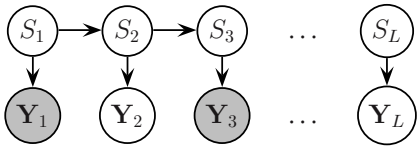

Figure 3. The structure of the jam tracing model.

This in turn is prerequisite for achieving our goal: estimating and predicting travel times, with a statistically based error check included.

3.2. Evaluation of the method

Let $\Pi_1, \ldots, \Pi_L$ be L regularly spaced locations on the road enumerated from downstream to upstream. We assume that locations $\Pi_k, k \in \mathcal{K}$, where $\mathcal{K}$ is a subset of $\{1, \ldots, L\}$, are equipped with detectors which measure velocity, flow and occupancy, among others. Data are signalled in the form of 1 min aggregates to a remote station for further processing. The resulting time series are interpreted as stochastic processes, denoted by $Y_k(t)$, which may be decomposed into

$$Y_k(t) = h(t) + f_k(t) + \epsilon_t \qquad k = 1, \ldots, L \tag{1}$$

where $h(t)$ is a baseline representing free flow for the quantity under consideration, ϵ_t are independent identically distributed random variables with variance σ^2 and $f_k(t)$ is the jam process, where jam is used in a somewhat loose manner including all forms of congested traffic (synchronized, stop and go, wide moving jam). The baseline can be obtained by averaging the quantity of interest for different subsets of the original data set (for example, all Mondays) in free-flow conditions. Jam process $f_k(t)$ is the stochastic approximation of the process $Y_k(t) - h(t)$ by means of radial basis function (RBF)

$$f_k(t) = w_k^0 + \sum_{j=1}^{M} w_k^j \exp\left[-\frac{\left(t - c_k^j\right)^2}{\left(\tau_k^j\right)^2}\right] \qquad t = 1, \ldots, T \tag{2}$$

where $\mathbf{w}_k$, $\mathbf{c}_k$ and τ_k represent weights, centres and variances of the RBF, respectively. These parameters form the hidden processes $S_k = (\mathbf{w}_k, \mathbf{c}_k, \tau_k)$. They are introduced to support the jam model and supposed to be Markovian. The architecture of the resulting network is sketched in figure 3. The hidden process passes all L locations giving rise (see (1)) to processes $Y_k(t)$, some of which are observable (black circles).

The transition distribution for S_k is chosen for simplicity to have a factorial form

$$\mathbf{P}(S_k|S_{k-1}) = \mathbf{P}(\mathbf{w}_k|\mathbf{w}_{k-1})\mathbf{P}(\mathbf{c}_k|\mathbf{c}_{k-1})\mathbf{P}(\tau_k|\tau_{k-1}). \tag{3}$$

We also assume that for some matrices B and D, positive definite matrices Q_1, Q_2, Q_3 and some vector $\mathbf{m}$ (all of which have to be estimated), the transition probabilities are given by

$$\mathbf{P}(\mathbf{c}_k|\mathbf{c}_{k-1}) = (2\pi)^{-M/2}|Q_1|^{-1/2}\exp\{(\mathbf{c}_k - \mathbf{c}_{k-1} + \mathbf{m})' Q_1^{-1}(\mathbf{c}_k - \mathbf{c}_{k-1} + \mathbf{m})\} \tag{4}$$

$$\mathbf{P}(\mathbf{w}_k|\mathbf{w}_{k-1}) = (2\pi)^{-M/2}|Q_2|^{-1/2}\exp\{(\mathbf{w}_k - B\mathbf{w}_{k-1})' Q_2^{-1}(\mathbf{w}_k - B\mathbf{w}_{k-1})\} \tag{5}$$

$$\mathbf{P}(\tau_k|\tau_{k-1}) = (2\pi)^{-M/2}|Q_3|^{-1/2}\exp\{(\tau_k - D\tau_{k-1})' Q_3^{-1}(\tau_k - D\tau_{k-1})\} \tag{6}$$

$$\mathbf{P}(Y(t)|S) = (2\pi\sigma^2)^{-1/2}\exp\left\{-\frac{(Y(t) - f(t;S))^2}{2\sigma^2}\right\} \tag{7}$$

where the distributions of $\mathbf{c}_0$, $\mathbf{w}_0$, τ_0 are assumed normal. Further, without loss of generality we suppose that the set $\mathcal{K}$ is arithmetic, that is for some integer $n > 0$ the observable but incomplete process $\tilde{Y}_k$ is related to Y_k in the following way:

$$\tilde{Y}_k = Y_{kn} \qquad k = 0, \ldots, N \tag{8}$$

where $N = [L/n]$ (note that observable quantities are marked by tilde '˜'). The problem of completing the process at the points of missing detectors thus entails two sub-problems:

(i) estimation of $\tilde{S}_k$ on the basis of $\tilde{Y}_k$,
(ii) estimation of transition probabilities $\mathbf{P}(S_k|S_{k-1})$ on the basis of $\mathbf{P}(\tilde{S}_{k+1}|\tilde{S}_k)$.

