Short-term Forecasts for Transport Models

Prévisions à court terme pour les modèles de transport

Kurzfristprognosen für Verkehrsmodelle

Prof. Dr. C. de Rham, International University of Monaco

B+S AG, Bern
R. Schwarz, Dipl. Ing. ETH
W. Schaufelberger, Dipl. Ing. ETH

Research Contract ASTRA 2006/019 commissioned
By The Federal Roads Office (FEDRO)

December 2007

Eidgenössisches Departement für Umwelt, Verkehr, Energie und Kommunikation UVEK
Bundesamt für Strassen ASTRA

Short-term Forecasts for Transport Models

Prévisions à court terme pour modèles de transport

Kurzfristprognosen in Verkehrsmodelle

Forschungsauftrag: ASTRA 2006/019

Forschungsstelle: Prof. Dr. C. de Rham, International University of Monaco

B+S AG, Bern
R. Schwarz, Dipl. Ing. ETH
W. Schaufelberger, Dipl. Ing. ETH

Begleitende Kommission:
Dörnenburg, Klaus, Sigmaplan AG, Bern, Präsident
Dr. Bischofberger, Nikolaus, Volkswirschartsdirektion Kt Zürich, Zürich
Petersen, Gerhard, ASTRA, Bern
Dr. von der Ruhren, Stefan, momatec GmbH, Aachen
Siegrist, Roger, ASTRA, Bern

CONTENT

In 2 words / en 2 mots / in 2 Worten

GB: Recipe to increase the predictive power of short-term forecasts

Initialization, preparation of a set of typical DVC (daily variation curves):

1 Extract 50-100 DVC from sets of single DVC with cluster analysis

2 Smooth the resulting DVC to eliminate random variations

Online, for each time step (3' – 5') and each link of the network:

3 Find the 8-12 best fitting DVC for the last 2 hours, calculate a mean DVC

4 Adjust the mean DVC at t0 with the 2-4 last time intervals

5 Use the adjusted mean DVC as forecast for the next two hours

Feedback, once per day:

6 Update the set of typical DVC with the new measured DVC of the day

The forecast quality for two hours lies around 10% MRE (mean relative error)

F : Recette pour augmenter la puissance prédictive de prévisions à court terme

Initialisation, préparation d'un ensemble de CVJ (courbe de variation journalière) typiques:

1 Extraire 50-100 CVJ à partir de CVJ individuelles grâce à la cluster analyse

2 Lisser ces CVJ pour éliminer les fluctuations aléatoires

En ligne, pour chaque intervalle de temps (3' – 5') et chaque tronçon du réseau:

3 Trouver les 8-12 CVJ les mieux adaptées aux 2h passées, calculer la CVJ moyenne

4 Ajuster la CVJ moyenne à t0 grâce aux 2-4 derniers intervalles de temps

5 Utiliser la CVJ moyenne ajustée comme prévision pour les 2 prochaines heures

Feedback, une fois par jour:

6 Mise à jour de l'ensemble de CVJ typiques avec les nouvelles CVJ de la journée

La précision des prévisions à 2 heure est d'environs 10% MRE (mean relative error)

D: Rezept für die Erhöhung der prädiktiven Leistung von Kurzfristprognosen

Initialisierung, Vorbereitung einer Menge typischer TGL (TagesGangLinien):

1 Extrahieren von 50-100 TGL aus Mengen einzelner TGL mit Cluster Analyse

2 Glättung dieser TGL um zufällige Abweichungen zu eliminieren

On line, für jedes Zeitintervall (3' - 5') und jede Strecke im Netz:

3 Finden der 8-12 best passenden TGL für die letzten 2h, mittlere TGL berechnen

4 Justieren dieser mittleren TGL für t0 mit den 2-4 letzten Zeitintervallen

5 Prognose für die nächsten 2 h basierend auf dieser justierten mittleren TGL

Feedback, ein mal pro Tag :

6 Update der Menge typischer TGL mit den während dem Tag gemessenen TGL

Die Qualität der Prognose auf 2 Stunden liegt um 10% MRE (mean relative error)

Summary

The new national traffic management centre VM-CH (Verkehrsmanagement Schweiz) will start its operations in 2008 and cover the entire Swiss motorway network. An efficient management is possible only if one knows the actual traffic situation and if one is able to estimate how this situation will evolve in the short-term.

The goal of this study is to identify the best possible methods and their parameterization, in order to apply short-term forecasts of traffic flow q [vhc/h] and speed v [km/h] to the online traffic model of VM-CH. Many methods were left aside from the beginning, because of their complexity and/or because there is no guarantee of better results. The forecast covers the next two hours, which represents a travel distances up to roughly 200 km on motorways.

The elements of the modeling process are the assignment, the calibration, the handling of daily variation curves (DVC), and the forecast.

The assignment assigns the origin-destination matrix of trips on the network links. The resulting traffic of the first assignment does not normally fit the measured reality very precisely. This is why the calibration step is necessary, as it modifies the OD matrix (as little as possible) to minimize the difference between assigned and measured traffic. Calibrated values for all links form the basis for the estimation of short-term forecasts. This step is important, because it allows forecasting for all links in the network, and not only for those equipped with traffic counters.

As traffic patterns repeat themselves periodically, the short-term forecast can take advantage of the history by using daily variation curves (DVC). A DVC is simply the pattern of traffic flow (expressed in vehicles per hour) for one day. Normally, such curves are recorded and stored with 15 minutes intervals or 96 data values per day. Large samples of such DVC are available at most traffic management centers. If a sample is too large for efficient handling within short-term forecast procedures, these DVC must first be aggregated, in order to reduce their number to an acceptable level. Cluster analysis techniques are particularly suitable, enabling reduction with a minimal loss of information.

Cluster analysis (CA) is a generic term for numerical methods and procedures. Its goal is to reduce large data and thereby losing the least possible information at each step of the process. The procedures of CA work iteratively. It takes the two most similar elements, aggregating them and repeats the procedure all over again. The user must choose when to stop, or the procedure will continue until all elements are merged to a single one.

The forecast consists of two distinct steps. First, one must find the best fitting DVC for the past period. Second, one has to apply this DVC to the forecast period. The past horizon is two hours, symmetrically to the future one.

A general remark: The quality of short-term traffic forecasts is limited by the inherent uncertainty of the traffic data itself. The precision of the forecast cannot be better than the precision of the traffic data. Measured in terms of mean relative error (MRE), this limit is around 10% for a time horizon of two hours. As long as random fluctuations of traffic lie within this range of precision, there is no hope of obtaining better results.

The results and conclusions of this study concern four areas:

Exponential smoothing is useful to “clean” measured data from random fluctuations and to estimate the error term of these measures. The smoothing parameter adapts automatically to the quality of the data by analyzing the residuals of the autocorrelation. If the residuals are evenly distributed, the coefficient can be lowered to increase dependency on past data. If residuals show some bias, the coefficient must be raised to decrease the dependency on past data.

The study shows that the quality of the forecast is increased by taking the mean of the k-best fitting curves instead of one. The mean of the k-best curves reduces the inherent variability of a single curve. An analog idea is to adjust the DVC not only to time t0, but also with some more intervals in the past. The fluctuations of the measured data are such, that it is impossible to give exact optimal values for k and n. On the other hand, the optimum is very flat, so that values of k = 8 to 12 best fitting clusters and n = 2 to 4 time periods seem appropriate.

The question, if DVC should be identified according to days of the week and special events was also addressed. As the forecast period is only of two hours, this is not necessary. The program will always find a suitable (or k suitable) curve for the past two hours. This greatly simplifies the procedures and eliminates the need for operators to feed the system with information about sport events, holiday departures, etc. The clustering procedure will aggregate all DVC’s independently of their identification.

The goal of cluster analysis is to produce typical DVCs out of many individual ones by successive aggregation. The method of reciprocal pairs developed by de Rham produces an equilibrated set of clusters, with a representative number of different types of clusters. The assumption, that it is good to give a great number of clusters (thousands of curves) to the system was invalidated by the tests. The system then seems to compensate by choosing dozens of k-best fitting curves, which raises run time to critical values. It turned out to be better for the forecasts to reduce the number of curves to about 50 to 100, and then to smooth them with a simple mean-value algorithm. The additional advantage is that the run time will not pose any problem, even for large applications with tens of thousands of links to be forecasted in every 3’ or 5’ interval.

Each day, the newly measured DVC are added to the existing cluster data set. This data set is then reduced by cluster analysis until its dimension reaches the predefined threshold. This self-learning procedure guarantees stability. Either the new clusters are different from the old ones and will be the seed of new typical curves, or the differences are small and the new curves will find similar existing clusters to merge with.

The traffic management center VM-CH will start its operations in 2008. **The research team strongly recommends to initiate a pilot project to implement these methods and monitor the results during about six months.** This time will be divided in three test cycles, each of them with data collection, analysis of results and corrective action.

Résumé

La nouvelle centrale nationale de gestion du trafic VM-CH (Verkehrsmanagement Schweiz) va fonctionner à partir de 2008 et couvrir l'ensemble du réseau des routes nationales suisses. Une gestion efficace n'est possible que si l'on connaît la situation de trafic actuelle et que si l'on est capable d'estimer comment cette situation va évoluer dans le court terme.

Le but de cette recherche est d'identifier les meilleures méthodes et leurs paramètres pour l'application de la prévision à court terme (deux heures) au modèle de trafic VM-CH. Plusieurs méthodes ont été laissées de côté d'emblée, car trop complexes et/ou sans garantie de meilleurs résultats. Les deux heures correspondent à des distances d'environ 200 km sur autoroute.

Les éléments du modèle sont l'affectation, le calage, la gestion des courbes de variation journalières et la prévision.

L'affectation affecte le trafic des matrices origine-destination (OD) aux arcs du réseau. Il y aura donc des différences entre les valeurs calculées par le modèle et les mesures de trafic réel (comptages). C'est ici qu'intervient le calage, qui consiste à modifier (le moins possible) la matrice OD afin de minimiser la différence entre le calcul et la réalité pour chaque arc du système. Ce seront ces valeurs estimées de trafic (volume et vitesse) qui seront la base des prévisions à court terme. Cette étape est importante car elle permet de faire des prévisions pour tous les arcs, et pas seulement pour ceux équipés de compteurs.

Comme les variations de trafic se répètent avec régularité, les prévisions à court terme auront avantage à se baser sur des données historiques telles que les courbes de variation journalières (CVJ). Une CVJ donne simplement des valeurs de trafic (exprimées en nombre de véhicules par heure) pour une journée. Normalement, ces données sont enregistrées et stockées par intervalle de 15 minutes, soit 96 valeurs par jour. De tels ensembles de données sont disponibles dans les centres de gestion de trafic. Si l'ensemble de CVJ est trop grand pour le traitement en ligne (dizaines de milliers de courbes), il faudra préalablement les agréger pour réduire leur nombre. Les méthodes de cluster analyse (CA) sont particulièrement bien adaptées pour réduire l'ensemble étape par étape, tout en minimisant la perte d'information.

La cluster analyse (CA) est un terme générique qui englobe des méthodes et procédures numériques dont le but est de réduire successivement de larges ensembles de données, tout en minimisant la perte d'information à chaque étape du processus. Les procédures de CA consistent à trouver les deux éléments les plus proches, à les agréger en un seul, puis de continuer ainsi. C'est à l'utilisateur de choisir le nombre optimal de CVJ qui lui convient, sinon, le processus continuera jusqu'à ce que tous les éléments aient été agrégés en un seul.

Les prévisions se font en deux étapes distinctes. La première consiste à trouver la CVJ qui correspond le mieux à la période passée. La deuxième à appliquer cette CVJ à la période de prévision. La période passée est de deux heures, par symétrie à la période future.

Une remarque générale : La qualité des prévisions à court terme est limitée par la précision inhérente aux données de trafic elles-mêmes. La précision des prévisions ne peut pas être meilleure que la précision des données de trafic. Mesurée en termes d'erreur moyenne relative (MRE), cette limite se situe aux alentours de 10% sur deux heures. Aussi longtemps que les fluctuations de trafic seront du même ordre, il n'y pas d'espoir de faire mieux.

Les résultats et conclusions de cette recherche concernent quatre domaines :

Le lissage exponentiel sert à « nettoyer » les données mesurées de leur composante aléatoire et à estimer l'erreur de ces mesures. Le paramètre de lissage s'adapte automatiquement à la qualité des données en analysant les résidus de l'autocorrélation. Si les résidus sont bien distribués, le coefficient pourra être abaissé, afin de mieux tenir compte du passé. Si les résidus ont un biais, il faudra augmenter le coefficient pour diminuer l'influence du passé.

L'étude montre que la qualité des prévisions est améliorée en choisissant non pas une, mais la moyenne des k CVJ les plus proches des données historiques. Cette moyenne permet de réduire la variabilité inhérente à une seule courbe. Une idée analogue est appliquée à l'ajustement de la CVJ à t0. L'ajustement sur plusieurs périodes n du passé, au lieu d'uniquement t0, améliore aussi les prévisions. Les fluctuations des valeurs de mesures sont telles qu'il n'est pas possible de donner des valeurs exactes pour k et n. Mais comme l'optimum est assez « plat », des valeurs de k = 8 à 12 courbes et de n = 2 à 3 périodes semblent appropriées.

La question de savoir s'il était nécessaire d'identifier les CVJ par type de jour et par type d'évènement a également été adressée. Mais comme les prévisions ne concernent que deux heures, ce n'est pas nécessaire. Le programme trouvera toujours une (ou k) CVJ correspondant à l'historique des deux dernières heures. Ceci simplifie grandement les procédures et élimine également la nécessité pour l'opérateur de devoir introduire des indications sur les manifestations sportives, les départs en vacances, etc. L'utilisateur peut choisir un seuil pour le nombre de CVJ (= clusters) restantes, p.ex. 1000 à 2000. Le nombre exact n'a que peu d'importance, tant qu'il y en a suffisamment pour s'adapter à toutes les situations.

Le but de la cluster analyse est de former des CVJ typiques à partir de multiples courbes individuelles. La méthode des voisins réciproques développée par de Rham produit un ensemble équilibré de clusters où chaque type de courbe est bien représenté. L'hypothèse qu'il était bon de fournir un maximum de courbes (des milliers) au système a été contredite par les essais pratiques. En effet, le système semblait compenser la variabilité par un grand nombre k (plusieurs douzaines) de courbes les plus proches avec pour conséquence négative d'augmenter dangereusement le temps calcul. Les tests ont montré qu'il valait mieux réduire de nombre de clusters à env. 50 – 100, et de lisser ces courbes avec un simple algorithme de moyenne. L'avantage supplémentaire est que le temps de calcul ne pose plus aucun problème, même pour la prévision de dizaines de milliers de tronçons toutes le 3' ou 5' minutes.

