Predictions of Urban Volumes in Single Time Series

Tom Thomas, Wendy Weijermars, and Eric van Berkum

***Abstract*—Congestion is increasing in many urban areas. This has led to a growing awareness of the importance of accurate traffic-flow predictions. In this paper, we introduce a prediction scheme that is based on an extensive study of volume patterns that were collected at about 20 urban intersections in the city of Almelo, The Netherlands. The scheme can be used for both short- and long-term predictions. It consists of 1) baseline predictions for a given preselected day, 2) predictions for the next 24 h, and 3) short-term predictions with horizons smaller than 80 min. We show that the predictions significantly improve when we adopt some straightforward assumptions about the correlations between and the noise levels within volumes. We conclude that 24-h predictions are much more accurate than baseline predictions and that errors in short-term predictions are even negligibly small during working days. We used a heuristic approach to optimize the model. As a consequence, our model is quite simple so that it can easily be used for practical applications.**

***Index Terms*—Demand forecasting, error measures, evaluating, Kalman filter, uncertainty.**

I. Introduction

Congestion has significantly increased over the last few decades. The efficient use of existing infrastructure by dynamic traffic management is one of the strategies to reduce congestion. An important requirement is the availability of detailed information about travel demand. In general, demand cannot directly be measured but must be estimated using information on volumes, i.e., traffic counts. On Dutch highways, data are collected by a high concentration of detection loops that yield information on both volumes and velocities. This information enables the use of artificial neural networks, e.g., [1] and [2], pattern-matching models, e.g., [3], or extrapolation models, e.g., [4], to make short-term predictions of the traffic circulation.

In urban areas, traffic information is much scarcer, and only recently have traffic data become available in traffic information centers, e.g., [5]–[7]. For these areas, the traffic circulation is usually estimated by a combination of volume measurements and (macroscopic) traffic models. Some authors, e.g., [8], have suggested that, with reliable volume predictions, it will be possible to improve model forecasts of urban traffic circulation.

Different approaches exist for predicting volumes. Extrapolation models (both spatial and temporal) are often used for short-term predictions, e.g., [9]–[13]. Extrapolations can give accurate predictions but only for prediction horizons smaller than 15–20 min. For longer prediction horizons, volume measurements can also be matched to historical patterns. For these predictions, neural networks, e.g., [14] and [15], or clustering methods, e.g., [16] and [17], are applied.

Short-term volume predictions can be used, for example, in flexible control systems that anticipate on rapid changes in demand. Long-term predictions are used in general management applications, for example, in the optimization of traffic signal plans. In this paper, we develop a scheme for both short- and long-term predictions. Unlike most methods, we make an explicit distinction between systematic variation and noise. Contrary to systematic variation (e.g., weekly, seasonal, or weather-related variation), noise in different measurements is uncorrelated and, therefore, unpredictable. Because we separate the noise from the systematic prediction error, we are able to improve our predictions and are also able to evaluate the prediction scheme in a more objective way.

In Section II, we introduce the data. In Section III, we describe the used method. In Section IV, we develop a prediction scheme for the next day, and in Section V, we improve these predictions for short prediction horizons. In Section VI, we estimate the quality of our predictions. In Section VII, we will provide conclusions, and in Section VIII, we will discuss our results.

II. Data

The study area for this research consists of the urban network of Almelo, The Netherlands, which is a medium-sized city with about 70 000 inhabitants and a cross section of about 5 km. Data were collected at about 20 signalized urban intersections from September 2004 until September 2005. Vehicles were detected by inductive loop detectors. The data were processed into volume measurements per link per time interval. A link represents a unidirectional road segment that could contain more than one lane. We define a volume profile $Q_{dl} = (q_{dl1}, \ldots, q_{dln})$ as a time series of n intervals for day d and link l. In most cases, measurements were provided in 5-min intervals so that $n = 288$. However, for about 30% of the links, only measurements in 30-min intervals were provided. For these links, $n = 48$.

In Fig. 1, we show the study area, which covers the whole city of Almelo. The links for which data were collected are marked in the figure. These links are part of the main urban roads for which the maximum allowed speed lies between 50 and 70 km/h. Because congestion is relatively rare, travel times in the study area are often smaller than 15 min.

The volume measurements were inspected, and invalid data were rejected as was suggested in [16]. Invalid data are the result of errors in the measurements (e.g., by failures in the electronics). These errors were detected using certain criteria

Manuscript received September 18, 2008; revised April 27, 2009. First published August 25, 2009; current version published March 3, 2010. This work was supported by Transumo. The Associate Editor for this paper was M. Brackstone.

The authors are with the Centre for Transport Studies, University of Twente, 7500 Enschede, The Netherlands (e-mail: t.thomas@utwente.nl; wendyweijermars@yahoo.com; e.c.vanberkum@utwente.nl).

Digital Object Identifier 10.1109/TITS.2009.2028149

Fig. 1. City of Almelo, The Netherlands. The thick lines correspond with intersection links for which traffic data were collected.

(e.g., volumes should be equal to or larger than 0, volumes should not exceed a certain maximum threshold, and 24-h volumes should be larger than 0). Due to the malfunctioning of detectors during several days or even weeks, a significant fraction of the volume profiles was rejected. It is worth stressing that the remaining volume profiles, which are used in this paper, only contain actual measurements. Thus, we excluded bad data but did not include "artificial" data or in any other way manipulated the measurements. Note that this procedure does not lead to a loss of information, because bad profiles do not contribute to our analyses.

To create a homogeneous sample with good statistics, we only included links with at least ten volume profiles per weekday (Mondays, Tuesdays, etc.) outside the school holiday period. For links with less than ten profiles, the predictions become too sensitive to the noise in individual time series, and the reliability of the predictions can no longer be tested. As a result, the remaining data set is a sample of 72 intersection links, of which 48 have 5-min time series.

Within our sample, there are profiles from (un)official public holidays (Christmas, New Years Day, Eastern, Queensday, Ascension Day and Pentecost, Christmas Eve, the days between Christmas and New Years Day, Good Friday, the day after Good Friday, and the day after Ascension Day). Variations between individual national holidays can be quite large. On Queensday, for example, traffic starts earlier than on New Years Day. We therefore decided to exclude (un)official public holidays from the sample.