3.2.1. Estimation of $\tilde{S}_k$. The problem can be resettled as one of approximating the given noisy time series by Gaussian RBF and at the same time holding the number of parameters (e.g., the dimension of the process S) as small as possible.

Suppose we want to approximate an arbitrary function $f(x)$ by a set of M radial basis functions $\phi_j(x)$, centred on the centroids c_j. The approximation of the function $f(x)$ (denoted by $\hat{f}$) may be expressed as a linear combination of the radial basis functions:

$$\hat{f}(x) = \sum_{j=1}^{M} w_j \phi_j(x - c_j) \tag{9}$$

where w_j are real-valued weight factors. A typical choice for the radial basis functions is a set of one-dimensional Gaussian kernels:

$$\phi_j(x - c_j) = \exp\left(-\frac{(x - c_j)^2}{2\tau_j^2}\right) \qquad \tau_j > 0 \quad c_j \in \mathbb{R}. \tag{10}$$

Once the number and the general shape of the radial basis functions $\phi(x)$ is chosen, the RBF network has to be trained properly. Given a training data set T of size N_T,

$$T = \{(x_p, y_p) \in \mathbb{R} \times \mathbb{R}, 1 \leqslant p \leqslant N_T : y_p = f(x_p)\} \tag{11}$$

the training algorithm consists in finding the parameters c_j, τ_j and w_j, such that $\hat{f}(x)$ fits the unknown function $f(x)$ as closely as possible. This is realized by minimizing a cost function. After the best-fit function is calculated, the performance of the RBF network is estimated by computing an error criterion. Consider a validation data set V, containing N_V data points:

$$V = \{(x_q, y_q) \in \mathbb{R} \times \mathbb{R}, 1 \leqslant q \leqslant N_V : y_q = f(x_q)\}. \tag{12}$$

The error criterion can be chosen to be the mean square error:

$$\mathrm{MSE}_V \equiv \frac{1}{N_V} \sum_{q=1}^{N_V} (y_q - \hat{f}(x_q))^2 \tag{13}$$

where y_q are the desired outputs. Often, the training algorithm is decoupled into a three-stage procedure:

- determine the centres c_j of the Gaussian kernels,
- compute the widths of the Gaussian kernels τ_j,
- compute the weights w_j.

Since we are mainly interested in tracing such extreme events as jams we can place the centres of the corresponding basis functions at points where local maxima (for density time series) or local minima (for velocity time series) are located. The number of basis functions can be the taken as the number of local maxima (respectively local minima). Other parameters should

be chosen accordingly. The weights can be found by solving the linear equation (9). For the width factors, typically two alternatives are considered. The first one consists in taking the widths τ_j equal to a constant for all Gaussian functions [12, 13]. In [13], for example, the widths are fixed as follows:

$$\tau = \frac{d_{\max}}{\sqrt{2M}} \tag{14}$$

where M is the number of centres and $d_{\max}$ is the maximum distance between these centres.

The second option consists in estimating the width of each Gaussian function independently. This can be done, for example, by simply computing the standard deviation of the distance between the data and their corresponding centres. An iterative procedure to estimate the standard deviation is suggested in [14]. On the other hand, the computation of the width factors τ_j by the 'r-nearest neighbours heuristic' is proposed in [15]:

$$\tau_j = \frac{1}{r}\left(\sum_{i=1}^{r} |c_i - c_j|^2\right)^{1/2} \tag{15}$$

where the c_i are the r-nearest neighbours of centres c_j. A suggested value for r is 2. This second class of methods offers the advantage of taking the data variance into account. In practice, the latter methods are able to perform much better, as they offer a greater adaptability to the data than a fixed-width procedure.