Chaque jour, les CVJ nouvellement mesurées sont ajoutées à l'ensemble des courbes existantes. L'ensemble est alors réduit par la même procédure de cluster analyse que pour l'ensemble initial jusqu'à que le seuil prédéfini soit atteint. La procédure d'apprentissage garantit la stabilité. Si une ou plusieurs courbes sont spécifiques à la situation locale, elles auront tendance à créer un nouveau cluster à part. Par contre, des CVJ similaires aux courbes existantes vont simplement s'y ajouter.

Le centre de gestion du traffic VM-CH sera opérationnel à partir de début 2008. **L'équipe de recherche recommande vivement de démarrer un projet pilote pour implanter ces méthodes et suivre les résultats pendant environs six mois**. Ce temps sera divisé en trois périodes, chacune avec une phase d'observation et de collecte de données, suivie par l'analyse des résultats et les mesures correctives.

Zusammenfassung

Ab 2008 wird die neue nationale Verkehrsmanagementzentrale VM-CH den Betrieb aufnehmen und das gesamte schweizerische Nationalstrassennetz abdecken. Die unabdingbare Voraussetzung für ein effizientes Management sind Kenntnisse über die aktuelle Verkehrslage sowie die Schätzung wie sich diese Lage in der nahen Zukunft entwickeln wird.

Die vorliegende Arbeit hat zum Ziel, die bestmöglichen Methoden und deren Parametrierung für die Anwendung von Kurzfristprognosen von Verkehrsmenge q [Fz/h] und Geschwindigkeit v [km/h] bis zwei Stunden für das online Verkehrsmodell VM-CH zu empfehlen. Manche Methoden wurden von vornherein weggelassen, weil sie zu komplex sind oder keine besseren Resultate garantieren. Die zwei Stunden entsprechen einer Fahrdistanz von etwa 200 km auf Autobahnen.

Die Elemente des Modells sind die Umlegung, die Kalibrierung, das Management der Tagesganglinien (TGL) und die Kurzfristprognose.

Die Umlegung legt den Verkehr der Quell-Ziel (QZ) Matrizen auf die Strecken des Netzes um. Diese Berechnung wird kaum genau mit den Messungen (Zählungen) übereinstimmen. Deshalb ist die Kalibrierung notwendig, mit der die Matrix (so wenig wie möglich) verändert wird, um die Differenzen zwischen Berechnung und Realität zu minimieren. Für die Kurzfristprognose werden diese kalibrierten Werte pro Strecke genutzt. Dies ist wichtig, denn damit können Prognosen für alle Strecken erstellt werden, nicht nur für diejenigen, die mit Zähler ausgerüstet sind.

Da der Verkehr eines Tages sich mit grosser Regelmässigkeit wiederholt, ist es naheliegend, die Tagesganglinien (TGL) für die Prognosen einzusetzen. Eine TGL ist einfach eine Folge von Verkehrsmengen (in Fahrzeug pro Stunde ausgedrückt) für einen Ort (Zähler oder Messquerschnitt) während eines Tages. Normalerweise werden diese Werte pro 15 Minuten (also 96 Werte/Tag) aufgenommen und gespeichert. Solche Daten sind in Verkehrszentralen verfügbar. Falls die Menge zu gross ist (mehrere Zehntausend Kurven), muss diese vor der Verarbeitung reduziert werden. Dazu sind Methoden der Cluster Analyse geeignet, mit denen diese Reduktion Schritt für Schritt und mit minimalem Verlust an Information durchgeführt wird.

Die Cluster Analyse ist ein genereller Begriff für numerische Methoden und Prozeduren, deren Ziel ist, grosse Datenmengen unter gleichzeitiger Minimierung des Informationsverlustes sukzessive zu reduzieren. Das iterative Vorgehen besteht darin, die zwei Elemente zu finden, die einander am meisten ähnlich sind, diese zu einem neuen Element zu vereinen und so weiter zu fahren, bis alle Elemente zu einem vereint wurden.

Die Prognose passiert in zwei Schritten. Zuerst wir die TGL gesucht, die den Verkehrsfluss einer bestimmten zurückliegenden Periode am besten abbildet. Dann wird diese TGL für die Prognose benützt. In Symmetrie zum Prognosezeitraum wurde die betrachtete zurückliegende Periode ebenfalls auf zwei Stunden festgelegt.

Eine generelle Bemerkung: Die Genauigkeit der Kurzfristprognosen kann nicht besser sein als die Genauigkeit der gemessenen Verkehrsdaten. Für Prognosen bis zwei Stunden kann man einen mittleren relativen Fehler (MRE, mean realtive error) von ca. 10% erreichen. Solange die Verkehrsdaten zufällige Abweichungen dieser Grössenordnung zeigen, gibt es keine Hoffnung auf bessere Resultate.

Die Resultate und Folgerungen dieser Forschung betreffen vier Bereiche:

Die exponentielle Glättung dient dazu, die gemessenen Daten q von zufälligen Komponenten zu trennen und den Fehler der Messung zu schätzen. Der Glättungsparameter passt sich automatisch an die Qualität der Daten an, indem die Residuen der Autokorrelation analysiert werden. Falls die Residuen gut verteilt sind, kann der Parameter gesenkt werden um besser der Vergangenheit Rechnung zu tragen. Falls die Residuen einen Bias enthalten, muss der Koeffizient erhöht werden, um den Einfluss der Vergangenheit zu mindern.

Die Studie zeigt, dass die Qualität der Prognose erhöht werden kann, indem nicht eine, sondern das Mittel der k am besten passenden TGL genommen wird. Eine analoge Idee ist, die Anpassung der TGL zur Zeit t0 auf mehrere Intervalle n der nahen Vergangenheit auszudehnen. Die lokalen Abweichungen der Messwerte sind so, dass es nicht möglich ist, exakte Werte für k und n zu benennen. Aber das Optimum ist „flach“ und Werte von k = 8-12 für die besten TGL und n = 2-4 für die Intervalle in der Vergangenheit scheinen vernüftig.

Die Frage ob die TGL nach Tagestyp und nach speziellen Ereignissen aufgesplittet werden solle, wurde auch untersucht. Da aber die Prognosen nur den Zeitraum bis zwei Stunden abdecken, ist es nicht notwendig. Das Programm wird immer eine (oder k) TGL finden, die gut zu den letzten zwei Stunden passen. Dies vereinfacht die Prozeduren massgeblich und eliminiert auch die Notwendigkeit, dass der Operateur das System mit Angaben über Sportanlässe, Ferienbeginn usw. füttert. Die Cluster Analyse wird die gegebenen TGL unabhängig von ihrem Typ aggregieren.

Die Cluster Analyse soll typische TGL aus der Menge aller TGL aggregieren. Die von de Rham entwickelte Methode der reziproken Paare produziert Clusters, wo die verschiedenen TGL ausgewogen vertreten sind. Die Annahme, dass es gut sei, dem System möglichst viele typische TGL zur Verfügung zu stellen, wurde durch die Tests widerlegt. Das System scheint zu kompensieren, indem Dutzende von besten Kurven gewählt werden. Das hatte einen negativen Einfluss auf die Rechenzeit. Neue Tests zeigten, dass es vorteilhafter ist, die Anzahl typische TGL auf ca 50-100 zu beschränken und diese mit einem einfachen Mittelwert-Algorithmus zu glätten. Der zusätzliche Vorteil dieses Vorgehens ist, dass die Rechenzeit keine Rolle mehr spielt, auch für grosse Anwendungen mit zehntausenden von Strecken, die alle 3’ oder 5’ Intervalle prognostiziert werden.

Jeden Tag werden die neuen TGL der neusten Messdaten zu der Menge der existierenden TGL addiert. Anschliessend wird die Menge reduziert, bis der vorgegebene Schwellenwert erreicht wird. Dieses Lern-Verfahren garantiert Stabilität. Falls eine oder mehr neue TGL sich von den existierenden unterscheiden, werden sie neue Cluster bilden. Hingegen werden TGL, die den existierenden nahe liegen, einfach zu diesen ergänzt.

Die Verkehrsmanagementzentrale Schweiz (VM-CH) geht anfangs 2008 operativ in Betrieb. **Das Forschungsteam empfiehlt sehr, in einem Pilotprojekt diese Methoden zu implementieren und die Resultate während ca. 6 Monate zu verfolgen**. Diese Zeit wird in drei Perioden unterteilt, je mit einer Phase des Datensammelns, der Auswertung der Resultate und der korrektiven Massnahmen.

1. Introduction

1.1 What is the problem?

Definition: With “short-term forecasts”, we define the ability to predict traffic flow and speed for the time interval from now on, to two hours from now. This corresponds to maximum travel distances of about 200 km on motorways (at a speed of 100km/h).

Short-term forecasts play a central role in traffic management and the actuality of the subject is stressed with one question:

“How will the Federal Office of Roads be sure to use the best possible short-term forecasts in the new traffic management centre (VM-CH) by the year 2008?”

The quality of traffic information, travel times, routing advices, dynamic setting of variable message signs, etc. depends directly on the quality of short-term forecasts.

However, there exists no coherent methodology. Many methods are (re-) invented or developed with no background knowledge, simply because of the immediate necessity of specific applications.

Most of these methods are very probably sub-optimal, even if their output seems to fulfill the need of their user.

The number of applications relying on short-term forecasts will grow in the next years and with it the need for robust methods with measurable quality.

1.2 State of the art, need for research

As far as we know, there is no study or guideline about forecast methods for traffic management systems, in spite of the fact that they are widely used.

If an administration wants to implement short-term forecasts, it will have difficulties in writing specifications, defining quality standards to be met, and comparing bids from suppliers.

1.3 Applied method

Survey about short-term forecasts in Switzerland

A small survey was organized among SVI Members to get an overview of the use of short-term forecasts in Switzerland. The number of expected answers was between 5 and 10. The reality was worse: only one instance seemed to be concerned with short-term forecasts! The details of the survey and the only answer are attached at the end of this report.

How to transfer knowledge from the past to the future

The great advantage of short-term forecasts is that values forecasted at t0 for time t0 + t can be compared with measured values at the latest at t0 + t !

Various questions are to be answered:

How to build knowledge from the known past with daily variation curves.

How information is extracted from past data of daily variation curves

How this information is stored.

What the system “knows”.

If it is able to learn from past, and how.

If it is able to react to changes, and how.

How the knowledge about the past is applied to the future.

Quality is integrated to the study, answering questions, such as:

Which quality can be guaranteed at t0 for time t0 + t.

How quality is measured at time t0 + t

How error propagation will influence results.

Can quality control be used to:

Measure how well a method performs in forecasting.

Optimize parameter settings.

Compare different methods among each other.

Choice of methods

Three methods are candidates for a deeper analysis: cluster analysis, regression analysis and Fourier series. If the survey or the literature search reveals that other methods are useful for short-term forecasting, they will be included too. However, this research will not be extended to methods a) existing only in theory, b) needing highly complex calculations, like neuronal networks, Kalmann filters, Box-Jenkins, hybrid models, etc. or c) with no reasonable chance to be of any utility for traffic applications.

Cluster analysis

Cluster analysis has been used for years by C. de Rham. It was also the subject of his PhD thesis. The method works well, but is very probably implemented sub optimally. This research will help find the best options among metrics, distance functions, parameter settings, number of clusters, etc.

Regression analysis

Regression analysis is also widely used for short-term forecasting. Here too, it would be of great value to have at least some recommendations about parameters, acceptable R^2 levels, number of time intervals etc. to assure a near-optimal implementation.

Fourier series

These are new to traffic engineers and no applications are known to the research team. C. de Rham is an electrical engineer and well off with Fourier techniques.

The idea behind Fourier series is that every continuous curve can be expressed as a sum of sine waves of increasing frequency and varying amplitude. The whole information content of the curve is given by the “series” of frequencies and amplitudes of the single sine waves which, once added, reproduce the original curve.

This has many advantages for the analysis of time series analysis. Low frequencies belong to real trends. The higher the frequencies the higher the probability that it is just noise. It is possible that forecasts could be estimated with frequencies instead of the curve itself.

Dr. S. von der Ruhren from Momatec GmbH, Aachen, has developed a demo to calculate Fourier coefficients from daily variation curves. It is astonishing to see how well the method approximates existing curves with even a small number of frequencies. An interesting question is: Are forecasts in the frequency domain more accurate than in the time domain? A positive answer would be a break-through in favor of the method.

The comparison of the 3+ methods will be central to this study. The following questions are to be answered:

What is the optimal setting of parameters, etc. for each method?

Does one method consistently produce better forecasts?

If not, which one can be recommended to users?

Error Propagation and Short-term Forecasting

A preliminary definition:

“Errors” are not “bugs” but unavoidable uncertainties, which are inherent to measured data.

The research project SVI2004/015 „Error Propagation in Transport Models“ produced interesting results for the assignment and calibration of online traffic situations.

One of the conclusions is that error propagation can directly be used as a tool for quality management. The idea is simple: if the quality (= errors) of the input is known, and if the software is able to propagate this quality (= errors) throughout the whole calculation process, the output quality (= errors) will be known too. This is what has been done in the present case.

There is no reason not to apply this idea to short-term forecasts as well. We would like to show that by integrating error propagation throughout short-term forecasting calculation, one could directly get the output data needed for quality control. This hypothesis must be verified, but we actually do not see why it should not work.

The great advantage of this integrated error propagation would be that any additional and/or external software to monitor the quality of results would become obsolete. Each step, where a result can be obtained with less complexity and less software is a step in the right direction.

1.4. Data availability

At least this problem will not be one! Heaps of urban / inter urban, lightly / heavy traffic daily variation curves are available for the city of Zurich, from the Federal Office of Roads and many other sources.

The research team had the chance to have access to a large historical data set of daily variation curves from HLSV (Hessisches Landesamt für Strassen- und Verkehrstechnik, D). This data set includes volume and speed for truck and car in 15’ interval for each weekday from over 1500 traffic counters.

The LST BW (Landesstelle für Strassentechnik from Land Baden-Württemberg, D) made another very useful and complementary source of data accessible. A model built up to monitor the traffic on the rectangle A5 / A6 / A8 / A81 between Stuttgart and Mannheim used these data. The size of the model made it possible to run full days within minutes and still obtain results based on a realistic situation.

1.5. General work program, milestones

Step 1: Survey, collection of historical data, experiment set up

Survey among members of SVI about the use of methods
Literature search on internet
Decision about which methods are to be analyzed in depth
Extension of existing software with tools to run and monitor test cases
Collection of daily variation curves for different situations
Experiment setup, for individual and comparative analysis of all methods
First meeting with the steering committee

Step 2: Calculations, optimizations, comparison

Run of tests, production of results for analysis
Parameter optimization for each method
Conclusion for each individual method
Second meeting with the steering group

Step 3: Conclusions, report

Comparison of all methods
Advice about the use of methods
Third and last meeting with the steering committee

1.6. Expected results and utility

The program should:

Determine which method is most appropriate to a) build knowledge from past data, b) apply this knowledge to the future, c) apply quality monitoring by error propagation, d) learn from past experience through feedback mechanisms

Give an overview about available methods for public and private administrations, software houses and other instances concerned with short-term forecasts.