Our time series show the presence of recurrent variations with 30-min periods [18]. Time intervals of 30 min are too long to follow these recurrent variations, and they are also too long to follow sharp changes in demand during the rush hours. We therefore did not consider links with 30-min time series in our analyses.

The strength of the volume fluctuations also depends on the average number of traffic signal cycles per time interval. If this number is low, the relative variation in the total green time per time interval will increase, which may lead to a significant variation in volumes when traffic loads are high. To reduce this effect, we decided to aggregate our measurements to 10-min rather than 5-min time intervals. As a consequence, we may lose information. However, in regular situations, volumes generally do not so dramatically change that this would impose a problem for our predictions.

III. Method

Autoregressive models (e.g., ARIMA models) are common in time-series forecasts. Their forecasts are based on linear combinations of measurements from previous time intervals. Travel demand variations are often nonlinear. Several authors, e.g., [11], [12], and [16], have therefore indicated that they prefer to use the average historical profile of a whole day for predictions of a future day. In this case, nonlinear features may already be captured by the historical profile. These authors also found day-to-day variations in the shape of volume profiles and therefore classified various days in different groups. Based on these results, our first prediction for day d, link l, and time interval t is equal to the historical mean of the group to which day d belongs

$$q_{dlt}^{\text{base}} = \frac{\sum\limits_{d' \in D} q_{d'lt}^{\text{obs}}}{n_D}. \tag{1}$$

In (1), we call q_{dlt}^{base} the baseline prediction, and q_{dlt}^{obs} is the measured volume on day d, link l, and time interval t. The group of days to that day d belongs is denoted by D. We classified our days in the following groups: Mondays, Tuesdays,

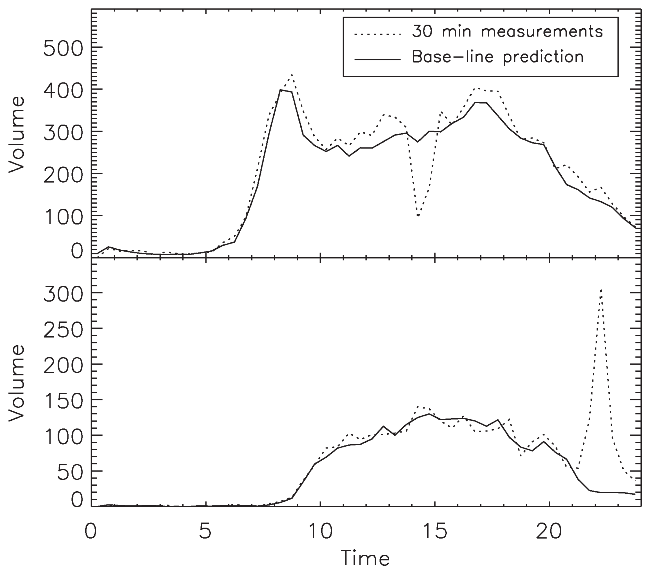

Fig. 2. (Top) Incident. (Bottom) Event. The baseline prediction does not follow the measurements. The measurements are 30-min time series.

Wednesdays, Thursdays, Fridays, Saturdays, Sundays, and the school holiday period. Group D consists of n_D days.

In real-time applications, D contains all days from the previous year, i.e., from day d of the previous year until day $d-1$. When we would test our prediction scheme for all days in one year, the training and test sets would have to contain at least two years of data. Since our data set only contains one year of data and since we think that a baseline prediction should be based on the yearly average, we did not distinguish between a test set and a training set. Although we affirm that this is not correct, we do not think that it will influence the results in a significant way. First, traffic flows usually do not dramatically change from year to year. Second, the yearly average remains more or less the same if it is based on many profiles, e.g., if $n_D \geq 10$.

When large events take place, traffic flows may be influenced by the visitors of these events. At certain locations, this will lead to a significant increase of traffic just before the event has started and after the event has finished. In Almelo, home matches of the local professional football club can be counted among such events. We excluded the match-related volumes, and we do not consider them in this paper. Traffic flows can also abruptly change as a result of incidents. Incidents are left into the sample but only form a marginal fraction of all measurements (about 0.1%). In Fig. 2, we show an example of an incident and an event. In both cases, baseline predictions are well off the measured time series. Note that, for this illustration, we used 30-min aggregates to minimize the disturbing influence of noise.

Baseline predictions sometimes also do not follow the measurements in a regular situation. This is illustrated in the upper panel of Fig. 2. In this example, the baseline prediction systematically lies below the measurements. Apparently, traffic counts show systematic variations in time, which cannot be described by the regular day-to-day variation alone. Such variations can have different causes, for example, seasonal effects, changing weather conditions, or road works.

From a visual inspection of the 10-min time series, we suspect that a large amount of the variation between successive time intervals is random. This variation is called noise. The amount of noise is an important quantity. If the amount of noise increases, systematic variations can less easily be detected. It also gives a lower limit for the predictive power, because noise cannot be predicted. Noise can have different causes. It can be caused by the random arrival process of cars. This process results in different headways between the following cars, which is an important source of variation on highways. In urban areas, traffic flows are interrupted by traffic signals. In this case, variable and unknown green times of these signals may contribute to the noise. In practice, all variations that have short time scales and that do not follow a recurrent pattern can be considered to be noise.

A measurement on day d, at link l, and in time interval t can thus be described in the following terms:

$$q_{dlt}^{\text{obs}} = q_{dlt}^{\text{pred}} + \varepsilon_{dlt} + \nu_{dlt} \quad (2)$$

with q_{dlt}^{pred} being the predicted volume (e.g., the baseline prediction), ε_{dlt} being the systematic variation between measurement and prediction, i.e., the prediction error, and ν_{dlt} being the noise on day d, at link l, and in time interval t.

The objective is to develop a prediction scheme that minimizes the systematic variation or prediction error. There are two extreme approaches to reach this objective. First, the external processes that lead to systematic variations can be studied in detail, so that the relation between the two can be modeled (e.g., the relation between weather and travel demand). The advantage of this approach is that it provides insight into the variation of travel demand. The disadvantage is that it is complicated and requires many reliable data sources, which are often not available. Another approach is a black-box approach. In this approach, correlations in historical data are found by certain mathematical techniques (e.g., neural networks), and these correlations are used in the prediction scheme.