3.2.2. Estimation of transition probabilities. After estimating $\tilde{S}_k$ we can turn to the second problem of estimation transition probabilities $\mathbf{P}(S_t|S_{t-1})$. Let X_t denote one of the two processes $\mathbf{w}_t$, τ_t, then we have for the corresponding incomplete hidden process

$$\tilde{X}_k = X_{kn} \qquad k = 1, \ldots, N \tag{16}$$

where $N = [L/n]$. If we set $p(\tilde{X}_k|\tilde{X}_{k-1}) = \mathcal{N}(C\tilde{X}_{k-1}, R)$ for some matrices C (corresponding to B or D) and R (corresponding to Q_2 or Q_3) and assume a Gaussian initial state density for $\tilde{X}_1$ with covariance matrix V_1 and mean vector π_1, namely

$$\mathbf{P}(\tilde{X}_1) = \exp\left\{-\tfrac{1}{2}[\tilde{X}_1 - \pi_1]' V_1^{-1}[\tilde{X}_1 - \pi_1]\right\}(2\pi)^{-k/2}|V_1|^{-1/2} \tag{17}$$

then the corresponding log likelihood can be written as

$$\begin{aligned}\mathcal{Q} = \log P(\{\tilde{\mathbf{X}}\}) = &-\sum_{t=1}^{N-1}\left(\frac{1}{2}[\tilde{X}_{t+1} - C\tilde{X}_t]' R^{-1}[\tilde{X}_{t+1} - C\tilde{X}_t]\right) - \frac{N-1}{2}\log|R| \\ &- \frac{1}{2}[\tilde{X}_1 - \pi_1]' V_1^{-1}[\tilde{X}_1 - \pi_1] - \frac{1}{2}\log|V_1| - NM\log(2\pi).\end{aligned} \tag{18}$$

Maximization of $\mathcal{Q}$ with respect to C is equivalent to solving the equation

$$\frac{\partial \mathcal{Q}}{\partial C} = -\sum_{t=1}^{N-1} R^{-1}\tilde{X}_t\tilde{X}'_{t+1} + \sum_{t=1}^{N-1} R^{-1}C\tilde{X}_t\tilde{X}'_t \tag{19}$$

which yields the following estimate (in what follows we denote by the '$\hat{}$' an estimated quantity)

$$\hat{C} = \left(\sum_{t=1}^{N-1}\tilde{X}_t\tilde{X}'_{t+1}\right)\left(\sum_{t=1}^{N-1}\tilde{X}_t\tilde{X}'_t\right)^{-1}. \tag{20}$$

For covariance matrix R, which is assumed to have inverse R^{-1}, differentiation leads to the equation

$$\frac{\partial \mathcal{Q}}{\partial R^{-1}} = \frac{N-1}{2}R - \frac{1}{2}\sum_{t=1}^{N-1}(\tilde{X}_{t+1}\tilde{X}'_{t+1} - C\tilde{X}_t\tilde{X}'_{t+1} - \tilde{X}_t\tilde{X}'_{t+1}C' + C\tilde{X}_t\tilde{X}'_tC') = 0 \tag{21}$$

and

$$\hat{R} = \frac{1}{N-1}\left(\sum_{t=1}^{N-1}\tilde{X}_{t+1}\tilde{X}'_{t+1} - \hat{C}\sum_{t=1}^{N-1}\tilde{X}_{t+1}\tilde{X}'_t\right). \tag{22}$$

If we assume multiple observation sequences we can effectively estimate the initial mean and covariance. Assume K observation sequences of length N, and let $\tilde{X}_t^{(i)}$ be the estimate of state at time t given the ith sequence, then we have

$$\frac{\partial \mathcal{Q}}{\partial \pi_1} = \big(\tilde{X}_1^{(i)} - \pi_1\big)V_1^{-1} = 0 \tag{23}$$

$$\hat{\pi}_1 = \frac{1}{K}\sum_{i=1}^{K}\tilde{X}_1^{(i)} \tag{24}$$

$$\frac{\partial \mathcal{Q}}{\partial V_1^{-1}} = \frac{1}{2}V_1 - \frac{1}{2}\big(\tilde{X}_1^{(i)}\tilde{X}_1^{(i)'} - \tilde{X}_1^{(i)}\pi_1' - \pi_1\tilde{X}_1^{\prime(i)} + \pi_1\pi_1'\big)V_1^{-1} = 0 \tag{25}$$

$$\hat{V}_1 = \frac{1}{K}\sum_{i=1}^{K}\big(X_1^{(i)} - \pi_1\big)\big(X_1^{(i)} - \pi_1\big)'. \tag{26}$$