Help administrations in formulating specifications and evaluating bids

Help software houses to choose among methods

Help users in evaluating short-term forecast results

2 Data and Processes for Short-Term Forecasts

2.1 Overview

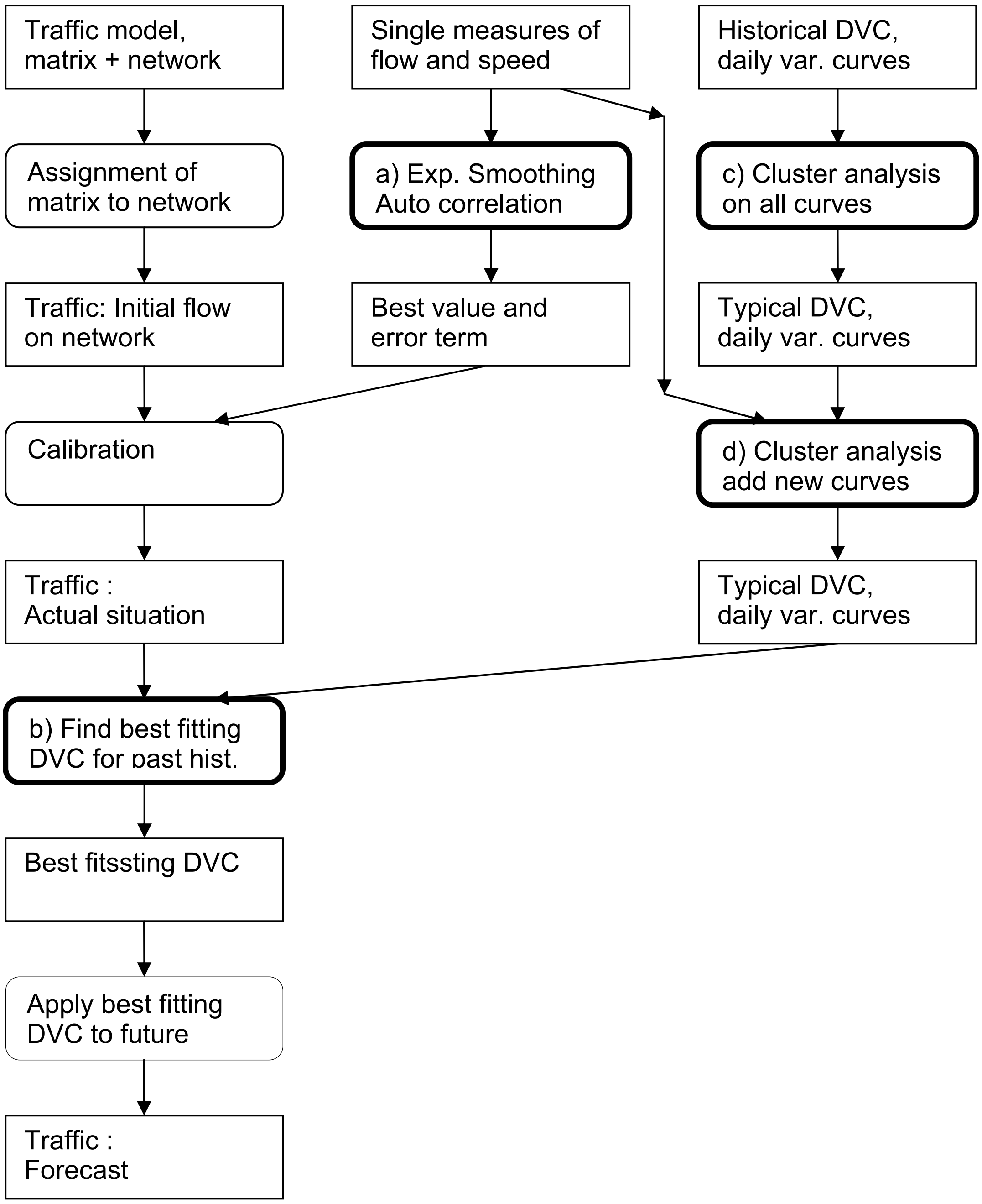

Fig. 1, Data and processes for short-term forecasts

This study is mainly concerned with methods used by the processes in the rectangles above in bold:

a) Exponential smoothing to “clean” measured data from local random fluctuations and to estimate the error term of these measures.

b) Measures of fit to compare DVCs with the past two hours and to choose the most probable DVC to be applied to the next two hours.

c) Cluster analysis to produce typical DVCs out of many individual ones by successive aggregation around reciprocal pairs.

d) Cluster analysis to add new curves to the set of existing typical curves.

2.2 Definitions

The time axis is divided in discrete intervals dt. This interval dt is normally equal to 15’ for historical daily variation curves (DVC). The interval for online data collection is smaller, about 3’ or 5’. Therefore, the short-term forecast will also be updated every 3’ or 5’, even if the interval of the DVC is larger. This method is called “rolling forecast”.

We adopt the following convention to number the past, present and future intervals, whih extend from the two past hours to the two future ones.

Horizon	**←**			**past**				**→**	**←**			**future**				**→**
Interval #, (15’ each)	**t-7**	**t-6**	**t-5**	**t-4**	**t-3**	**t-2**	**t-1**	**t0**	**t+1**	**t+2**	**t+3**	**t+4**	**t+5**	**t+6**	**t+7**	**t+8**
time from (minutes)	**-120**	**-105**	**-90**	**-75**	**-60**	**-45**	**-30**	**-15**	**0**	**15**	**30**	**45**	**60**	**75**	**90**	**105**
time to (minutes)	**-105**	**-90**	**-75**	**-60**	**-45**	**-30**	**-15**	**0**	**15**	**30**	**45**	**60**	**75**	**90**	**105**	**120**
forecast Period									**t15**	**t30**		**t60**				**t120**

Fig. 2, Past and future horizons

The past horizon goes from t0 - 120’ to t0 inclusive, with steps of 15’ each. As measurements at t0 are known, t0 will also belong to the past. The future horizon goes from t+15’ to t+120’, also with steps of 15’.

The following indices will be used throughout the text for the flow q[vhc/h] and speed v[km/h]:

qc	calibrated value
qd	value from daily variation curves
qf	forecasted value
qi	initial value from first assignment before calibration
qm	measured value from sensors, FCD, etc.
qs	smoothed measured value
e	error term = qs - qm
es	smoothed error term e

When assessing the goodness of fit of a forecast from a traffic model, one is confronted with different types of discrepancies in the chain from measured data qm to forecasted data qf. It is important to separate the different effects and to isolate only those the forecasting method may influence.

The step from measured data qm to smoothed measurement qs is necessary to separate the measurement from the error term e = qs - qm.

The step from smoothed measurement qs (only for links with sensors) to calibrated values qc (for all links) is the calibration process described in [de Rham 06]. This actual situation is qc and serves as base for traffic information, online representations (Google Earth), etc.

The actual calibrated traffic situation on all links (values qc) will be the reference for the tests with forecasting methods and parameter settings.

2.3 Exponential smoothing

Exponential smoothing will be analyzed for two purposes, first as a method to calculate error terms of measurement data, second as a way to create short-term forecasts.

See also www.wikipedia.org for more details.

Exponential smoothing assigns weights that are exponentially decreasing, as the observations get older. The general formula is:

$$qs(0) = qm(0)$$

$$qs(t) = \alpha\ qm(t) + (1-\alpha)\ qs(t-1) = qs(t-1) + \alpha\ (qm(t) - qs(t-1))$$

where α is the smoothing constant.

The formula can be expanded as

$$qs(t) = \alpha\ qm(t) + (1-\alpha)\ qs(t-1)$$

$$qs(t) = \alpha\ qm(t) + \alpha\ (1-\alpha)\ qm(t-1) + (1-\alpha)^2\ qs(t-2)$$

$$qs(t) = \alpha\ [qm(t) + (1-\alpha)\ qm(t-1) + (1-\alpha)^2\ qm(t-2) + (1-\alpha)^3\ qm(t-3) +] + (1-\alpha)^t\ qm(0)$$

In words (cited from [Wikipedia]): “As time passes, the smoothed statistic qs(t) becomes the weighted average of more and more values of the past observations qm and the weights assigned to previous observations are proportional to the geometric progression {1, (1-α), $(1-\alpha)^2$, $(1-\alpha)^3$, … } which illustrates the exponential behavior”.

With values of α close to 1, recent values have a great weight and dampening is quick, whereas values of α close to 0 will induce slow dampening. This is illustrated in the table below:

➔ towards past observations

α	(1- α)	$(1-\alpha)^2$	$(1-\alpha)^3$	$(1-a)^4$
.9	.1	.01	.001	.0001
.5	.5	.25	.125	.0625
.1	.9	.81	.729	.6561

Fig. 3, Exponential smoothing weights with different values of α

In the example below, the values are random numbers in range {0..1}. Exp(α) are three exponential smoothing results for α = 0.1, α = 0.5 and α = 0.9.

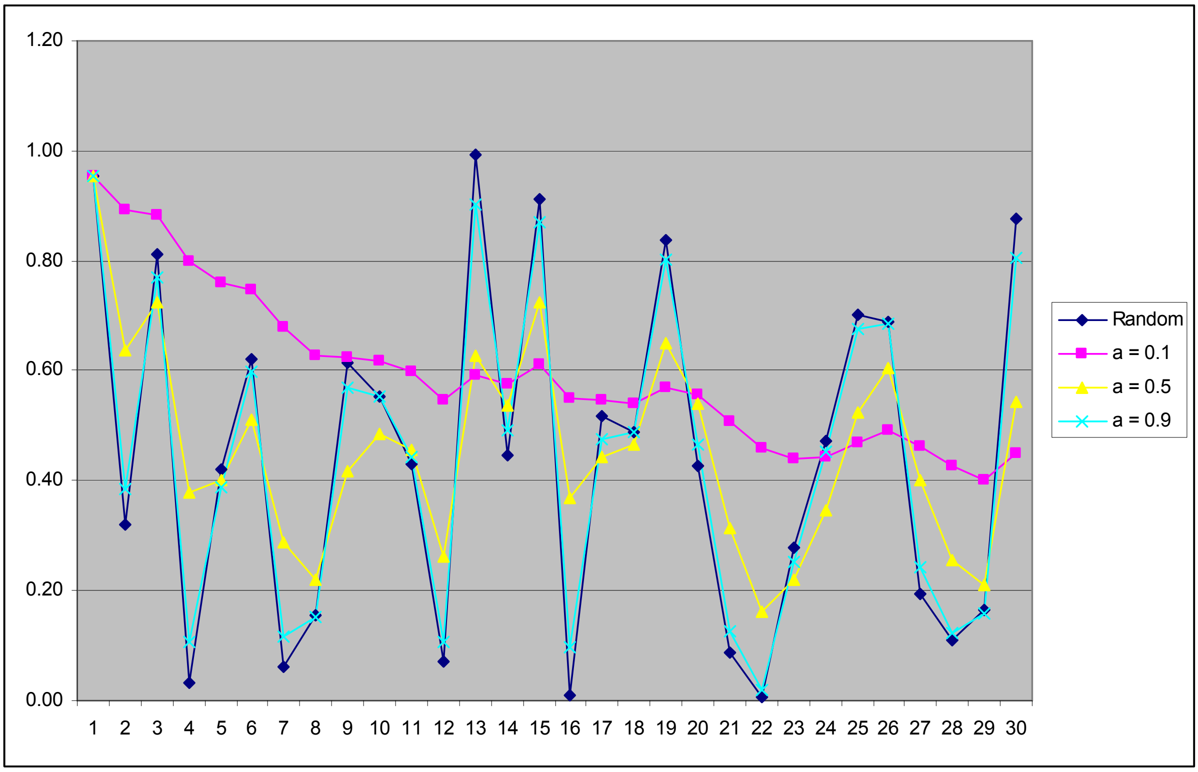

Fig. 4, Exponential smoothing of “Random” with α = 0.1, 0.5 and 0.9

A smoothing parameter α = 0.9 has nearly no effect on the data. With α = 0.5 , the smoothing is quite efficient in eliminating peaks but still keeping fluctuations. With an α = 0.1, the original pattern of the curve disappears nearly completely.

The question of course is: what is the optimal value for α?

The optimal value for α used for error estimations will be the one that best separates error from the underlying trend.

When doing forecasts, the optimal value of α will produce the most accurate forecasts.

The simplest, even if somewhat fastidious way, is to try a range of values for α and to keep the one which produces the “best” results, measured by some criteria.

Another way is to let α « float » and have a feedback mechanism that continuously corrects α towards some optimal value. This is the adopted solution described in the next chapter.

2.4 Exponential smoothing to estimate error terms

Traffic volumes are never smoothly distributed in space (road sections) and time (measurement interval). Raw data from sensors can show large local fluctuations. The cause is the high sensitivity of traffic flow to small disturbances. A simple truck or camping car being somewhat slower than average may induce the formation of local groups of cars that produce data values not significant for the whole picture.

One way to avoid having the forecast method depend too much on such fluctuations is to smooth the raw measurement data before use as a basis for forecasts. The following method was developed during the research project “Error Propagation in Transport Models” [de Rham 06] to smooth measurement values and calculate the error term necessary to subsequent error propagation.