In this paper, we apply an intermediate approach. Our method is based on the following assumption. The single most important temporal correlation in the systematic variation is that between *successive* epochs (which can have different time scales). In particular, we assume that, due to seasonal effects, there is a positive correlation between the systematic variations, i.e., ε_{dlt} and $\varepsilon_{d+1,lt}$, of successive days. If there is more traffic than average on a particular day, then the probability is high that the next day will also show more traffic.

The improvement of the baseline prediction is quite simple in this case. The relative systematic variation c (with $\varepsilon = cq^{\text{pred}}$) results from the ratio between the observation and the baseline prediction. Suppose that this ratio is 1.10, i.e., c is estimated to be 10%. Depending on the strength of the correlation between the relative systematic variations of successive days, the updated 24-h prediction for the next day then lies between 1.00 (in case of no correlation) and 1.10 (in case the correlation coefficient is 1) times the baseline prediction of that day. In Section IV, we explain how we optimize the 24-h prediction.

We also assume a positive correlation between the systematic variations, i.e., ε_{dlt} and $\varepsilon_{dl,t+1}$, of successive time intervals. If there is more traffic than normal during the present time interval, then there will also probably be more traffic during the next time interval. In fact, we assume that, in general, this correlation $\rho(\varepsilon_{dlt}, \varepsilon_{dlt+1}) \approx 1$. The logic behind this assumption is that systematic variation slowly changes. Their time scales can be longer than a day (e.g., seasonal variations) and are often longer than an hour (e.g., weather-related variations). This is illustrated by the systematic difference between the measurement and the baseline prediction in the upper panel of Fig. 2.

The update of the 24-h prediction is called the short-term prediction. The estimate of this prediction is comparable to that of the 24-h prediction. The relative systematic variation c, which is estimated in the present and previous time intervals, is used to update the prediction for the next time interval(s). In Section V, we explain how we optimize the short-term prediction.

There may also exist a spatial correlation between two links l and k that are incident, i.e., ε_{dlt} and ε_{dkt}. If there is more traffic than normal on a certain link, then there probably will also be more traffic on a neighboring link. However, we did not consider these spatial correlations, because the spatial sampling of links is limited in our sample.

In the previous paragraphs, we explained the basic concept of our prediction scheme, which will be described in more detail in the next sections. We do not consider other correlations, nor do we try to model these covariances.

The update of the short-term prediction is complicated due to the noise in the measurements. The most important characteristic of this noise is that the noise in different measurements is uncorrelated so that $\rho(\nu_i, \nu_j) = 0$ for $i \neq j$. This implies that, for aggregated measurements, the variance of the noise adds up. The total volume is proportional to the number of measurements (each measurement within a fixed time interval). Thus, the variance in the noise is proportional to the expected volume for a given traffic regime. In fact, $q^{\text{pred}} + \nu$ is Poisson distributed when the volumes are small [18]. In that case, for $\text{var}(\nu)$, which is the variance in ν, it holds that

$$\text{var}(\nu) = q^{\text{pred}}. \tag{3}$$

The root mean square (RMS) of the residuals, i.e., $q^{\text{obs}} - q^{\text{pred}}$, is an indicator for the quality of the prediction scheme. In Sections IV and V, we will present predictions for which we tried to minimize the RMS of the residuals. However, the residuals also contain noise. For a fair evaluation of the prediction scheme, we therefore need to separate the prediction error (remaining systematic variation) from the noise. This is done in Section VI.

According to the previous approximation of the noise, the mean square of the residuals can be approximated by

$$\frac{1}{n}\sum_{dlt}\left(q_{dlt}^{\text{obs}} - q_{dlt}^{\text{pred}}\right)^2 = (c\bar{q}^{\text{pred}})^2 + \bar{q}^{\text{pred}} \tag{4}$$

where $\bar{q}^{\text{pred}} = \frac{1}{n}\sum_{dlt} q_{dlt}^{\text{pred}}$, and n is the total number of residuals. The left-hand side of the equation describes the total quadratic variation with respect to the prediction. The first term of the right-hand side describes the total quadratic systematic variation, and the second term describes the variance of the noise. Thus, the (average) relative prediction error c can be estimated from (4).

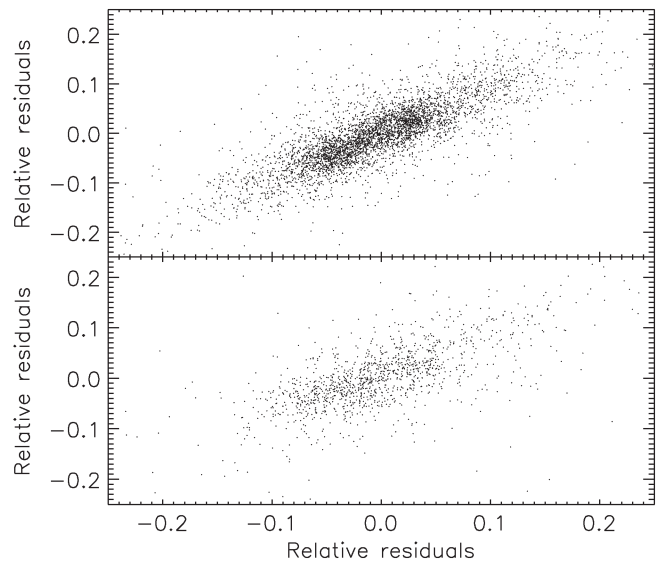

Fig. 3. Daily volume residuals (relative to the baseline prediction) for successive working days. (Top) Successive weekdays. (Bottom) Mondays versus Fridays.

The assumptions and findings about systematic variation and noise form the basis of the prediction scheme described in this paper. However, it should be stressed that these are assumptions. Equation (3) is an approximation. It is plausible that the variation in green time increases toward intersections that serve high volumes, e.g., due to the fact that the number of signal cycles per time lag decreases. In that case, the contribution to the "noise" will increase. For large volumes, we therefore expect $\text{var}(\nu) > q^{\text{pred}}$. As a result, (3) probably gives a lower limit for the variance of the noise. Moreover, the correlation between systematic variations in successive time intervals cannot always be close to 1, because systematic differences between predictions and measurements are not constant all the time. Finally, there may be other correlations than those between successive epochs. However, in the following sections, we show that our assumptions lead to a prediction method that is simple and effective.