Upon setting $p(X_t|X_{t-1}) = \mathcal{N}(AX_{t-1}, \Gamma)$ for some matrices A and Γ (see (5) and (6)) we have due to (16)

$$\mathbf{P}(\tilde{X}_{t+1}|\tilde{X}_t) = \mathcal{N}\left(\tilde{X}_{t+1}; A^{n-1}\tilde{X}_t, \sum_{k=0}^{n-1}A^k\Gamma(A^k)^N\right). \tag{27}$$

Hence, the estimation for A can be expressed as ($\hat{C}$ as given by (20) is clearly positive definite)

$$\hat{A} = (\hat{C})^{1/(n-1)}. \tag{28}$$

Assuming, for example, that Γ is a diagonal matrix, we derive for the estimated Γ

$$\hat{\Gamma} = \hat{R}\left(\sum_{k=0}^{n-1}\hat{A}^k(\hat{A}^k)^T\right)^{-1}. \tag{29}$$

For component $\tilde{\mathbf{c}}$ we set

$$\mathbf{P}(\tilde{\mathbf{c}}_{t+1}|\tilde{\mathbf{c}}_t) = \mathcal{N}(\tilde{\mathbf{c}}_t + \mathbf{a}, F) \tag{30}$$

and the following estimates can be easily obtained:

$$\hat{\mathbf{a}} = \frac{1}{2}\sum_{t=1}^{N-1}[(\tilde{\mathbf{c}}_{t+1} - \tilde{\mathbf{c}}_t)' + (\tilde{\mathbf{c}}_{t+1} - \tilde{\mathbf{c}}_t)] \tag{31}$$

$$\hat{F} = \frac{1}{N-1}\left(\sum_{t=1}^{N-1}\tilde{\mathbf{c}}_{t+1}\tilde{\mathbf{c}}'_{t+1} - \sum_{t=1}^{N-1}\tilde{\mathbf{c}}_{t+1}\tilde{\mathbf{c}}'_t\right). \tag{32}$$

Finally, Q_1 and $\mathbf{m}$ (see (4)) can be estimated by

$$\hat{Q}_1 = n^{-1}\hat{F} \qquad \hat{\mathbf{m}} = n^{-1}\hat{\mathbf{a}}. \tag{33}$$

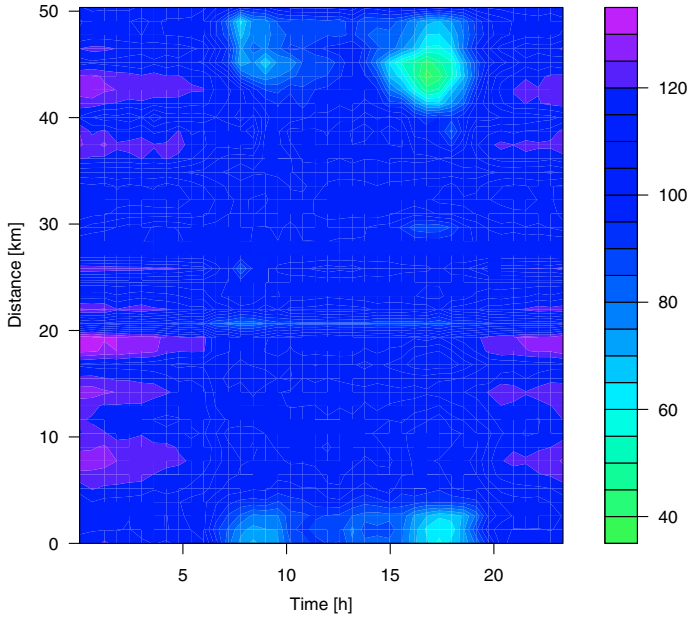

Figure 4. Results of completion for velocity time series during Friday (Cologne ring, driving direction towards the south, left lane).

Table 1. Normalized mean square errors (NMSE) for different completion methods.

	NMSE on test set	
Number of detectors in training set	Jam tracing method	Temporal spatial kriging
15	0.033	0.059
13	0.039	0.065
10	0.048	0.072
5	0.052	0.085

4. Results

We are now in the position to test our nonlinear data analysis methods by means of real traffic data which have been described in section 2.