At each interval dt (3’ or 5’), the raw measured value qm is used within exponential smoothing to get the best estimate qs and its error term dqs:

qs(t)	= (α * qm(t))	+ ((1- α) * qs(t-1))		smoothed
e(t)	= abs(qm(t)	–	qs(t))	error term or residual
es(t)	= (α * e(t))	+ ((1- α) * es(t-1))		smoothed error term

From then on, error propagation rules are applied with:

$$q(t) = qs(t) \pm es(t)$$

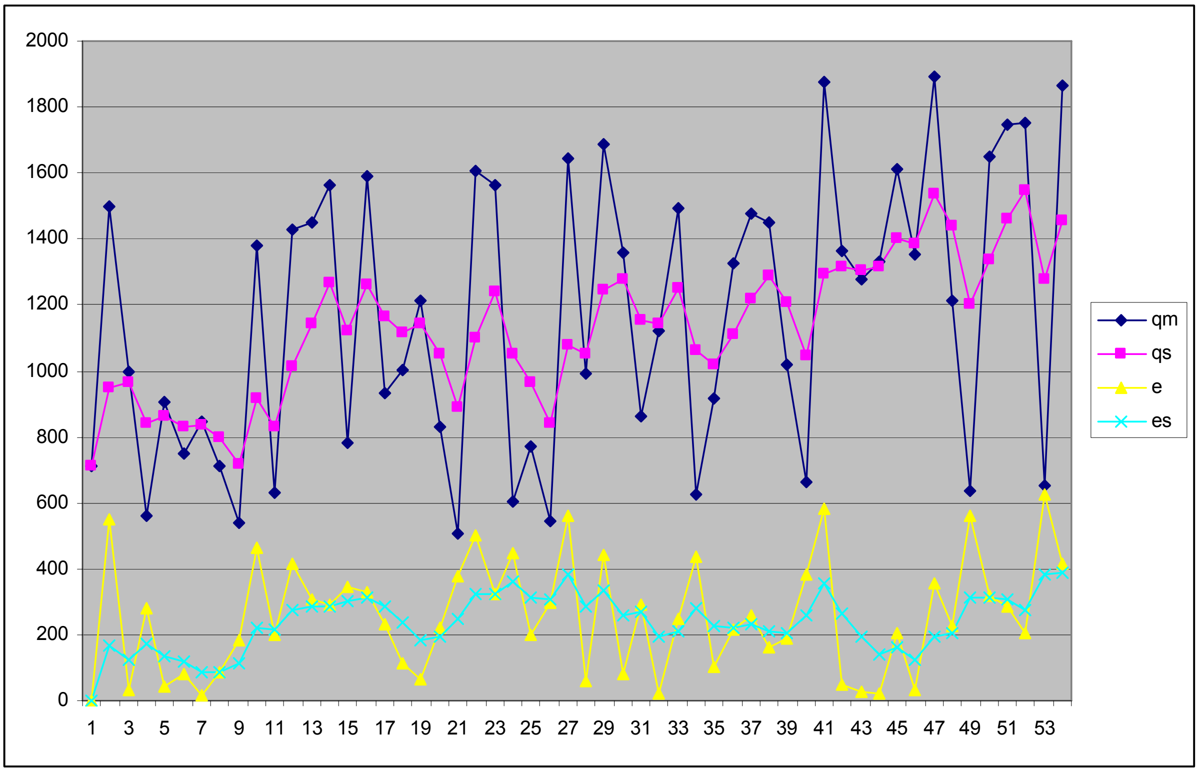

Fig. 5, Exponential smoothing to estimate the error term

The figure above illustrates the dependencies: qm are the raw measurements, qs their smoothed value, e the error term |qm – qs| and es the smoothed value of this error term.

The question now is how to estimate the best fitting value for α. This will be done by analyzing the behavior of the error term e(t). By definition, the error terms e(t) should be independent from each other for different time intervals t. This means that they must not be autocorrelated. This is how autocorrelation analysis of the residuals will help finding the optimal value of α.

If the residuals are highly autocorrelated, it means that past values depend too much on each other. Therefore, α should be increased to shorten the influence of the past. On the other hand, a weak autocorrelation of residuals means that α can be decreased to depend more on past values.

2.5 Measures of fit between DVCs

The fit between two curves or group of curves can be expressed by so called “distance measures”. There are many possible definitions of such distance measures. The most used ones are variants of the L-metric defined as:

$$\mathbf{L_n = (\sum (xi - yi)^n)^{1/n}}$$

L_1 is the so-called Manhattan metric,

$$\mathbf{L_1 = \sum |xi - yi|}.$$

If you move on a rectangular grid (like in Manhattan), the distance between two points will be the sum of the distances of the single arcs connecting the two points.

L_2 is the formula of Pythagoras,

$$\mathbf{L_2 = (\sum (xi - yi)^2)^{1/2}}$$

L_∞ is equal to the maximum difference between elements,

$$\mathbf{L_\infty = \max |xi - yi|}.$$

These distance measures may be combined with weights. One weighting measure for L_2 consists to multiply it with a factor F depending on the numbers na and nb of elements of clusters a and b:

$$\mathbf{L_2 = L_2 * ((na * nb) / (na + nb))^{1/2}}$$

This weighting ensures monotonic behavior of the distances. This means that no distance will decrease during the clustering process.[1]

If one divides L_1 by the number N of differences (= dimensions), one gets the MAE or mean absolute error:

$$\mathbf{MAE = L_1 / N = (\sum |xi - yi|) / N}$$

Similar to the MAE is the MRE or mean relative error. The MRE is suited when comparing results y to some reference x:

$$\mathbf{MRE = (\sum |xi - yi| / xi) / N}$$

The advantage of the MRE is that it will treat curves with high and low traffic volumes similarly. A difference of 4000 and 5000 vehicles at peak hours will produce the same relative value as 40 and 50 vehicles at 3 AM. The MRE is definitely more suitable for the analysis of daily variation curves than the MAE.

Another widely used measure is the root mean square error or RMSE:

$$\mathbf{RMSE = (\sum (|xi - yi| / xi)^2 / N)^{1/2}}$$

Its disadvantage is that it is a squared measure, which may overreact to large single differences.

[1] More in [de Rham 80].

The linear correlation coefficient is also interesting, when comparing two sets of data. The definition of R^2 or RSQ is:

$$\mathbf{R^2 = (\sum (xi * yi) * \sum (xi * yi)) / (\sum (xi * xi) * \sum (yi * yi))}$$

The comparisons of results will include the three measures of fit: **MRE, RMSE and R^2**.

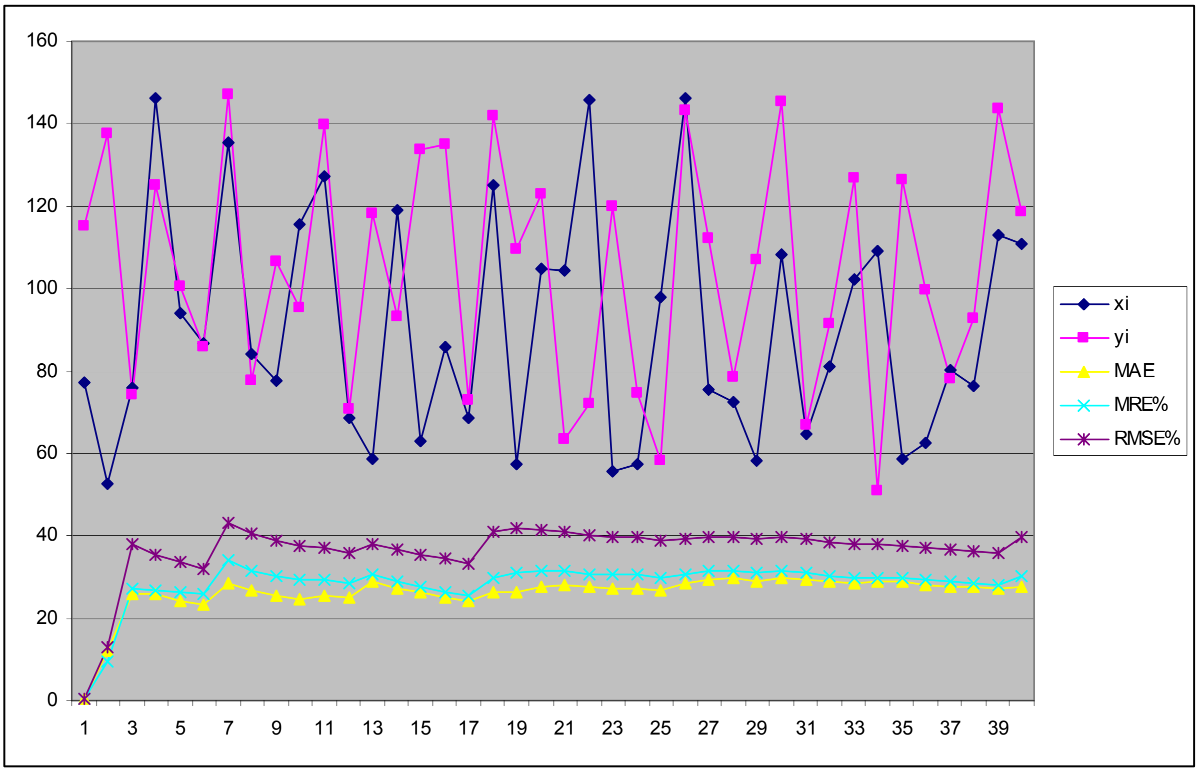

Fig. 6, Behavior of MAE, MRE and RMSE with random data

This figure shows how the value of the criteria MAE, MRE and RMSE change for 1 to 40 pairs of (xi, yi) chosen randomly between 50 and 150. The three criteria behave very similarly; they are practically interchangeable. The correlation coefficient is not included here; its value has no meaning because of the randomness of xi and yi.

2.6 Cluster Analysis

Clustering algorithms came up in the early seventies, when a new problem appeared to statisticians and other specialists handling data. The increased use of computers caused the cost of data collection to fall sharply. The consequences were immediate: surveys produced huge amounts of data, which could not be directly analyzed with conventional methods. Suddenly, statisticians were confronted with the problem of too much data! They had to invent methods to reduce those heaps of information, mostly with multiple dimensions, before being able to analyze them. This was a fruitful period with many “inventions”, such as correspondence analysis, multidimensional scaling, etc.

The basic procedure of cluster analysis is very straightforward and is illustrated by the following example. Intuitively, the six thin curves c1 to c6 of the diagram below can be grouped into two typical bold ones CA and CB. Expressed in the language of cluster analysis:, this means the elements c1, c2 and c3 are best grouped to form a cluster CA and elements c4, c5 and c6 are best grouped to form a cluster CB.

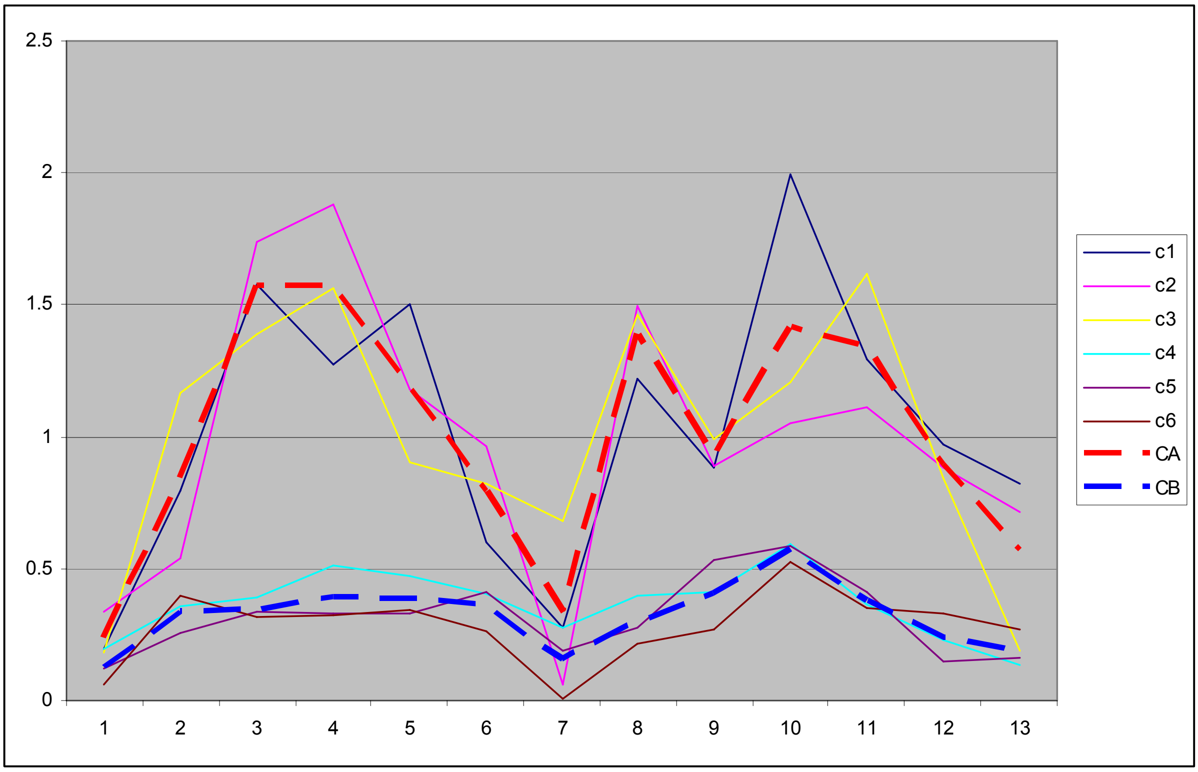

Fig. 7, cluster analysis

The field of cluster analysis is concerned with the measurement of likeliness or proximity of elements, and with procedures to aggregate these elements to clusters. One of the objectives is to choose the next elements which are to be aggregated in such a way, that loss of information content is minimal at each aggregation step.

It is natural, logical and self-evident to base forecasts on historical data, like daily variation curves. Traffic management centers tend to be well equipped with such historical curves available from databases. Even if such a database does not exist locally, there is always the possibility of taking a DVC set from other sources.

One question, which frequently arises is how to aggregate single DVCs to typical curves (for Mondays, for holidays, special events, etc. The method used is cluster analysis. The same cluster analysis methods are very helpful to find which DVC best fits some time series of historical data.

The basic cluster algorithm for hierarchical classification works as follows:

```
Initialize:   Calculate all distances between elements
Loop          Find the smallest distance between two elements, say i and j
              Aggregate i and j to a new element k
              If some stopping criteria is reached: EXIT loop
              (Re)-calculate distances from k to all other elements
End loop
```

This algorithm needs storage in the order of $O(n^2)$ (for the matrix of distances) and also runs with times of order $O(n^2)$. This is unaffordable for large amount of data, like tens of thousands or more elements. This is where one of us came up with the idea of using reciprocal pairs [de Rham 82] to reduce the order towards $O(n * \log(n))$.

If two elements constitute a reciprocal pair, it means that both of them are nearer to each other than to any other element in the set. A classic ballroom with dancing couples (not a disco!) is a good example. Each individual out of a couple (or pair) is closer to its partner than to any other person in the room. Now we can reformulate the algorithm:

```
Initialize:   Find all p reciprocal pairs (*)
Loop          Aggregate all p reciprocal pairs to p new elements
              If some stopping criteria is reached: EXIT loop
              (Re)-calculate all q reciprocal pairs (*)
              Set p = q
End loop
```

This new algorithm needs not to store the whole distance matrix, but only an array of the reciprocal pairs. Thus, the storage will be of order $O(n)$ instead of $O(n^2)$. As the number of reciprocal pairs is decreasing exponentially (ideally by a factor of 2 at each step), the run time is of order $O(n * \log(n))$.

2.7 Cluster analysis for the reduction of large sets of DVC

The research team had the chance to work with a high quality data base from Land Hessen, Germany, with data from more than 1500 section counters. The data set includes volume and speed data for car and truck for each quarter of an hour and every day of the week.