IV. Twenty-Four-Hour Prediction

In the previous section, we assumed that systematic variations in traffic flows are correlated on successive days. In Fig. 3, we show that this correlation indeed exists for urban volumes in Almelo. Per link, we show the relative residuals (between baseline predictions and measurements) of daily traffic volumes for successive working days. The top panel shows the residuals for successive weekdays. The bottom panel shows the residuals for successive working days that are not successive weekdays, i.e., Fridays versus Mondays. In both cases, there is a positive correlation between the residuals (correlation coefficients of 0.81 and 0.64, respectively).

As explained in the previous section, this correlation implies that we can use the ratio between the measurement and the baseline prediction of the present day to update the prediction for the next day. This was done as follows:

$$q^{24}_{d+1lt} = q^{\text{base}}_{d+1lt} \left(\frac{\sum_{t'=t-s}^{t+s} q^{\text{obs}}_{dlt'}}{\sum_{t'=t-s}^{t+s} q^{\text{base}}_{dlt'}} \right)^{p}. \tag{5}$$

In (5), the 24-h prediction q^{24}_{d+1lt} for day $d+1$, link l, and time interval t is equal to the baseline prediction multiplied by an update factor. This update factor is an exponent p of the ratio between the (central moving) averages of measurements and baseline predictions from day d. The central moving averages are taken in the interval $[t-s, t+s]$ with $2s$ as the box width. To limit the disturbing influence of noise in the update, the box width of this interval was chosen to be at least 1 h.

The exponent p lies between 0 and 1. If there would be no correlation between the residuals of two successive days, it would be impossible to update the prediction with measurements from the previous day $(p = 0)$. On the other hand, if this correlation is equal to 1, the prediction is simply multiplied by the ratio between the measurement and the baseline prediction $(p = 1)$.

By trying different values for p and s, we found that the RMS of the residuals for the 24-h prediction is relatively low when the box width for the central moving average is larger than 1 h and when p lies between 0.5 and 1.0. Although results within these ranges do not very strongly vary, we conclude that the RMS of the residuals was minimized for our data set when the box width is 3 h $(s = 9)$ and $p = 0.8$ and 0.5 for successive and nonsuccessive weekdays, respectively. Note that p is smaller for nonsuccessive weekdays, because the correlation between the residuals is weaker for these days (see the bottom panel of Fig. 3).

The 24-h prediction is, in general, updated by volumes from the previous day (with $p = 0.8$). The exceptions are Mondays and Saturdays, because a Monday follows the weekend, and a Saturday follows a working day. The prediction for these days were updated with the volumes of the previous Friday and Sunday, respectively (with $p = 0.5$). It is also possible to update these predictions with the volumes of the previous Monday and Saturday, respectively. We did not find significant differences between these different updates.

V. Short-Term Prediction

In addition to day-to-day variations and seasonal variations, there may be other variations that have shorter time scales (for example, weather-related variations). These variations may not be included in the 24-h prediction. Actual measurements and 24-h predictions, however, can be combined to update the prediction for the short term, e.g., [12].

As explained in Section III, a positive correlation between residuals in successive intervals implies that we can use the ratio between the measurement and the 24-h prediction of the previous interval(s) to update the prediction for the next interval(s). Comparable with (5), we could update the short-term prediction

$$q^{\text{st}}_{dlt+T} = q^{24}_{dlt+T} \left(\frac{\sum_{t'=t-r}^{t} q^{\text{obs}}_{dlt'}}{\sum_{t'=t-r}^{t} q^{24}_{dlt'}} \right)^{h(T)}. \tag{6}$$

In (6), the short-term prediction q^{st}_{dlt+T} on day d, at link l, and in time-interval $t+T$ is equal to the 24-h prediction multiplied by an update factor. This update factor is a power $h(T)$ of the ratio between the (moving) averages of measurements and 24-h predictions from the previous time intervals. The number of time intervals is defined by the interval $[t-r, t]$ with $r \geq 0$. The prediction is for the near future with time horizon $T \geq 1$ (with one unit corresponding to 10 min in this case).

The exponent $h(T)$ lies between 0 and 1. If there would be no correlation between the residuals of successive intervals, it would be impossible to update the prediction for the next interval with measurements from the previous interval(s). In that case, $h(1) = 0$. On the other hand, if this correlation would be equal to 1, the prediction is simply multiplied by the ratio between the measurement and the 24-h prediction $(h(1) = 1)$. The power $h(T)$ is also decreasing with T, because the correlation between the residuals will become weaker for larger time lags. Beyond a certain time horizon $h(T) = 0$, the short-term prediction will be equal to the 24-h prediction.

The problem with this prediction and with any extrapolation method that uses moving averages (how sophisticated they might be) is that the measurements contain noise that contaminates the prediction. Noise can be reduced when the length of the interval $[t-r, t]$ is increased. Unfortunately, by increasing the length of this interval, the correlation between historical flows and future flows will decrease. This will diminish the quality of the short-term predictions as well. In fact, we could not find any interval length r for which the prediction was an improvement compared with the 24-h prediction.

To tackle this problem, we tried to filter the noise. For this purpose, we used a Kalman filter [19]. The principle of this filter is that the noise in the measurements is smoothed by expected model values, which are given by a state equation. The measurements are given by the measurement equation

$$q^{\text{obs}}_{dlt} = q_{dlt} + \nu_{dlt} \tag{7}$$

with q_{dlt} being the true volume and ν_{dlt} being the measurement noise on day d, at link l, and in time interval t. We used the following state equation, in which we estimated the volume of the next time step by the expected increase (or decrease) in the 24-h prediction:

$$q^{\text{est}}_{dlt} = q^{\text{kal}}_{dlt-1} + \left(q^{24}_{dlt} - q^{24}_{dlt-1} \right). \tag{8}$$

The filtered volume q^{kal}_{dlt} is calculated by taking the linear combination of the state estimate q^{est}_{dlt} and the measured volume q^{obs}_{dlt}, in which the total variance due to model errors and measurement noise is minimized (for details, see [19]). For this recipe, we need an estimate for the variance of the noise R_{dlt}

and for the variance of the model error Q_{dlt}. Because the state equation does not contain other variables besides the volume q, no covariances need to be considered. According to (3), the variance of the measurement noise can be approximated by the expected volume. Thus

$$R_{dlt} = q^{24}_{dlt}. \tag{9}$$

Following the reasoning behind (3) and (4), we estimated the variance of the model error in the following way:

$$Q_{dlt} = \left(c' q^{24}_{dlt}\right)^2 + \left(q^{24}_{dlt-1} + q^{24}_{dlt}\right)/N_D. \tag{10}$$

The model error contains some measurement noise, because it is based on historical measurements. This is described by the second term in (10). If the number of days N_D in group D is large, this noise term becomes negligibly small. The first term in (10) describes the systematic model error due to imperfections in the 24-h prediction. According to the state equation (8), we look for the error in the difference between two successive predictions. The variance due to the (systematic) error in a single prediction is given by the first term of the right-hand side of (4). Similarly, the first term of the right-hand side of (10) gives the variance due to the error in the difference between two successive predictions.