4.1. Completion

Figure 4 shows the result of completion. Plotted are contour lines in the (x, t)-plane, where $x = 0$ is the northernmost point, corresponding to Kreuz Leverkusen. Cars are heading towards the south, i.e., increasing x. The region of low velocities in the afternoon hours is most prominent around $x = 0$ and $x = 45$ km (Kreuz Leverkusen and Leverkusen, respectively). In order to estimate the goodness of our completion method, we have divided all detectors into two groups (training and test sets). Table 1 shows normalized mean square errors (NMSE) on test sets for our method along with the widely used method of spatial–temporal kriging (see, e.g., [18]). It turns out that errors are considerably smaller for our method, regardless

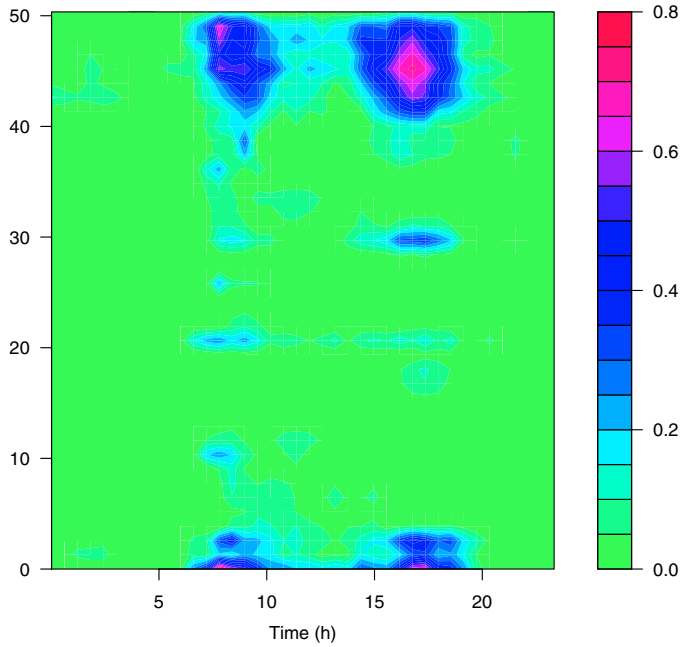

Figure 5. Probability of jam formation on the Cologne ring (middle lane, cars move anticlockwise).

of the number of detectors involved. Once the velocities are completed, it is not difficult to estimate the jam probability, which is depicted in figure 5. We recall that a jam is assumed to occur when velocities drop below some threshold limit, typically of the order of 50 km h^{-1}. Figure 5 incorporates information for all Mondays over three years and covers the entire Cologne ring, where $x = 0$ is defined as before. Note that the basic features of jam probabilities are not affected by the specific threshold. What changes are the numbers (or colours), not structures. So higher thresholds will entail lower probabilities, but will produce similar patterns regardless of the specifics of jams.

Closely related to the above is the time it takes to travel a specified distance. Due to inhomogeneities in space and time of the traffic process, travel time depends on the starting and end points, as well as on day (Monday, Tuesday) and time of the day. Figure 6 gives an idea of the shape of the harmonic velocity on Friday. The harmonic velocity is defined as

$$\mathcal{V}_k(s) = \frac{1}{V_k(s)} \qquad k = 1, \ldots, L \quad s = 1, \ldots, T \tag{34}$$

where $V_k(s)$ is the aggregated velocity measured at the kth location during time $(s, s+1)$, with s $s+1$ being separated by $\Delta t = 1$ min.

Harmonic velocity gives a greater weight to small velocities and therefore makes congested traffic regime more pronounced.

Suppose that we are able to complete the data at all locations $\Pi_1, \ldots, \Pi_L$. Then in terms of $\mathcal{V}$ the arrival time at the ith location, provided the travel has begun at $i = 1$ location, can be

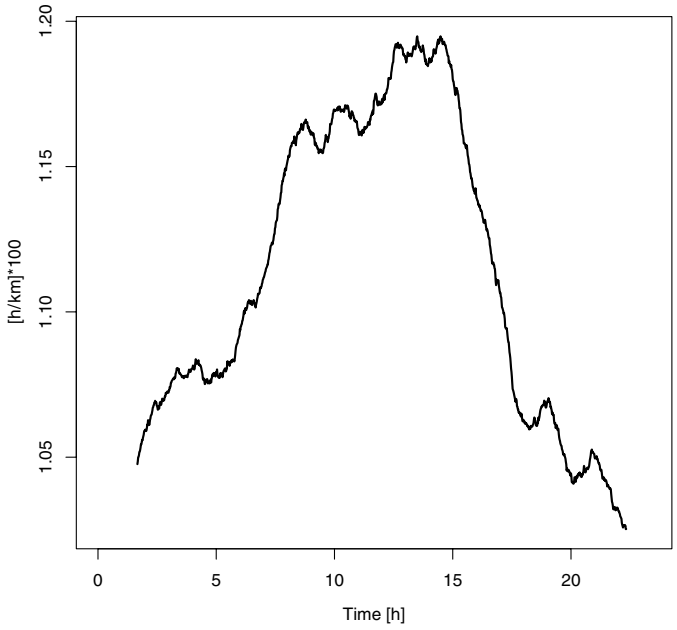

Figure 6. Harmonic velocity baseline for Friday on A3 (Cologne ring).