Ganglinientyp	Klasse	Ereignisname	Klassengrösse	Wochentag	MQ-Name	Datenart	00:00	00:15
Standard	0	Standard	20	0	B43/20RN	qKfz	313	298
Standard	0	Standard	20	0	B43/20RN	qLkw	27	20
Standard	0	Standard	20	0	B43/20RN	vPkw	79	79
Standard	0	Standard	20	0	B43/20RN	vLkw	70	71
Standard	0	Standard	20	0	B43/20RS	qKfz	206	213
Standard	0	Standard	20	0	B43/20RS	qLkw	7	7
Standard	0	Standard	20	0	B43/20RS	vPkw	66	66
Standard	0	Standard	20	0	B43/20RS	vLkw	60	56
Standard	0	Standard	20	0	B43/21RN	qKfz	101	76
Standard	0	Standard	20	0	B43/21RN	qLkw	6	6
Standard	0	Standard	20	0	B43/21RN	vPkw	64	64
Standard	0	Standard	20	0	B43/21RN	vLkw	48	46
Standard	0	Standard	20	0	B43/21RS	qKfz	56	60
Standard	0	Standard	20	0	B43/21RS	qLkw	8	7
Standard	0	Standard	20	0	B43/21RS	vPkw	66	67
Standard	0	Standard	20	0	B43/21RS	vLkw	50	52
Standard	0	Standard	20	0	B43/23RN	qKfz	37	33
Standard	0	Standard	20	0	B43/23RN	qLkw	6	4
Standard	0	Standard	20	0	B43/23RN	vPkw	45	45

Fig. 8, Upper left corner of the data set of daily variation curves.

There are more than 20’000 records with volume data (ca. 1500 section counts, 2 types of vehicle, 7 types of days). This is far too much to handle for short-term forecast, where the short-term history of each link is to be compared with each available curve. This is where cluster analysis is useful, reducing the number of single curves to a set of typical curves.

Theoretically, one may choose between N individual curves and one single mean cluster. As there are no criteria to choose the optimal number of clusters, this number may be chosen with practical criteria. In our case, we had the algorithm stop as soon as the number of remaining clusters is below 1000. It is enough to guarantee a large choice of typical curves to match the history of measurements, and is still easily manageable from the computing time aspect.

As the daily variation curves were available per weekday, we had to choose if the typical curves were to be aggregated by weekday too, or could be aggregated neutrally, without considering their “origin”. The argument in favor of weekdays is that it could lead to better forecasts, as it differentiates more in detail. The argument against is that the number of curves is large enough to find a good fitting one anyway, and that the differentiation by days of the week could even produce erroneous forecasts, when a day does not “behave” according to its normality.

2.8 Cluster analysis for the update with new DVCs

Fresh DVCs are available for every link at the end of each day. Links with section counts have measured values, links without have the calibrated flow values. These DVCs are automatically integrated in the set of existing clusters. The procedure is analogous to the initial aggregation. First, all new DVCs are added to the existing set. Second, cluster analysis is applied to reduce the number until the threshold is reached.

If one or more curves are specific to the local situation, they will tend to build a new cluster of their own. Curves which are very similar to existing ones will simply join existing clusters. This is how the system will adapt to the local behavior of daily traffic.

The speed at which old clusters are replaced by new ones can be roughly estimated with a mean percentage of new clusters at each renewal step.

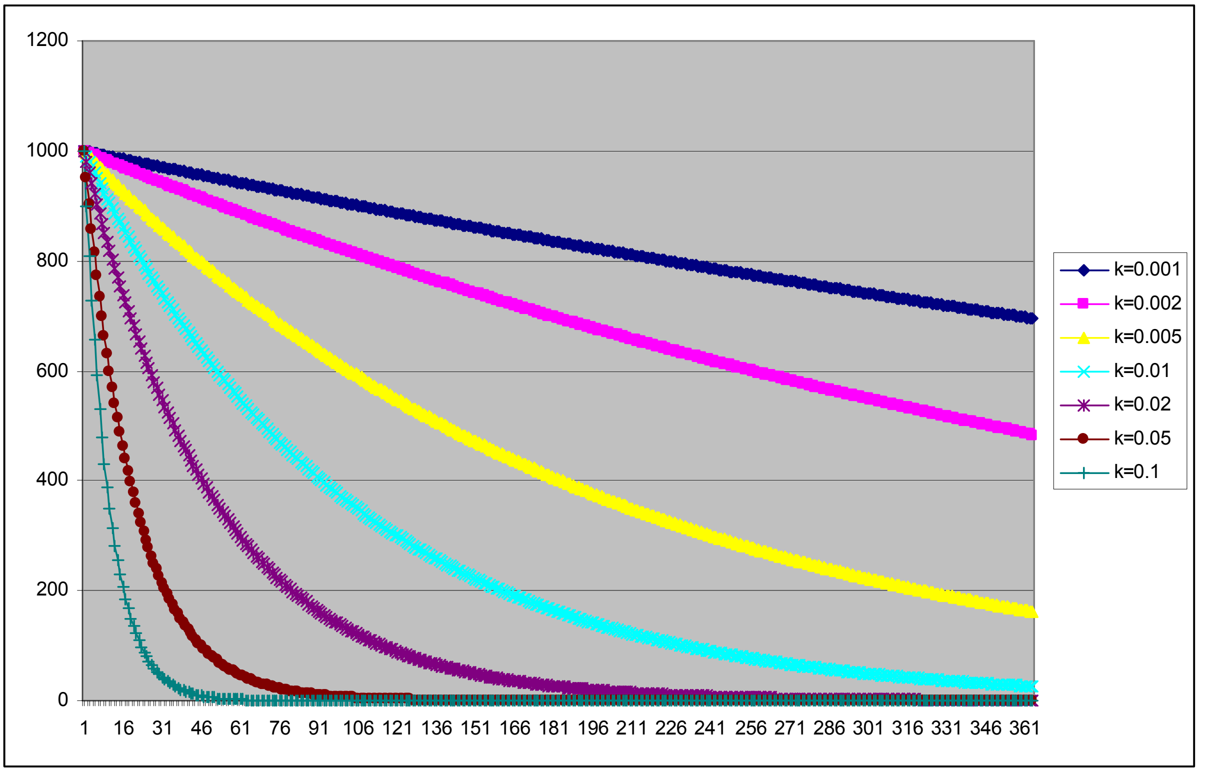

Fig. 9, Replacement speed of clusters as a function of parameter k

If only one thousandth of clusters is replaced each day, roughly two thirds will be replaced after 1 year. But the renewal time shortens fast with growing parameter k. With k = 1%, nearly all clusters will have been replaced after one year, and with k = 10%, it takes only 18 days to renew half of them.

Anyway, the method is effective in both situations: either the new DVCs are different from the base set and the renewal rate will be fast, or the new DVCs are similar to the existing ones and there is no need to renew the original data set.

2.9 Smoothing DVC data around midnight

The typical daily variation curves will be used for forecasts round the clock. The history extends two hours back and the future two hours ahead. This poses a problem when the actual time “now” is before two o’clock in the morning, or after ten o’clock in the evening. To ensure the continuity of the daily variation curves two hours before and after midnight, the curves are smoothly corrected to meet the same value for the first and last interval of 15’.

Assume the head (first) and tail (last) values of a daily variation curve are qh and qt. A correction factor is applied to the eight first and the eight last values of the curve to ensure that qh equals qt, and that these values will be approached smoothly.

We define the mean value qm of qh and qt; that is where head and tail will meet at midnight.

qm = (qh + qt) / 2

The two factors fh and ft are:

fh = (qm / qh) -1

ft = (qm / qt) -1 by the way: ft = - fh

A loop will apply fh and ft gradually to both ends of the curve:

```
for k in 0..7 loop
        F = (8 – k) / 8
        q(head + k) = q(head + k) * (1 + (fh * F))
        q(tail    – k) = q(tail    – k) * (1 + (ft * F))
end loop
```

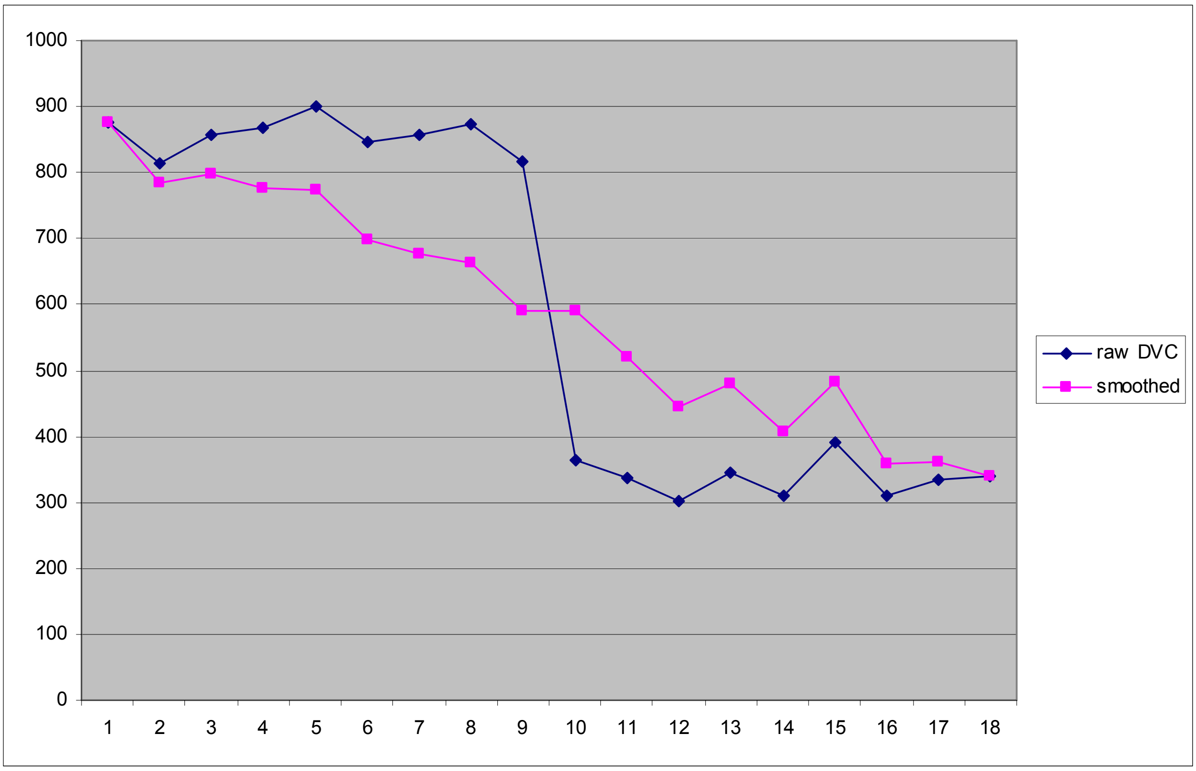

Fig. 10, Smoothing data of DVC around midnight

This method ensures that both ends of the curve will meet at midnight and that forecasts between ten o’clock in the evening and two o’clock in the morning will pass the midnight line smoothly.

2.10 Smoothing DVC data through the day

Even the clustered DVC have local variations, which are not representative for the behavior of the traffic. The following diagram shows a hundred of such DVCs chosen at random among the original set.

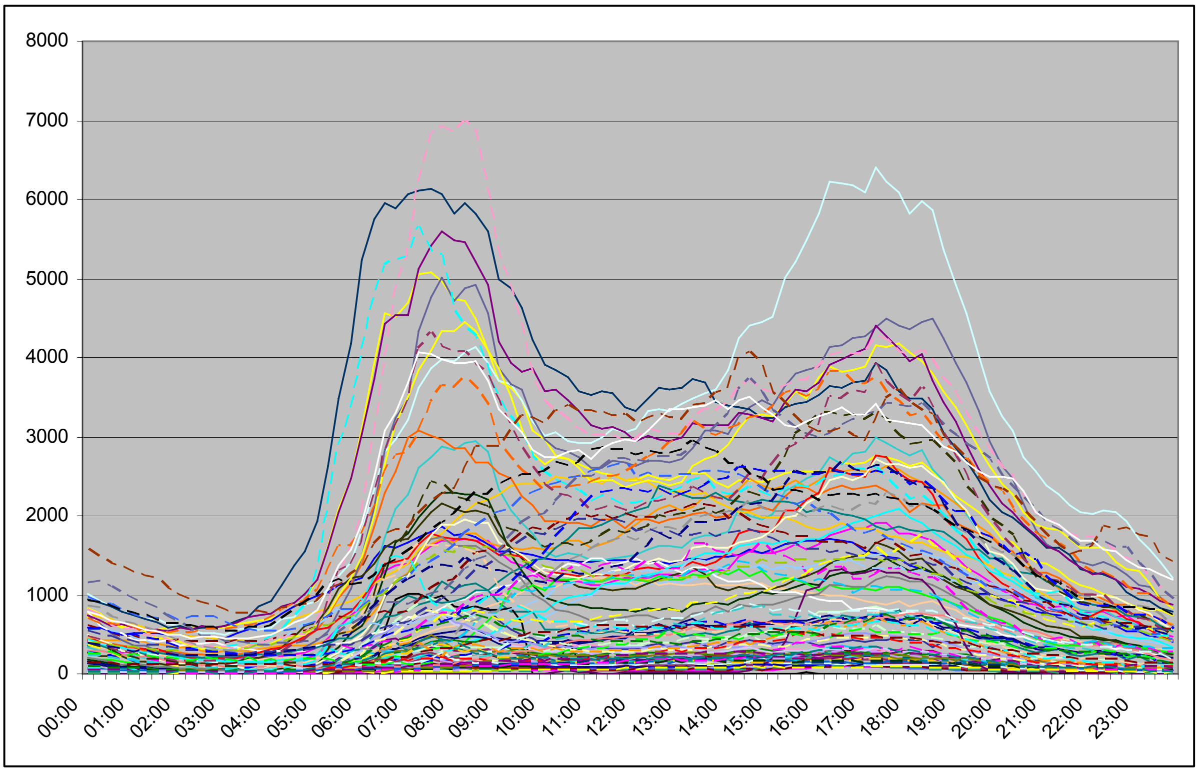

Fig. 11, DVCs with local variations

As these local fluctuations do not contribute to the quality of the forecast, they can be eliminated by a simple though very efficient smoothing method typically used in image processing. To be consistent for the first and last element, this smoothing should be executed after the smoothing around midnight. The transformation goes like this:

q1(t) = (q0(t-1) + q0(t) + q0(t+1)) / 3

And for the first and last element:

q1(t_first) = (q0(t_last) + q0(t_first) + q0(t_first+1)) / 3

q1(t_last) = (q0(t_last-1) + q0(t_last) + q0(t_first)) / 3

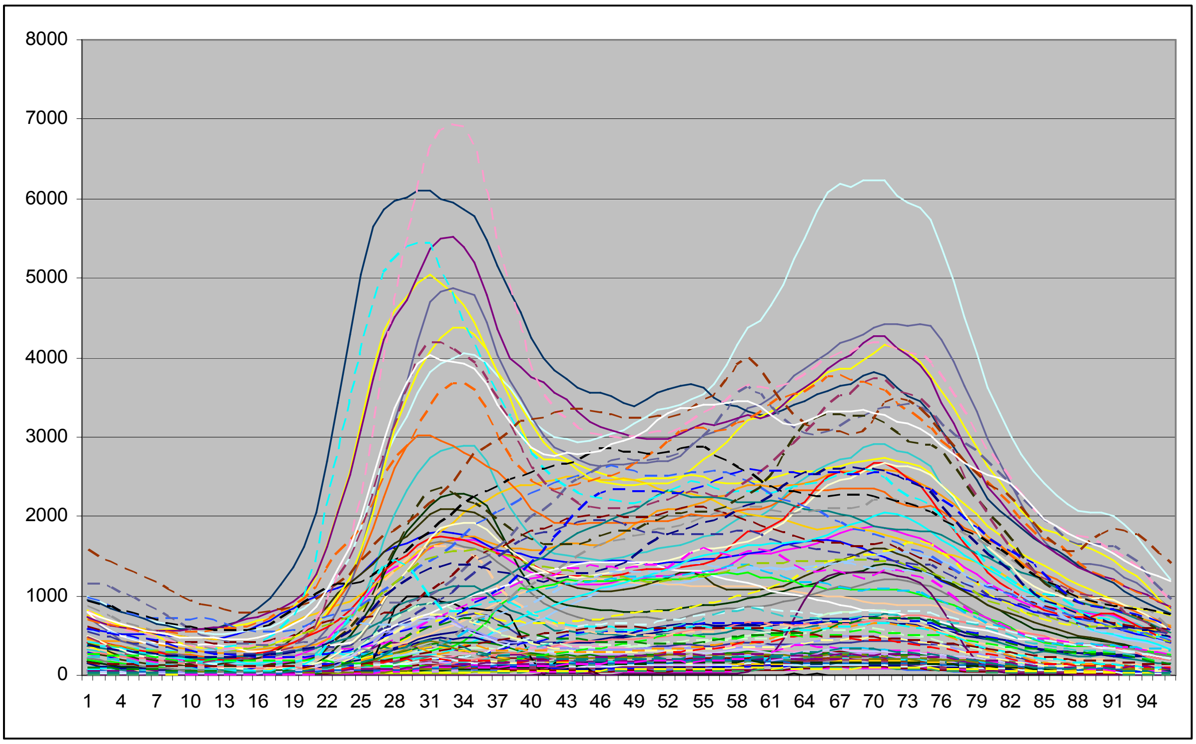

Fig. 12, DVCs after one smoothing step

The advantage of this method is that it can be applied successively to the same data, until all local variations have completely vanished