Because the relative error in a single prediction c only varies slowly with time, we expect that the relative error in the difference c' should be small. However, it is almost impossible to directly estimate c'. We therefore indirectly estimated c' by trying to minimize the RMS of the residuals in the short-term prediction (see below), i.e., we tried different values for c'. We found low RMS values when c' lies around 0.01 (1%) for working days and around 0.1 (10%) for weekends. The estimate of c' thus appears to be dependent on the day of the week. However, we also find that residuals in short-term predictions are not very sensitive to variations in c'. For values of c' between 0.01 and 0.1, the RMSs of the residuals are comparable, and they are all significantly smaller than the RMS of the 24-h prediction residuals. We set c' at a fixed value of 0.03, because we found that the overall RMS of the residuals in the short-term prediction is minimal for this value. We stress that, since c' is different for various groups of days, the chosen value can hardly be seen as a "fit" to the data.

Given the filtered data, we tried different values for the maximum time horizon and for the power $h(1)$, which is required for the prediction of the next time interval. From this, i.e., by minimizing the RMS in the residuals, we derived the following equation for the short-term prediction:

$$q^{\mathrm{st}}_{dlt+T} = q^{24}_{dlt+T} \left(\frac{\sum_{t'=t-5}^{t} q^{\mathrm{kal}}_{dlt'}}{\sum_{t'=t-5}^{t} q^{24}_{dlt'}} \right)^{0.8-0.1T}. \tag{11}$$

This equation is comparable with (6), but here, the raw measurements $q^{\mathrm{obs}}_{dlt'}$ are replaced by the filtered measurements $q^{\mathrm{kal}}_{dlt'}$. The power $h(T) = 0.8 - 0.1T$. The correlation between systematic variations in two time intervals decreases with the length of the time lag between these intervals. As a consequence, the power of the update is maximal for the shortest prediction horizon (next time interval) and gradually decreases with prediction horizon T. The maximum horizon is eight intervals, i.e., in 80 min, the prediction will be equal to the 24-h prediction. The length of the historical time interval $[t-r, t]$, which is used to update the short-term prediction, is 1 h (six intervals). Note that the RMS of the residuals for the short-term prediction is not very sensitive to the power $h(T)$. The RMS is relatively low when $h(1)$ lies between 0.5 and 1.0 and when the maximum horizon lies between 50 and 100 min.

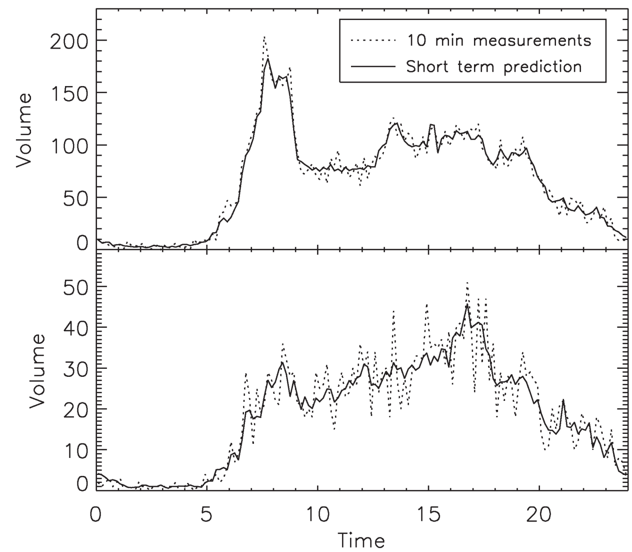

Fig. 4. Two examples of predictions with a 10-min horizon compared with the measurements. The measurements are 10-min aggregates of raw data.

In Fig. 4, we show two examples of predictions with a 10-min horizon. The predictions appear to follow the measurements quite well. In both cases, the remaining variation is mainly caused by the measurement noise.

VI. Accuracy of Predictions

It was already mentioned in Section III that, for this evaluation, we do not separate between a test and a training set. It was argued that, for a baseline prediction, this will not cause any big problems. For a 24-h prediction and a short-term prediction, the consequences are also limited. Although these predictions are direct or indirect updates of the baseline prediction, they are otherwise derived in exactly the same way as in a real-time prediction. We affirm that we used the whole data set to fine-tune the free parameters in these predictions without using an independent data set to test them. However, the number of free parameters is very small compared to the number of profiles. Moreover, we used the same values for all predictions, despite the fact that there are significant differences between different days, links, and times of the day. We therefore argue that the results from the evaluation in Section VI are also valid for real-time predictions.

Predictions can be validated by an *a posteriori* comparison between predictions and measurements. In the previous sections, we used a standard measure for validation, i.e., the

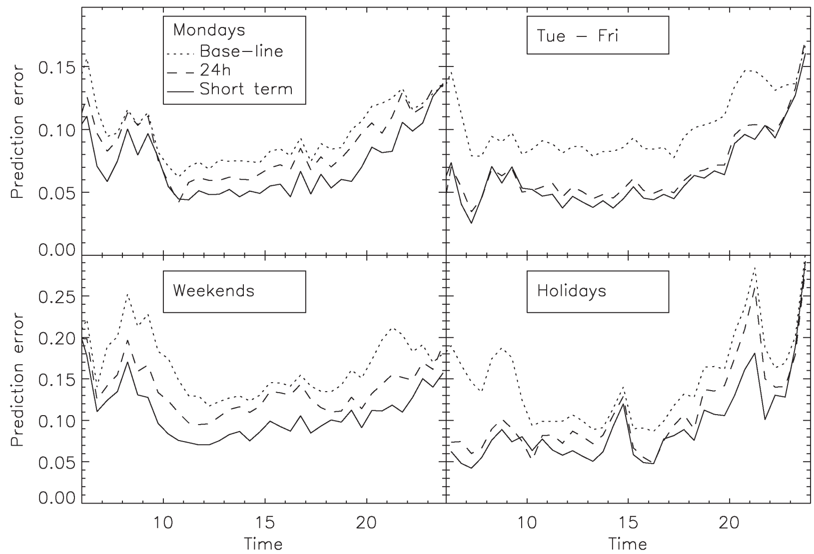

Fig. 5. Relative prediction errors in baseline predictions, 24-h predictions, and short-term predictions for Mondays, Tuesdays to Fridays, weekends, and the holiday period.