expressed as

$$t_i = t_{i-1} + d\mathcal{V}_{i-1}([t_{i-1}]) \tag{35}$$

where t_0 is the departure time, d is the distance between consecutive locations and $[t]$ is the integer number of minutes in t. The corresponding travel times are

$$T_i = t_i - t_0 \qquad i = 1, \ldots, L. \tag{36}$$

Some estimates are shown in figure 7. For definiteness, the starting point is fixed at $x = 0$. Travel times are given as a function of distance, for different departure times. Note that the slope of the approximately straight lines depends on departure time (and starting point). Results are summarized in form of a matrix (see table 2) where the corresponding slopes for different combinations of starting points and departure times are calculated all along the Cologne ring.

4.2. Continuation

The problem of continuation of a given time series parallels that of completion and can therefore be tackled in a similar manner. More specifically, the processes $Y_k(t)$ are estimated on $[0, T_0]$ for some $T_0 < T$; on $[T_0, T]$ we set $Y_k(t) = h_k(t) + \hat{f}_k(t)$, where $\hat{f}_k(t)$ is the predicted jam component. Predicted velocities in figure 8 are represented by the grey curve, while the measured time series is represented by the black curve. Comparison of the two curves shows that the trend, including the jam between 20 and 30 min, can be well reproduced; however the details, in particular the shape of the jam, are not yet well captured. The following information

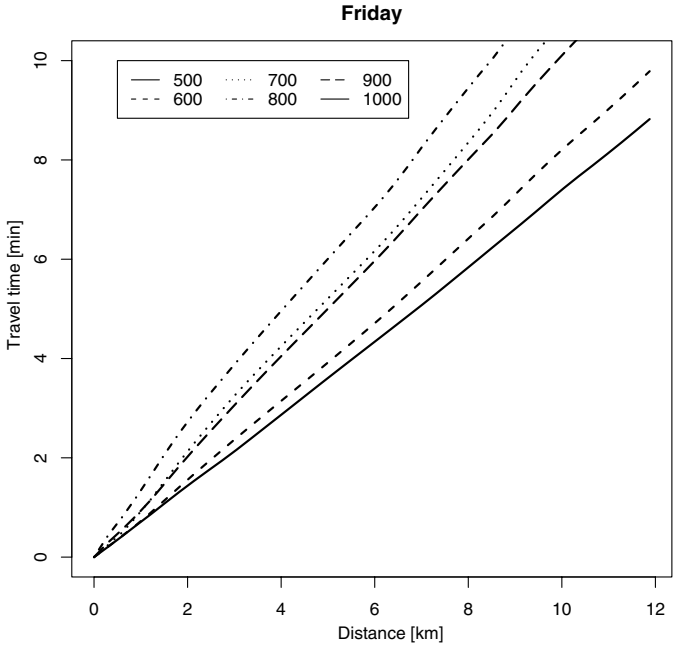

Figure 7. Travel time estimation for Fridays. Different curves correspond to the different departure times (in minutes). These estimates are obtained using completed velocity time series (see figure 4).

Table 2. Slopes of the corresponding travel time curves for different departure times and starting points.

	Departure time (min)							
Start (km)	100	200	300	400	500	600	700	800
0	0.192	0.201	0.204	0.242	0.287	0.229	0.219	0.215
1	0.192	0.201	0.203	0.241	0.282	0.228	0.211	0.211
2	0.194	0.201	0.202	0.242	0.267	0.232	0.211	0.212
3	0.195	0.200	0.201	0.246	0.268	0.237	0.212	0.213
4	0.195	0.200	0.202	0.250	0.270	0.245	0.213	0.214

turns out to be necessary for proper continuation:

- time series of past and present from the detector under consideration,
- time series of past and present from adjacent detectors for both downstream and upstream directions,
- temporal information such as time of day and day of week