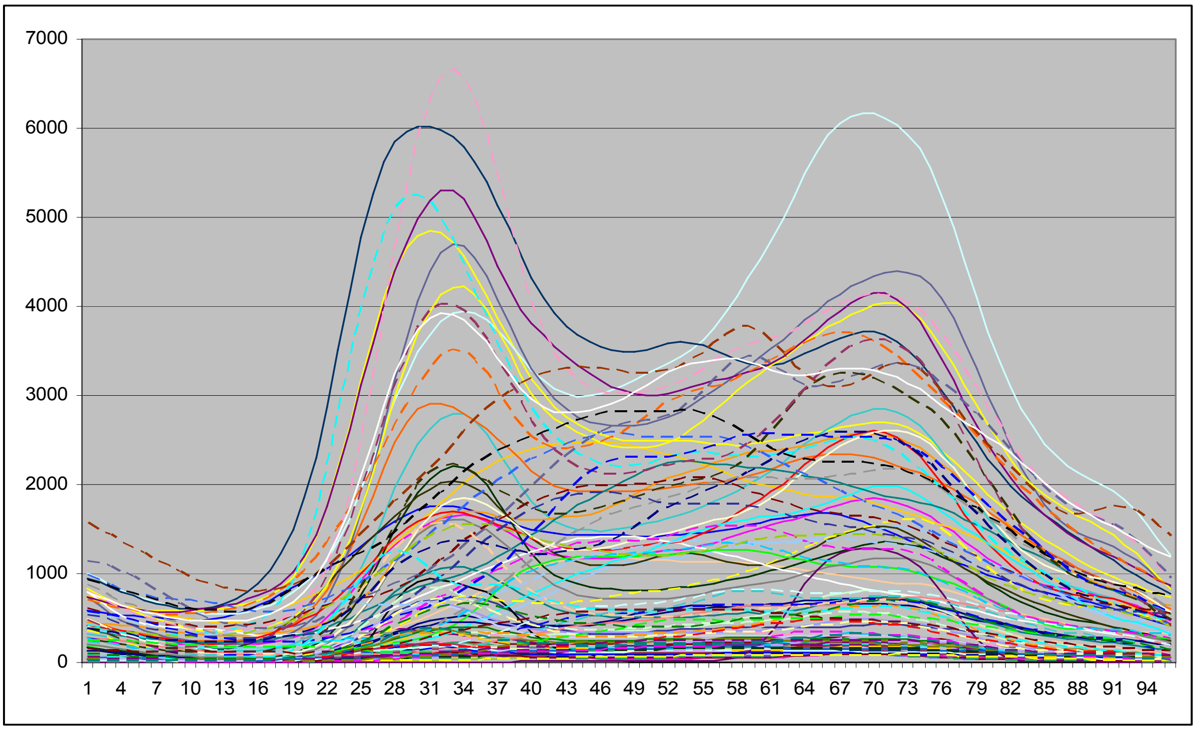

Fig. 13, DVCs after six smoothing steps

The positive effect of this smoothing method is clearly visible. All local variations have disappeared and only the basic information content of the curve remains. It is to be noticed that the successive smoothing steps do not modify the maxima and minima of the curves.

3 Other Forecasting Methods

The PhD thesis of Stefan [von der Ruhren] gives an excellent overview of existing methods. We limit our overview to methods applicable to VM-CH, and exclude methods based on sophisticated mathematical procedures.

3.1 Forecast with a moving average

Definition: the forecasted value qf for time t+1 is the x+1 point moving average of measured values from t-x to t0:

qf(t+1) = SUM(qm(t-x) … q(t0)) / (x+1)

When applied to traffic forecasting, this method has two disadvantages:

The first is that all values are equally weighted, which does not reflect the reality of traffic behavior.

The second is that it does not adapt to changes fast enough. When traffic volumes increase or decrease sharply before or after of a peak period, the time lag of the moving average will be too important. Predictions will then be systematically too conservative and/or too late.

This is why the moving average method is not appropriate for short-term forecast and will not be studied further in this project.

3.2 Forecasts with exponential smoothing

Definition of the forecasted value qf for time t+1:

qf(t+1) = α * qs(t0) + (1- α) * qf(t0)

Even if this looks somewhat better than moving average, this method isn't suitable for efficient forecasting of traffic either, for reasons similar to those of the moving average method: it does not adapt to changes fast enough. When traffic volumes increase or decrease sharply before or after of a peak period, the time lag is too important. As with the moving average method, predictions are systematically too conservative and/or too late.

Just as with the moving average method, exponential smoothing is not appropriate for short-term forecast and will not be studied further in this project.

3.3 Fourier series

There seem to be no applications of Fourier transforms to short-term forecast in traffic modeling. One reason is very probably that most traffic engineers have a civil engineering background without lectures about Fourier transforms and applications.

[Kirschfink] describes an application where Fourier transforms are used to reduce the amount of data for the storage of daily variation curves. This could have been of some interest when storage capacity was scarce and expensive, but such times are over. The same application also mentions the comparison of two variation curves in the frequency domain, but without mentioning its relative merit to the same operation in the time domain.

On one side, the idea is interesting, but on the other, there are no tangible results. The conclusion is that it has been tried, but that something seems to make it impossible to realize. This is what we wanted to find out.

The first step of estimating the Fourier coefficients from a DVC is very straightforward and poses no problem. The difficulty starts when a part of a DVC (like the past two hours) has to be compared with a known DVC. There is no way to estimate the Fourier coefficients of the two hour interval which can be matched to those of a 24h DVC.

From the time aspect, information about when the partial curve is to be compared with the full DVC is always at hand. However, if one transforms a partial curve, the information about the beginning and the end of the interval is definitively lost, and the coefficients are not comparable with those of the full DVC any more.

This is the reason why there is no traffic forecast based on Fourier transforms.

3.4 Differentiation by day of the week

The question if DVC should be identified according to days of the week and special events was also addressed. [Peters 05] wrote an interesting paper about building and monitoring a database of DVCs for traffic applications. However, the goal of this study is only to forecast within the next two hours. It is not necessary to differentiate by days of the week to reach this goal.. The program will always find a suitable DVC for the past two hours.

This greatly simplifies the procedures and eliminates the need for operators to feed the system with information about sport events, holiday departures, etc. The clustering procedure will aggregate all DVCs independently of their identification. The user can select the threshold number of remaining clusters. The exact number is not very important, as long as there are enough of them to adapt to all normal situations.

3.5 Forecast with DVC and re-calibration

The individual forecast of every flow value of every link has one drawback: the sum of ingoing and outgoing traffic at nodes will not be equal any more (like Kirchhoff's law for electrical current). The only way to correct it is to (re-) assign and (re-) calibrate the matrix for each future forecast period.

This additional assignment and calibration will guarantee a coherent future state, where link value has been forecasted with the best possible DVC, and Kirchhoff's law is respected in every node as well.

Nevertheless, this "Gedankenexperiment" did not work, because the quality of the forecast was reduced by the additional calibration.

When analyzed more precisely, this result is no surprise. The calibration consists of adapting the trip values from the origin destination matrix to obtain the best possible fit with observed link values. This means that the assigned trips will slightly change the "best possible" forecasted values. This change introduces additional dispersion of the forecasted data, which decreases its forecasting precision.

One has to choose between two options: Either a best possible forecast, knowing that there may be small inconsistencies in the nodes, or a completely coherent model, but with forecast of lesser quality.

For practical applications, good forecasts are more important than perfect consistency. The re-calibration option is abandoned.

3.6 More complex methods

Many other methods are applied in forecasting, some highly complex, other suiting to a particular field. These methods appear here and there in the literature, mostly within single research projects. Among those are Box-Jenkins ARIMA autoregressive correlations, Kalman filters, neural networks and hybrids that mix different methods.

All these methods have in common that they are not widespread for traffic application. This is due to their complexity, and to the fact that they are not usable without a specific software.

4 Tests and Results

4.1 Measured data: best estimate and error term

The auto correlation coefficient r^2 is calculated for the last 40 values of es(t). This corresponds to the last two hours (with intervals of 3'), or the last 3h20' (with intervals of 5').

The value α is checked individually for each link and at every time interval. To avoid instabilities, the steps were set to 0.1 (up or down).

Start	with α = 0.2
If R_coeff > 0.7	then α = α + 0.1
If R_coeff in 0.3..0.7	then keep α as it is
If R_coeff < 0.3	then α = α - 0.1

The new value α is limited to the range 0.1..0.9.

Tests with many practical applications show that α converges rapidly towards the lowest possible value of 0.1. A small statistics keeps track of which percentage of α's have this minimal value of 0.1. This percentage normally stays around 95%.

This method has another advantage: it adapts automatically to change. If for one or another reason, the behavior of traffic changes, α may get larger during some intervals and return to 0.1 as soon as normality is back. The speed of change is limited to the step of 0.1 per 3'.

4.2 Forecast with daily variation curves, basic procedure

The base value for the forecast will be the calibrated flow value qc, because qc is the most probable actual traffic flow value on every link of the network. By the way: the measured values qm would not suit because they are limited to only a fraction of the links. The values qm are only used to calibrate the model and to update the set of DCVs.

The forecast takes the following steps:

- find the curve which best fits the known history from t-x to t0 with MRE
- take the forecasted values from t+1 to t+y from the curve
- wait 15', 30', 60' and 120' and compare the forecasted values with the actual ones
- evaluate the result, again with the MRE criterion

This procedure is the reference from which all others will be derived.

DateTime	MRE1	MRE2	MRE4	MRE8	RMSE1	RMSE2	REMS4	RMSE8	RSQ1	RSQ2	RSQ4	RSQ8
Average	**4.95**	**6.46**	**8.71**	**12.19**	7.48	9.61	12.86	17.41	0.98	0.97	0.96	0.93
Stdev	1.42	2.06	2.61	2.67	2.44	3.34	4.16	4.15	0.02	0.02	0.03	0.04
Min	2.58	3.49	4.62	6.64	3.69	5.21	6.15	8.43	0.89	0.84	0.84	0.84
Max	9.04	11.96	15.86	18.40	13.97	19.88	22.82	26.36	1.00	1.00	1.00	0.99

Fig.14, Test results with basic procedure (res-070530-2157-360-00-00.csv)

The most important data of this table is highlighted in row "Average" and columns MRE1 to MRE8. It represents the mean relative error between actual calibrated link flow and forecasted link flow for the four time horizons of 15', 30', 60' and 120'.

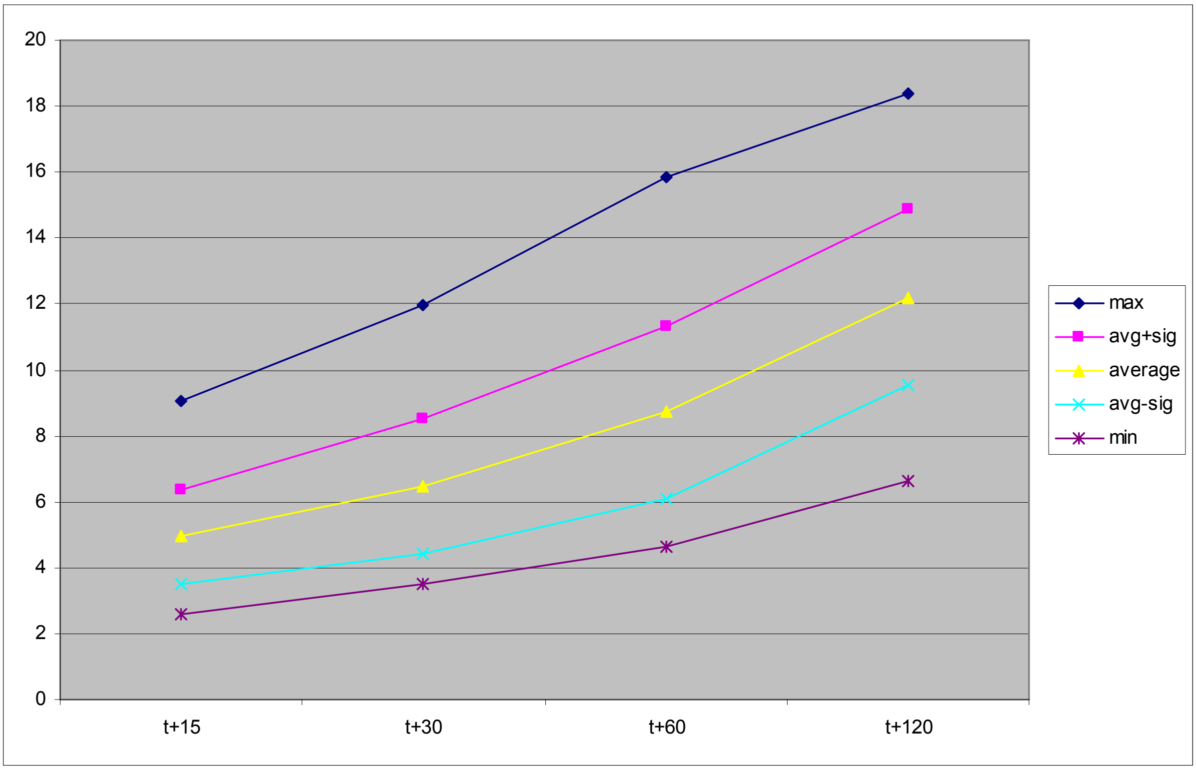

Fig. 15, Forecast errors for t+15 to t+120

The average error varies from 5% for t+15 to 12% for t+120. It is probable that these results are acceptable for traffic management purposes. However, it is also probable that they can be refined, at least to some part. This is the subject of the next chapter.