RMS of the residuals. Another measure is the mean absolute value of the residuals. Often, relative values (with respect to the prediction itself) are also given. Relative measures can be put in percentages, and they give complementary information. Absolute measures are often called absolute prediction errors, while the relative measures are called relative prediction errors.

Unfortunately, these measures are not real prediction errors, because the residuals also contain (measurement) noise [ν in (2)]. The standard deviation in the noise is proportional to the square root of the expected volume. The absolute "prediction error" therefore increases with the volume, while the relative "prediction error" decreases with the volume. In other words, the "prediction error" depends on the magnitude of the link flow. This result is nicely illustrated in Fig. 4. The volumes in the upper panel are four times larger than those in the bottom panel. The relative variation in the residuals is indeed significantly smaller in the upper panel, while the absolute variation is significantly smaller in the bottom panel. In fact, the real prediction errors [ε in (2)] are negligibly small compared with the noise, and the quality of both predictions is actually quite comparable. The measures that are referred to as absolute and relative prediction errors in the literature might as well be considered to be measures for the noise. They are therefore not suitable to evaluate the quality of a prediction scheme.

The noise and prediction errors are uncorrelated so that both variances add up, which is indicated by (4) in Section III. According to this equation, the average quadratic prediction error can thus be estimated as the difference between the mean square of the residuals (total variance) and the variance of the noise. For a fair evaluation of the prediction scheme, it is therefore crucial to have some knowledge about the amount of noise. In Section III, we explained that the noise of the first order can be approximated by a Poisson distribution but that this estimate is probably a lower limit. However, without a better estimate, we used this to estimate the prediction errors.

In Fig. 5, we show the average relative prediction errors for Mondays (upper left), Tuesdays until Fridays (upper right), weekends (bottom left), and school holidays (bottom right). For a smooth result (to minimize the effects of noise in the figure), we show the prediction errors for 30-min aggregates (in which predictions with a 30-min horizon are included). From Fig. 5, we can conclude that both short-term predictions (solid lines) and 24-h predictions (dashed lines) are significantly more reliable than baseline predictions (dotted lines). Short-term predictions are significantly more reliable than 24-h predictions on Mondays, on weekends, and during the holiday period, but the two predictions are quite comparable for the other working days. Predictions are probably less reliable on Mondays, because they cannot be updated by traffic information from the previous day (which would be a Sunday). Note that, even for short prediction horizons, it is difficult to estimate reliable predictions for the Monday morning rush hour.

Apart from the Monday morning rush hour, estimated errors of short-term predictions are around 5% during the working day, while predictions become less reliable in the evening period and in the weekends. This is not very surprising, since less predictable recreational traffic is dominant in these periods.

In theory, predictions can be improved when errors are around 5%. However, as mentioned before, the noise is probably underestimated. The prediction errors may therefore be overestimated. The predictions could be considered to be optimal when no systematic trends (larger than 10 min) are left in the residuals. This would be the case if residuals of successive time intervals are uncorrelated. The randomness of residuals is usually tested by a Ljung–Box test. The null hypothesis states that the residuals in the time series are uncorrelated. We performed the Ljung–Box test and estimated in how many cases the null hypothesis is rejected. We used a 95% confidence level. We found that the fraction of rejected profiles is about 90% for the baseline prediction, about 80% for the 24-h prediction, and less than 30% for the short-term prediction. In other words, for about 70% of the short-term predictions, there is no evidence that systematic variation is left in the residuals.

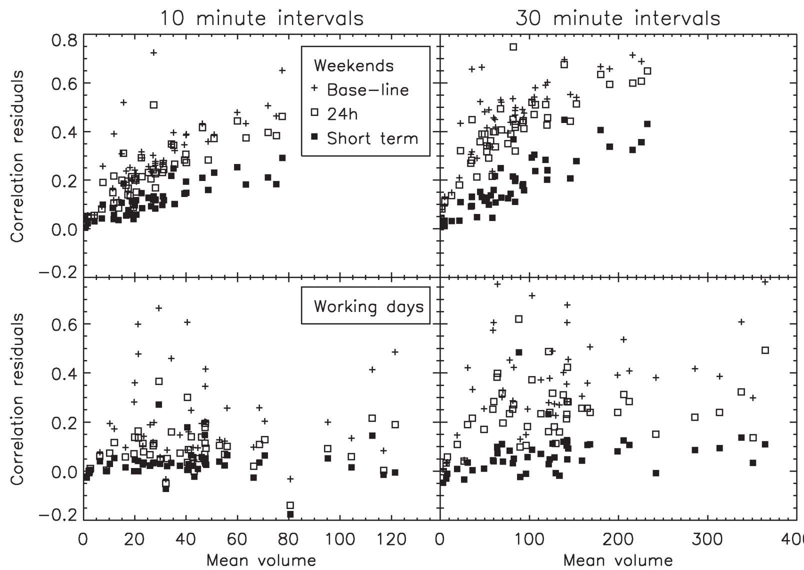

Fig. 6. Correlation between residuals of successive time intervals for baseline, 24-h, and short-term predictions during (upper panel) weekends and (bottom panel) working days. Left panel: 10-min time series. Right panel: 30-min time series. The symbols represent link values.

The Ljung–Box test also includes the autocorrelation of long time lags. Because residuals with long time lags are less correlated or uncorrelated, they may "overshadow" the correlation for short time lags. We therefore separately considered the autocorrelation of residuals with a time lag of one interval, i.e., we analyzed the correlation of successive residuals. Because the systematic variation relatively slowly changes, it is quite likely that, when the prediction is systematically too high in one interval, the systematic component will also be too high in the next interval. Thus, in that case, we would expect that successive residuals are positively correlated.