Travel time prediction is shown in figure 9. The grey line is the travel time estimation and shows how long on average a driver is on the road, when he starts at $x = 0$ (Leverkusen) at 7 am and leaves the highway at some exit $x > 0$. The prediction horizon is chosen as in figure 8, namely 30 min. The corresponding box plots of prediction for eight locations are based on 50 samples from the estimated distribution. Every box indicates median, quartiles,

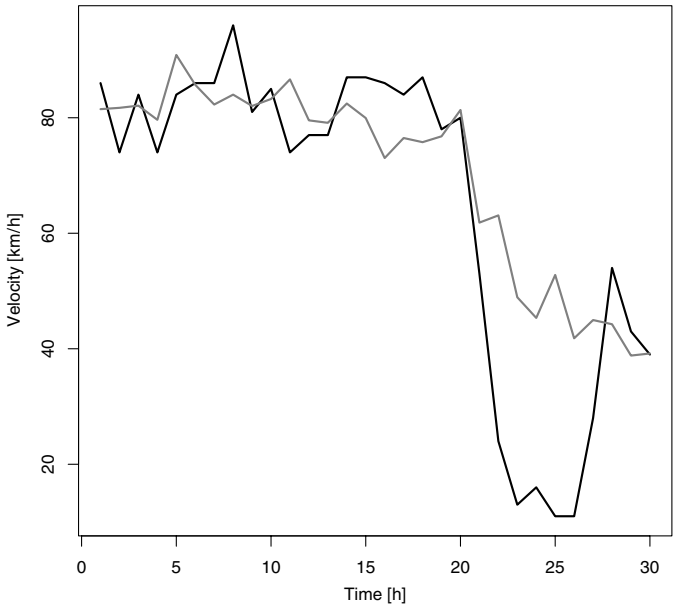

Figure 8. Prediction of velocity records (grey) compared to the measured records (black) (collected near Heumar).

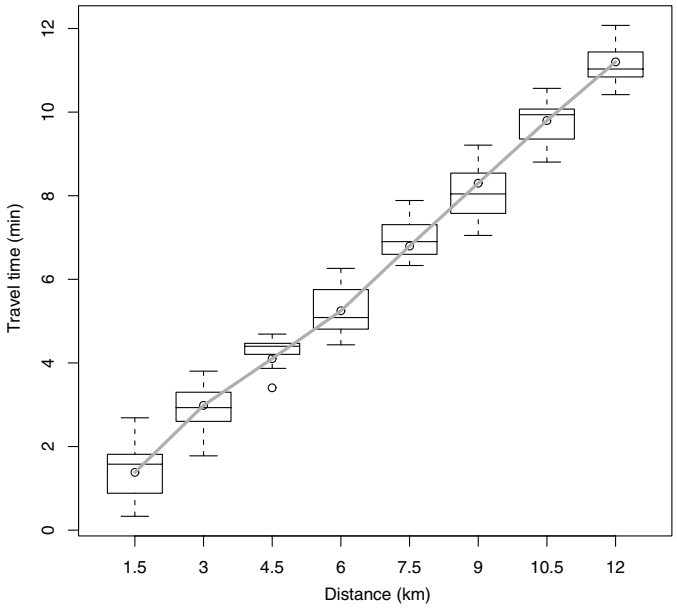

Figure 9. Box plots of predicted and estimated travel times (grey) for eight exit locations (based on 50 samples from the predicted distribution).

extreme values and outliers. The overall impression is that the predicted values are in good agreement with the estimations.

5. Discussion and summary

What is interesting in traffic is the jammed situation, when velocity breaks down to values near zero, and density goes up. The problem is to detect jams and mark endangered areas where jams are most likely to occur. These areas depend on the time of day and the position on the road. On the basis of traffic time series, we have developed a novel method that handles this problem. It completes the spatially incomplete and irregular data sets and thus can catch the hot spots encountered by drivers when they move along the Cologne ring.

Another problem is the prediction of jams. Predicting the future of a system is always hazardous, so much the more, as traffic variables are essentially random: the time series reveal a large amount of dynamical noise. This casts doubt on the predictability of the traffic system. Luckily, jam events also reveal systematic features such as the jam propagation velocities. This fact is exploited for both completion and prediction problems. However, only those jams are predicted in this paper that already exist; for in this case, we can use our completion method and reinterpret it as the continuation of (an already known) time series. Therefore, we do not tackle the problem of predicting an event which at the time of prediction has not yet occurred. Such an adventure requires the analysis of precursory signals (whatsoever exist) and will draw on extreme value statistics, which may (or may not) answer questions such as average waiting times for the next event, or the expected duration and depth of a jam. This is however another story and will be tackled in future.