4.3 Forecast with daily variation curves, refinements

Length of past horizon

At first, the past horizon was chosen to be variable with values of t-7, t-5, t-3 and t-1 which corresponds to -120’, -90’, -60’ and -30’ minutes. But soon it became obvious that this was not a good idea. In particular, -60’ and -30’ produced inconsistent results, because of the short interval. It was then decided that the past horizon would be constant and include the 8 intervals of 15’ from t0-105’ to t0. The search of the best fitting curve will be done by comparing the 8 past calibrated values qc(t0-105) to the values qd(t0) of the DVCs.

Comparison with the correlation coefficient

Another idea was to compare the curves by calculating the correlation coefficient r relating the two sets of values xi and yi:

$$R^2 = \sum(xi * yi) * (xi * yi) / (\sum(xi * xi) * \sum(yi * yi))$$

This approach is not useful, because r neither indicates a relative fit, nor an absolute proximity of the curves. If the values of one set are constant multiples of the other set (like yi = k* xi with k = constant), the value of r will be exactly equal to 1, even if both curves are far apart with a great value of k.

Mean of the k-best fitting variation curves

Searching for a best fitting curve is like trading at the stock exchange: An excellent fit with past records does not guarantee a good fit in the future. Nevertheless, there is a possibility to reduce the inherent variability of one single curve by taking more than one of them. Instead of choosing only the best fitting curve, one chooses the k-best ones and combines them to form a new curve. This new curve (or cluster) will be the one applied for the forecast. Different numbers of clusters were tested. A first conclusion is that the optimum is quite flat, and that it must lie around 5 to 10 curves. Anyway, the variability of all concerned data renders the search for an exact optimum illusory.

DateTime	MRE1	MRE2	MRE4	MRE8	RMSE1	RMSE2	REMS4	RMSE8	RSQ1	RSQ2	RSQ4	RSQ8
Average	**4.81**	**6.20**	**8.12**	**10.88**	7.33	9.28	12.02	15.45	0.98	0.97	0.97	0.95
Stdev	1.40	2.03	2.48	2.73	2.45	3.39	4.20	4.32	0.02	0.02	0.03	0.03
Min	2.27	3.15	4.37	6.69	3.43	4.65	5.92	8.84	0.89	0.84	0.86	0.86
Max	8.99	11.61	14.54	17.76	13.89	19.40	22.04	24.67	1.00	1.00	1.00	0.99

Fig.16, Test results with 10 best curves instead of only one

This refinement shows a small decrease of the forecast error for each of the horizons. Such an increase in forecast quality is barely worth the additional programming effort for itself, but it should be maintained, if it increases quality when combined with other refinements.

Adjust at t0, extension to t0-n

Even when a curve fits globally, it never fits the value at t0 exactly. Tests have shown that the forecast error may further decrease when the curve is adjusted to n time intervals or periods. That means that the future values are multiplied with a factor FQ:

$$\mathbf{FQ = \sum (qc(t0-n+1) .. qc(t0)) \ / \ \sum (qd(t0-n+1) .. qd(t0))}$$

For n = 3 the factor FQ will be:

$$FQ = (qc(t0-2) + qc(t0-1) + qc(t0)) / (qd(t0-2) + qd(t0-1) + qd(t0))$$

DateTime	MRE1	MRE2	MRE4	MRE8	RMSE1	RMSE2	REMS4	RMSE8	RSQ1	RSQ2	RSQ4	RSQ8
Average	**2.70**	**4.73**	**7.70**	**11.46**	4.28	7.29	11.57	16.66	0.99	0.98	0.97	0.94
Stdev	0.90	1.63	2.41	2.26	1.77	2.88	3.93	3.82	0.01	0.02	0.03	0.04
Min	1.28	2.20	3.65	6.95	1.62	3.03	4.96	8.73	0.96	0.92	0.86	0.82
Max	5.48	9.56	14.55	16.91	10.22	15.55	22.02	24.13	1.00	1.00	1.00	0.99

Fig.17, Test results with adjustments from t0 to t0-3+1 periods

Here too, the variability of the input data renders the search for an exact value of n illusory. Many tests show that a value of n between 2 and 4 will produce smaller forecast errors. As with the choice of k = 10 clusters instead of one, the increase in quality is small, but if both improvement are joined, the increase in quality is worthwhile.

Combination of k-best fitting curves and adjustment with n periods

The two combined improvements produced the following results:

DateTime	MRE1	MRE2	MRE4	MRE8	RMSE1	RMSE2	REMS4	RMSE8	RSQ1	RSQ2	RSQ4	RSQ8
Average	**2.45**	**4.26**	**6.83**	**9.92**	3.87	6.62	10.36	14.30	1.00	0.99	0.97	0.95
Stdev	0.87	1.53	2.28	2.51	1.77	2.83	3.91	4.11	0.01	0.01	0.02	0.03
Min	1.02	2.21	3.54	6.26	1.31	2.98	4.79	8.10	0.97	0.93	0.88	0.89
Max	5.08	9.13	13.15	15.22	10.02	14.95	20.82	23.24	1.00	1.00	1.00	0.99

Fig. 18 Forecast with k = 10 best curves and n = 3 period adjustment

The forecast error for 2h falls even below 10%, which shows that the effect of both improvements is cumulative.

To check these preliminary results, two other days are analyzed in exactly the same way. First by taking k = 10 best fitting clusters instead of one and second by adjusting with n = 3 past periods instead of one.

DateTime	MRE1	MRE2	MRE4	MRE8	RMSE1	RMSE2	REMS4	RMSE8	RSQ1	RSQ2	RSQ4	RSQ8
Average	**2.25**	**3.95**	**6.44**	**9.62**	3.14	5.44	8.70	12.63	1.00	0.99	0.98	0.97
Stdev	0.74	1.40	2.55	4.00	1.23	2.09	3.39	4.67	0.00	0.01	0.03	0.04
Min	1.24	1.70	3.27	4.68	1.67	2.32	4.08	6.01	0.97	0.90	0.83	0.83
Max	5.07	8.26	14.41	18.14	7.22	12.43	18.28	23.68	1.00	1.00	1.00	1.00

Fig. 19, Second test with k = 10 best curves and n = 3 period adjustment

DateTime	MRE1	MRE2	MRE4	MRE8	RMSE1	RMSE2	REMS4	RMSE8	RSQ1	RSQ2	RSQ4	RSQ8
Average	**2.33**	**4.14**	**6.82**	**10.86**	3.13	5.45	8.69	13.28	1.00	0.99	0.98	0.97
Stdev	0.89	1.72	3.05	5.11	1.20	2.20	3.71	5.82	0.00	0.01	0.01	0.02
Min	1.09	1.75	2.81	3.93	1.58	2.43	3.93	5.23	0.99	0.97	0.94	0.92
Max	5.01	9.77	17.13	24.09	7.05	10.96	20.01	28.76	1.00	1.00	1.00	0.99

Fig. 20, Third test with k = 10 best curves and n = 3 period adjustment

Both results show an improvement of about 2% for the 120' forecast and a total error of about 10%. This leads to a first conclusion that the quality of the forecasts may increase with the help of two measures:

- **taking the k (k ≈ 10) best clusters instead of only one**
- **adjusting at t0 with n (n ≈ 3) past values instead of only one**

4.5 Forecast with DVC and damped fitting

This method is based on an idea kindly communicated by Dr. Stefan von der Ruhren. The original method is similar to the adjustment at t0 and goes like this:

Mean difference between measurement and DVC for the near past intervals t0-n to t0.

DQ(t0) = ∑ (qc(t0-n) - qd(t0-n)) / (n + 1)

Apply a decreased weighting to the forecasted values ranging from t0+1 to the horizon th:

DQ(t) = DQ * (th-t+1)+1 / (th – t0+1)

Add this weighted difference to the forecasted values from t0+1 to horizon th:

qf(t) := qf(t) + DQ(t)

The additive term DQ corresponds to the multiplicative adjustment term FQ at t0 from above. Instead of applying additive values, the same method can be applied multiplicatively, this to be consistent with all other refinements. The damping is applied to future values of FQ for the next two hours. By the way: it is very similar to the adjustment around midnight.

The multiplicative factor FQ at t0 will be (as above):

FQ(t0) = ∑ (qc(t0-n) .. qc(t0)) / ∑ (qd(t0-n) .. qd(t0))

The successive FQ's is damped until zero within a time horizon of 2 hours or 8 quarters of an hour:

FQ(t) = 1 + ((FQ(t0) – 1) * (8 – t) / 8), with t in {0..7}

The last value of FQ will be equal to 1 + ((FQ – 1) * (1/8)).

4.6 First series of test: combinations of methods

A first systematic test was completed with the following 5 combinations of improvements. Combinations including both n - period adjustment and damped fitting are excluded, because these methods are exclusive.

Test	k - best fitting curves	n – period adjustment	damped fitting	figure
2	--	Yes	--	22
3	--	--	Yes	23
4	Yes	--	--	24
6	Yes	Yes	--	25
7	Yes	--	Yes	26

Fig. 21, Tests with combinations of improvements

The results of the five tests are below. The most important criteria is the average MRE, and among them the MRE8 for the 2h – forecast.

DateTime	MRE1	MRE2	MRE4	MRE8	RMSE1	RMSE2	REMS4	RMSE8	RSQ1	RSQ2	RSQ4	RSQ8
Average	**7.43**	**8.11**	**9.40**	**11.41**	10.07	10.92	12.55	15.24	0.98	0.97	0.96	0.94
Stdev	2.18	2.06	2.28	2.63	3.27	2.99	3.22	3.47	0.01	0.01	0.02	0.03
Min	3.80	4.56	5.70	7.32	4.69	5.57	7.02	9.78	0.94	0.93	0.91	0.77
Max	13.80	13.91	17.13	20.09	22.17	20.28	24.45	25.79	0.99	0.99	0.99	0.99

Fig. 22, Test with 3 – period adjustment

The n – period adjustment brings a small improvement, and the method will be kept for further combinations.

DateTime	MRE1	MRE2	MRE4	MRE8	RMSE1	RMSE2	REMS4	RMSE8	RSQ1	RSQ2	RSQ4	RSQ8
Average	**7.39**	**8.14**	**9.61**	**12.50**	10.01	10.91	12.71	16.32	0.98	0.97	0.96	0.93
Stdev	2.15	2.07	2.23	3.11	3.22	2.95	3.08	3.70	0.01	0.02	0.02	0.04
Min	3.83	4.75	6.05	6.66	4.73	5.73	7.53	9.05	0.94	0.93	0.90	0.73
Max	13.57	13.58	16.31	22.48	21.84	20.17	23.47	28.10	0.99	0.99	0.99	0.98

Fig. 23, Test with damped fitting

It does not seem that damped fitting improves the forecast quality. The decision to take or leave it is taken after test 7 below.

DateTime	MRE1	MRE2	MRE4	MRE8	RMSE1	RMSE2	REMS4	RMSE8	RSQ1	RSQ2	RSQ4	RSQ8
Average	**7.20**	**7.79**	**8.70**	**9.99**	9.84	10.59	11.60	13.00	0.98	0.98	0.97	0.96
Stdev	2.34	2.22	2.24	2.36	3.67	3.18	3.28	3.12	0.02	0.01	0.02	0.02
Min	4.11	4.53	5.44	6.00	5.42	6.52	6.94	7.88	0.92	0.93	0.93	0.89
Max	15.42	13.58	14.97	15.55	25.74	18.65	22.81	21.03	0.99	0.99	0.99	0.99

Fig. 24, Test with 10 – best fitting curves

This idea seems to be promising, also probably in combinations with other improvements.

DateTime	MRE1	MRE2	MRE4	MRE8	RMSE1	RMSE2	REMS4	RMSE8	RSQ1	RSQ2	RSQ4	RSQ8
Average	**7.23**	**7.73**	**8.57**	**9.80**	9.84	10.41	11.29	12.71	0.98	0.98	0.97	0.96
Stdev	2.18	2.09	2.38	2.51	3.27	3.08	3.37	3.27	0.01	0.01	0.01	0.02
Min	3.84	4.61	4.92	6.28	4.83	5.80	6.33	8.33	0.94	0.94	0.94	0.90
Max	13.27	13.30	15.98	15.86	21.72	19.80	23.79	22.15	1.00	0.99	0.99	0.99

Fig. 25, Test with 10 – best fitting curves and 3 – period adjustment

This is the best combination among all these tests. We shall keep it for further analysis.

DateTime	MRE1	MRE2	MRE4	MRE8	RMSE1	RMSE2	REMS4	RMSE8	RSQ1	RSQ2	RSQ4	RSQ8
Average	**7.23**	**7.94**	**9.67**	**14.66**	9.76	10.48	12.35	18.47	0.98	0.98	0.97	0.94
Stdev	2.15	2.08	2.34	2.97	3.21	2.92	2.88	2.88	0.01	0.01	0.01	0.02
Min	4.30	4.92	5.24	10.58	5.57	6.53	7.59	13.79	0.95	0.94	0.93	0.88
Max	13.00	13.69	15.66	22.02	21.48	19.51	21.77	25.90	0.99	0.99	0.99	0.98

Fig. 26, Test with 10 – best fitting curves and damped fitting

This last result was somewhat disappointing. The damped fitting method does not seem to have a positive effect on the quality of short-term forecasting. One explanation could be that the damped fitting corresponds to a sort of local “bending” of the DVC, and that this bending cannot possibly produce a more suitable curve. As damped fitting did not improve the results, the method was abandoned.

The positive fact is that the best method is obviously to combine the k - best fitting of curves with the n - period adjustment. Tests with four other days of data confirmed this result.

The value k for the test with the k – best fitting curves was k = 10. The value n for the n – period adjustment was n = 3. It may be interesting to search if the quality can increased even more by varying these two parameters. This is the object of the next chapter.