In Fig. 6, we show the correlation between the successive residuals. We distinguished between different links, so that for each link, we got an estimate of the average correlation. The results are shown for weekends (upper panel) and working days (bottom panel). We show the correlation for both 10-min time series (left panel) and 30-min aggregates (right panel). It can be seen that, for the short-term predictions (solid symbols), most links show a very small positive correlation during working days. For the weekends and for all baseline predictions (crosses), correlations between successive residuals are significantly larger (in some cases, with correlation coefficients of 0.8). The figure also nicely illustrates that the 24-h prediction (open symbols) is in between the baseline and the short-term prediction.

Since the number of residuals is very large (the number of days times the number of time intervals), even small correlations are still significant. There will always be some correlation left in the residuals, because no prediction scheme is perfect. We therefore do not claim that the residuals of short-term predictions are random. However, from Figs. 5 and 6, we can conclude that our short-term and 24-h predictions are an important improvement compared with baseline predictions. Moreover, apart from weekends, evenings, and the Monday morning rush hour, errors in short-term predictions and, to a lesser extent, 24-h predictions are negligibly small.

VII. Conclusion

We have developed a prediction scheme based on volume data that were collected at about 20 urban intersections in the Dutch city of Almelo.

Our prediction scheme consists of three steps. In the first step, we have made a baseline prediction for a given day, location, and time interval. We have classified days into working days (Monday to Friday), weekends, and the school holiday period and defined the baseline prediction as the historical (yearly) average over all days belonging to the same group. In the second step, we have made a prediction for 24 h ahead, in which the baseline prediction is updated by the profile from a previous day. In the third step, flows from the previous hour(s) have been used to update predictions for the short term.

We have found that prediction errors in short-term and 24-h predictions are significantly lower than those in baseline predictions. We have also found that predictions are less reliable for a Monday morning rush hour, evenings, and weekends. The Monday morning rush hour is less predictable, because it is the start of a new week for commuters. In the evenings and weekends, traffic is dominated by people with less-predictable recreational motives. However, apart from these periods, we have found that short-term predictions have negligibly small errors.

Prediction of link flows can be applied for the management of traffic-control systems, e.g., [20] and [21]. Predictions can be used to optimize intersection traffic light split times. In fact, some authors already have developed traffic control systems that can adapt to a changing travel demand, e.g., [22] and [23]. In these cases, artificial neural networks are being used. The prediction scheme that is introduced in this paper may also be used for that purpose.

VIII. Discussion

Our approach is useful for several reasons. First, the explicit assumptions about temporal correlations in link flows appear to be logical, and they therefore provide a simple method that produces reliable predictions. We did not use spatial correlations, because the spatial sampling of links is limited in our sample. We therefore concluded that the inclusion of spatial correlations would hardly improve our predictions. Although the predictions significantly improve when we adopt simple temporal correlations, this does not mean that the method can automatically be used for other traffic parameters like travel time, because these parameters may show different correlations.

Second, in this paper, we explicitly separated systematic variation from noise. Such an approach is not common in this research area. Current validation measures, which do not separate the noise from the prediction error, cannot be considered to be appropriate validation measures for prediction schemes. We have also shown that predictions can be improved when the (knowledge about the amount of) noise is taken into account. If the noise is not taken into account, extrapolation methods may deliver noisy predictions with an inferior quality.

An important issue is how the algorithm performs under different traffic conditions. Although, most of the time, congestion levels in the city of Almelo are low, many different traffic situations occur. In our sample, both periods with light traffic and rush hours are included. The sample also contains low- and high-capacity roads, which serve low and high flows, respectively. Moreover, the signaling can be different from intersection to intersection. In some cases, the signals are fixed and have quite long cycle times. In other cases, the signals are vehicle actuated. In these different situations, noise levels may also be different. Nevertheless, according to Fig. 5, the quality of the prediction is similar during rush hours when the loads are heavy and during the off peak when traffic is lighter. According to Fig. 6, the quality of the predictions during working days is also similar for roads that serve little traffic and high-capacity roads that tend to have signaling with longer cycle times.

The main difference in the quality of the prediction is between working days and weekends. The demand in weekends is mainly caused by recreational traffic, which has fewer regular patterns. It appears, however, that the prediction algorithm shows similar results for the aforementioned different traffic conditions.

It is quite remarkable that the 24-h prediction is such a strong improvement compared with the baseline prediction. The quality of this prediction only depends on the strength of the correlation between the demands on successive days. The predictions are very reliable for Tuesdays, Wednesdays, Thursdays, and Fridays, because these days follow days with similar demand patterns. Because of weekly variations and the fact that Saturdays, Sundays, and Mondays do not show similar demand profiles, 24-h predictions are less reliable for weekends and Mondays.

The success of the 24-h prediction is the result of the slow-changing systematic variation in demand, which does not depend on the traffic condition. The estimate of the 24-h prediction also hardly depends on the noise levels. This is different for the short-term prediction, in which we explicitly used noise levels to improve the prediction. Although these noise levels may vary for different traffic situations, we still find that the short-term prediction is almost always a significant improvement compared with the 24-h prediction.

The robustness of the short-term prediction can be explained as follows. The effectiveness of the Kalman filter depends on the estimated ratio of the noise and systematic error in the state equation. The latter one was described by c' in (10). The exact noise level does not have to be known, because c' is also unknown. We chose c' so that the short-term prediction became optimal. It is thus possible to improve the prediction without an exact estimate of the noise. Although it is quite likely that the optimal value of c' depends on the traffic situation, we found that the quality of the short-term prediction is not very sensitive to some variation in c'. The results show that we are able to improve the short-term prediction under different circumstances when we at least know the order of magnitude for the noise.

Although the algorithm appears to be robust under different traffic situations, it should actually be tested during structural congestion. We were not able to validate our algorithm for congested areas, because these measurements are lacking in our data set. However, the algorithm might be tested in other studies that provide traffic count predictions in large metropolitan areas.

Throughout this paper, we optimized our prediction scheme by trying different values for the free parameters. From a theoretical point of view, one might consider introducing more generic optimizing algorithms. In this paper, our main objective was to show that reliable predictions are possible by using this practical approach. However, we do not consider this paper as finished but rather as a starting point for exploring more applications, for example, travel time predictions. It is possible that for those applications a more generic extension of this algorithm, e.g., by including Bayesian inference statistics, is desirable.

Acknowledgment

The authors would like to thank Vialis and the municipality of Almelo for providing the flow volume data in Almelo.