As already outlined above, an important step towards prediction is the completion of time series. Normally, completion is handled by means of interpolation in space (or time). This largely ignores the processes involved. We do better by a shrewd combination of temporal and spatial data and appropriate combination of methods, i.e. historical method and jam tracing techniques based on radial basis function networks. This yields satisfactory results for completion, when records in space are sparse and irregularly distributed, and even good results in the case of continuation. To summarize:

- Our method combines different temporal–spatial scales. For example, the short time scale (of the order of minutes) is represented by noise, the meso time scale (of the order of hours) by jam (the average extension of a jam in time is about 1 h on the Cologne ring) and the long time scale (of the order of days) by the baseline or trend. Furthermore, the model adaptively chooses the most appropriate combination depending on the current traffic situation and the prediction horizon. This explains the fact that for predictions of more than 30 min ahead, our method gives better results than other approaches (such as artificial neural networks or historical profile method, see results in [5, 8, 9]).
- The structure of information in traffic flow is clearly reflected: the propagation of a shock wave can be traced and its shape can be reconstructed. This shows that the model is closely related to the underlying physical processes.
- The interpretation of results is straightforward. All three components (trend, jam component, noise) are estimated separately and therefore results can be displayed for every component.

Acknowledgments

We thank S Albeverio for numerous discussions and valuable hints concerning the mathematical part of the paper. We also thank C Tornau and R Chrobok for their assistance in data handling and the 'Landesbetrieb Strassenbau NRW' for supplying the data. One of us (DB) was supported by SPP 1114 of the Deutsche Forschungsgemeinschaft.

References

[1] Bovy P H L and Thijs R 2000 *Estimators of Travel Time for Road Networks, New Developments, Evaluation Results, and Applications* (Delft: Technical University)
[2] Nagel K and Schreckenberg M 1992 *J. Physique* I **2** 2221
[3] Kerner B S, Rehborn H and Aleksic M 1999 *Traffic and Granular Flow '99, Social, Traffic, and Granular Dynamics* p 339
[4] Kerner B S and Rehborn H 1996 *Phys. Rev.* E 1297
[5] Helbing D and Treiber M 2002 *Cooperative Transportation Dynamics* **I** 3.1
[6] Goemans J W 2000 *Afstudeerverslag sectie Verkeerskunde en sectie Civieltechnische* Informatica, Faculteit der Civiele Techniek, Technische Universiteit Delft
[7] Cheu R-L 1998 *Proc. Int. Conf. on Applications of Advanced Technologies in Transportation Engineering* (Reston, VA: ASCE) p 247
[8] Dia H 2001 *Eur. J. Oper. Res.* **131** 2 (Special Issue)
[9] Park-Dongjoo and Rilett-Laurence 1999 *Comput. Aided Civ. Infrastruct. Eng.* **14** 357
[10] Kantz H and Schreiber Th 1997 *Nonlinear Time Series Analysis* (Cambridge: Cambridge University Press)
[11] Helbing D 2001 *Rev. Mod. Phys.* **73** 4
[12] Park J and Sandberg I 1991 *Networks, Neur. Comput.* **3** 246
[13] Haykin S 1999 *Neural Networks a Comprehensive Foundation* (Englewood Cliffs, NJ: Prentice-Hall)
[14] Verleysen M and Hlavackova K 1996 *Proc. Int. Conf. on Neural Networks (ICNN) (Washington)* p 199
[15] J Moody and Darken C J 1989 *Neural. Comput.* **1** 281
[16] Anderson B D and Moore J B 1997 *Optimal Filtering* (Englewood Cliffs, NJ: Prentice-Hall)
[17] Baum L E, Petrie T, Soules G and Weiss N 1970 *Ann. Math. Stat.* **41** 164
[18] Ripley B D 1981 *Spatial Statistics* (New York: Wiley)
[19] Fraser A M and Dimitriadis A 1994 *Time Series Prediction: Forecasting the Future and Understanding the Past* ed A S Weigend and N A Gershenfeld (Reading, MA: Addison-Wesley) p 264
[20] Kerner B S and Rehborn H 1996 *Phys. Rev.* E **53** 5
[21] Rabiner L 1996 *Proc. IEEE* **77** 257
[22] Park D, Rilett L and Han G 1999 *J. Trans. Eng.* **125** 515