References

[1] A. Dharia and H. Adeli, “Neural network model for rapid forecasting of freeway link travel time,” *Eng. Appl. Artif. Intell.*, vol. 16, no. 7/8, pp. 607–613, Oct.–Dec. 2003.

[2] J. W. C. Van Lint, S. P. Hoogendoorn, and H. J. Van Zuylen, “Robust and adaptive travel time prediction with neural networks,” in *Proc. 6th Annu. Transp., Infrastruct. Logist. Congr.*, Delft, The Netherlands, 2000.

[3] S. I. Bajwa, E. Chung, and M. Kuwahara, “An adaptive travel time prediction model based on pattern matching,” in *Proc. 11th World Congr. Intell. Transp. Syst.*, Nagoya, Japan, 2004.

[4] M. Danech-Pajouh and M. Aron, “ATHENA: A method for short-term inter-urban motorway traffic forecasting,” *Recherce Transports Sécurité*, vol. 6, pp. 11–16, 1991. English issue.

[5] P. Hasberg and D. Serwill, “Stadtinfoköln—A global mobility information system for the Cologne area,” in *Proc. 7th World Congr. Intell. Transp. Syst.*, Turin, Italy, 2000.

[6] A. Kellerman and A. Schmid, “Mobinet: Intermodal traffic management in Munich-control centre development,” in *Proc. 7th World Congr. Intell. Transp. Syst.*, Turin, Italy, 2000.

[7] B. Leitsch, “A public-private partnership for mobility—Traffic management Center Berlin,” in *Proc. 9th World Congr. Intell. Transp. Syst.*, Chicago, IL, 2002.

[8] R. Camus, G. E. Cantarella, and I. Domenico, "Real-time estimation and prediction of origin-destination matrices per time slice," *Int. J. Forecast.*, vol. 13, no. 1, pp. 13–19, Mar.–May 1997.
[9] R. Chrobok, O. Kaumann, J. Wahle, and M. Schreckenberg, "Different methods of traffic forecasting based on real data," *Eur. J. Oper. Res.*, vol. 155, no. 3, pp. 558–568, Jun. 2004.
[10] Y. Kamarianakis and P. Prastacos, "Forecasting traffic flow conditions in an urban network: Comparison of multivariate and univariate approaches," *Trans. Res. Rec.*, vol. 1857, pp. 74–84, 2003.
[11] R. Van Grol, D. Inaudi, and E. Kroes, "On-line traffic condition forecasting using on-line measurements and a historical database," in *Proc. 7th World Congr. Intell. Transp. Syst.*, Turin, Italy, 2000.
[12] D. Wild, "Short-term forecasting based on a transformation and classification of traffic volume time series," *Int. J. Forecast.*, vol. 13, no. 1, pp. 63–72, Mar. 1997.
[13] M. Van de Voort, M. Dougherty, and S. Watson, "Combining Kohonen maps with ARIMA time series models to forecast traffic flow," *Transp. Res., Part C Emerg. Technol.*, vol. 4, no. 5, pp. 307–318, Oct. 1996.
[14] H. Yin, S. C. Wong, J. Xu, and C. K. Wong, "Urban traffic flow prediction using a fuzzy-neural approach," *Transp. Res., Part C Emerg. Technol.*, vol. 10, no. 2, pp. 85–98, Apr. 2002.
[15] H. Dia, "An object-oriented neural network approach to short-term traffic forecasting," *Eur. J. Oper. Res.*, vol. 131, no. 2, pp. 253–261, Jun. 2001.
[16] W. A. M. Weijermars, "Analysis of urban traffic patterns using clustering," Ph.D. dissertation, Univ. Twente, Enschede, The Netherlands, 2007.
[17] E. Chung, "Classification of traffic pattern," in *Proc. 10th World Congr. Intell. Transp. Syst.*, Madrid, Spain, 2003.
[18] T. Thomas, W. A. M. Weijermars, and E. C. van Berkum, "Variations in urban traffic volumes," *Eur. J. Transp. Infrastruct. Res.*, vol. 8, no. 3, pp. 251–263, Sep. 2008.
[19] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME, J. Basic Eng.*, vol. 82, no. 1, pp. 34–45, 1960.
[20] X. Wang, W. Cottrell, and S. Mu, "Using k-means clustering to identify time-of-day break points for traffic signal timing plans," in *Proc. 8th IEEE Conf. Intell. Transp. Syst. Conf.*, Vienna, Austria, 2005, pp. 586–591.
[21] B. L. Smith, W. T. Scherer, and T. A. Hauser, "Data-mining tools for the support of signal-timing plan development," *Trans. Res. Rec.*, vol. 1768, pp. 141–147, 2001.
[22] J. S. Yang, "Traffic signal timing control for a small-scale road network," *Control Appl.*, vol. 441, pp. 48–49, 2004.
[23] Z. S. Yang, X. Chen, Y. S. Tang, and J. P. Sun, "Intelligent cooperation control of urban traffic networks," in *Proc. 4th Int. Conf. Mach. Learn. Cybern.*, Guangzhou, China, 2005, pp. 1482–1486.

Tom Thomas was born in 1972. He received the M.Sc. and Ph.D. degrees in astronomy from the University of Leiden, Leiden, The Netherlands. His thesis was about the evolution of galaxies in rich nearby clusters.

After working as a Consultant, in 2006, he started his career in traffic engineering with the University of Twente, Enschede, The Netherlands, where he is currently with the Centre for Transport Studies. His main topics of interest are in the fields of travel behavior and dynamic traffic management.

Wendy Weijermars was born in 1977. She received the M.Sc. and Ph.D. degrees in civil engineering from the University of Twente, Enschede, The Netherlands. Her Ph.D. research was on analyzing urban traffic patterns using clustering. The research described in this paper partly elaborates on her Ph.D. research.

She is currently with the Institute for Road Safety Research, The Netherlands.

Eric van Berkum was born in 1959. He received the M.Sc. degree in applied mathematics from the University of Twente, Enschede, The Netherlands, and the Ph.D. degree in civil engineering from the Technical University of Delft, Delft, The Netherlands. His Ph.D. research was on the impact of traffic information on route and departure time choice.

He was a Software Engineer with Cap Gemini and a Transport Consultant with Goudappel Coffeng. Since 1998, he has been a Professor of traffic management with the Centre for Transport Studies, University of Twente. His main topics of interest are in the areas of travel behavior, transport modeling, and dynamic traffic management.