4.7 Second series of tests: optimizing good improvements

The goal of this second series of tests is to optimize the two values of k and n for the k – best fitting curves and the n – period adjustment respectively. A series of 96 tests with k in {5, 10, 15, 20, 25, 30, 35, 40} and n in {2, 3, 4} produce the following results:

k \ n	n = 2	n = 2	N = 2	n = 2	n = 3	n = 3	n = 3	n = 3	n = 4	n = 4	n = 4	n = 4
	MRE1	MRE2	MRE4	MRE8	MRE1	MRE2	MRE4	MRE8	MRE1	MRE2	MRE4	MRE8
k = 5	7.22	7.75	8.57	9.89	7.29	7.80	8.60	9.84	7.35	7.79	8.61	9.85
k = 10	7.19	7.70	8.48	9.72	7.23	7.73	8.54	9.66	7.31	7.72	8.55	9.64
k = 15	7.17	7.64	8.37	9.59	7.20	7.67	8.41	9.50	7.27	7.66	8.42	9.50
k = 20	7.17	7.62	8.33	9.46	7.21	7.67	8.37	9.40	7.26	7.64	8.36	9.39
k = 25	7.17	7.61	8.33	9.48	7.21	7.66	8.36	9.37	7.27	7.63	8.37	9.36
k = 30	7.16	7.60	8.29	9.40	7.21	7.65	8.35	9.34	7.27	7.63	8.36	9.33
k = 35	7.16	7.59	8.27	9.34	7.20	7.64	8.33	9.31	7.26	7.62	8.34	9.30
k = 40	7.16	7.59	8.27	9.33	7.20	7.64	8.33	9.29	7.26	7.61	8.34	9.28

Fig. 27, Test to optimize the k – best fitting curves and n – period adjustment

This table is limited to the average value of the MRE1 to MRE8 for the four time horizons. A first conclusion is that n = 3 seems to be very near to the optimum. Results for MRE1 are slightly better with n = 2, whereas results for MRE8 are slightly better for n = 4.

A value of n = 3 periods seems reasonable for the n – period adjustment at t0.

The astonishing fact was that the optimum for k seemed to lie beyond values of k = 40. The reason seems to lie in the high number of clusters that are candidate for the forecast. The system tends to compensate for the lack of typical clusters by choosing many of them and by reconstructing a typical one for each case. It may therefore not be the best solution to input as many clusters as possible. Instead of letting the system build his k – best clusters, one could reduce the number of initial clusters as well.

Another aspect is run time. For large applications like VM-CH, the search of the say 100 – best clusters for each of the 30’000 or more links at each time step could take too much time.

To check the influence of the number of clusters, a new series of 60 tests was programmed with cluster sets of c in {16, 32, 64, 128, 256, 512} and k – best fitting with k in {4, 8, 16}.

c \ k	k = 4	k = 4	k = 4	k = 4	k = 8	k = 8	K = 8	k = 8	k = 16	k = 16	k = 16	k = 16
	MRE1	MRE2	MRE4	MRE8	MRE1	MRE2	MRE4	MRE8	MRE1	MRE2	MRE4	MRE8
c = 512	7.32	7.89	9.10	11.01	7.34	7.92	9.03	10.64	7.28	7.85	8.88	10.47
c = 256	7.32	7.90	8.97	10.50	7.32	7.90	8.97	10.50	7.28	7.77	8.65	9.90
c = 128	7.32	7.85	8.79	10.12	7.32	7.85	8.79	10.12	7.32	7.85	8.79	10.12
c = 064	7.27	7.75	8.66	9.89	7.27	7.74	8.65	9.87	7.27	7.74	8.65	9.87
c = 032	7.46	8.00	8.91	10.01	7.46	8.00	8.91	10.01	7.46	8.00	8.91	10.01
c = 016	8.99	9.54	10.36	12.33	8.99	9.54	10.36	12.33	8.99	9.54	10.36	12.33

Fig. 28, Test to optimize the number c of clusters and the k – best fitting curves

The optimal number of clusters clearly lies around c = 64 and is quite flat. Anyway, it is interesting to discover that there is no need for a large set of thousands of clusters and that a few dozens are enough. There is a new way to paraphrase the well-known slogan: less (clusters) is more (quality).

The optimum for the k – best fitting is k = 8. An increase makes no sense. This is visible from the figures, as the results are the same for k = 8 and k = 16.

This is a nice confirmation of the hypothesis that a reduced number of clusters can be combined with a reduced number of k of best fitting clusters without reduction of the forecast quality.

The analysis will continue with the following values:

Choice of the k ≈ 8 best clusters in a set of c ≈ 64, adjustment at t0 with n ≈ 3 periods

It is not probable that forecasts for 2 hours will be much more accurate than the minimum value of about 10% for MRE. This value is anyway in the range of traffic measurement errors and will hardly ever decrease in the future.

4.8 Solving for the morning / evening asymmetry

As all typical DVC are together in one data set, a problem could arise because of the morning / evening asymmetry. Assume a forecast at 6 o'clock in the morning for an "inward" road directed towards the city centre. Among the best fitting DVC are some that have their origin on inward roads and others on outward roads. Even if the comparison period from 4 to 6 o'clock fits well, the DCV belonging to outward roads will produce poor forecasts for the morning peak on the inward road.

S. von der Ruhren, member of the steering committee, mentioned this potential source of poor forecasts. Thanks to the new fact that the number of typical DVC can be reduced to less than a hundred, the solution was not too hard to find: just keep one individual set of typical daily variation curves per link ! For links with section counts, this set will be updated once per day with measured DVC. For links without, the daily update will be done with the calibrated flow values.

It was not possible to add this capability and start a third series of tests if the project had to be terminated on time by the end of 2007. However, this elegant solution should be kept in mind and if possible added to the software used by VM-CH.

5 Forecasts with Error Propagation

The research report Error Propagation in Macro Transport Models [de Rham 06] is a valuable basis for the subsequent analysis. The study analyzes the behavior of errors through the different steps of transport models. The definition of error is repeated here, to avoid any misunderstandings:

Errors are not "bugs" but inherent and unavoidable uncertainties of measured data.

One of the findings of this study was that only a few variables are important for error propagation, mainly traffic volume q and time t (with speed proportional to 1/t). These results will now be applied to forecast methods and help estimate the forecast error at time t+15, t+30, t+60 and t+120.

5.1 From input data to present and forecasted value

Chapter 2.4 (Exponential smoothing to estimate error terms) develops a method to extract the error terms es of measured data directly by comparing raw values qm with exponentially smoothed values qs.

Chapter 4.1 (Measured data: best estimate and error term) shows how to introduce feedback to adapt the smoothing parameter α through the analysis of residuals. If residuals are large, α must be increased to get less dependent on past values. If residuals are small, α can be lowered to increase the weight of past values. This feedback is applied individually to every link to obtain measured smoothed flow values qs.

The result is a present value qs(t0) ± es(t0) for each measured link of the network. The assignment assigns the trips of the OD matrix and the calibration calibrates the matrix to get the best estimate of flow qc on every link of the network.

The qc values are then forecasted according to the methods described in 4.2 (Forecast with daily variation curves, basic procedure) and 4.3 (refinements). The result of these two steps are forecasted values qf for 15', 30', 60' and 120' ahead for each link of the network.

5.2 Estimation of the error term for the forecasted value

As the budget of this research does not allow to acquire a crystal ball (at least a good one), there is no possibility to estimate future errors in the future! One must wait until the future becomes the present to compare the forecast with reality and be able to project the comparison in future again. This means that the time interval is "lost" twice: first to know the error, second to apply it. As our greatest forecast horizon is 2 hours, this delay will be of a maximum of 4 hours.

This can be explained by the following rollover mechanism (forecast interval = x):

The forecast error ef(t0) is known from comparison of actual calibration data qc(t0) with forecasted data which are t0-x old:

$$ef(t0) = \Delta(qc(t0) - qf(t0\text{-}x)) / qc(t0)$$

The actual forecast is valid for t0+x in the future:

qf(t0+x) = qd(t0+x), from DVCs and cluster analysis

The forecast error is applied to the forecast of t0+x:

$$qf(t0+x) = qd(t0+x) \pm ef(t0)$$

or

qf(t0+x) = qd(t0+x) ± Δ(qc(t0) – qf(t0-x)) / qc(t0)

The forecast error for qf(t0+x) depends on forecast qf(t0-x). This means that the forecast error depends on data which were forecasted 2x back.

Here is an example for the maximum forecast interval of 2h: A forecast is estimated at 6 o'clock for 8 o'clock. The earliest time to know the forecast error will be 2h later, that is at 8 o'clock. This error will be applied for 2h in the future or 10 o'clock. The error will always lag behind the forecast by one forecast interval, that is 15', 30', 60' and 120' respectively.

5.3 Application to Short-term Forecast

The method developed in the previous chapter is illustrated with some real situations from heavy motorway traffic.

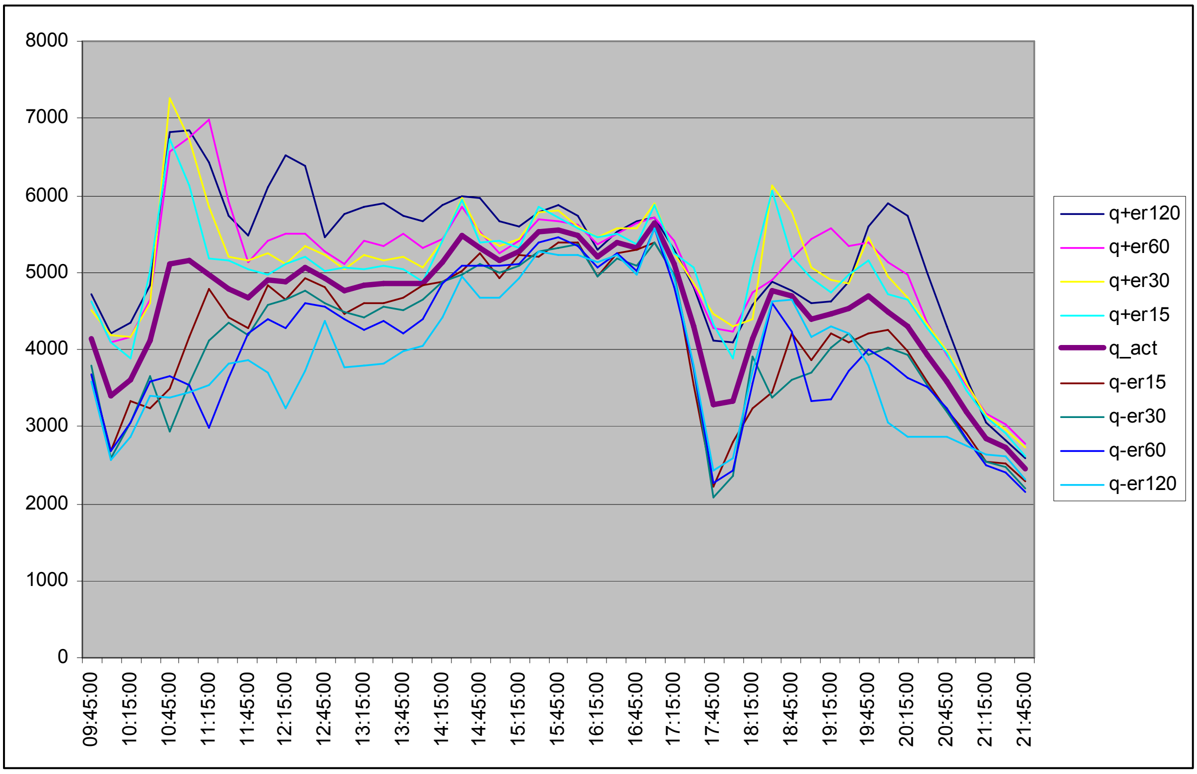

Fig. 29, Typical example with funnel effect from 10:45 to 16:45

The picture above shows a nice example of the behavior of the forecast error. The sudden increase of traffic from 3500 to over 5000 between 10:15 and 10:45 induces instability. From 10:45 on, the traffic oscillates around 5000 and the forecast error narrows steadily until 16:45. Then a sharp fall to about 3200 provokes a new period of instability recovering around 21:45.

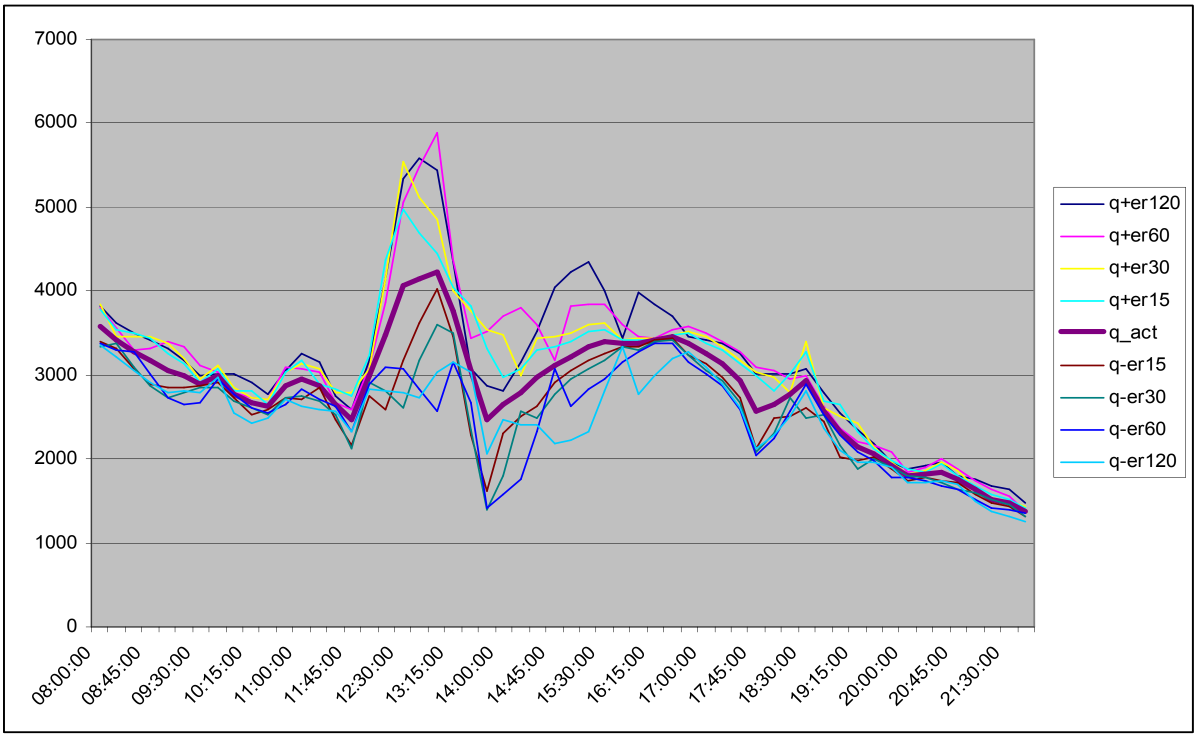

Fig. 30, Another example with funnel effect from 14h to 18h

Here too, the system takes some time to recover from the sharp peak around 13-14h. The quiet period after 19h shows that forecasts can be very precise, as long as traffic counts vary too much.

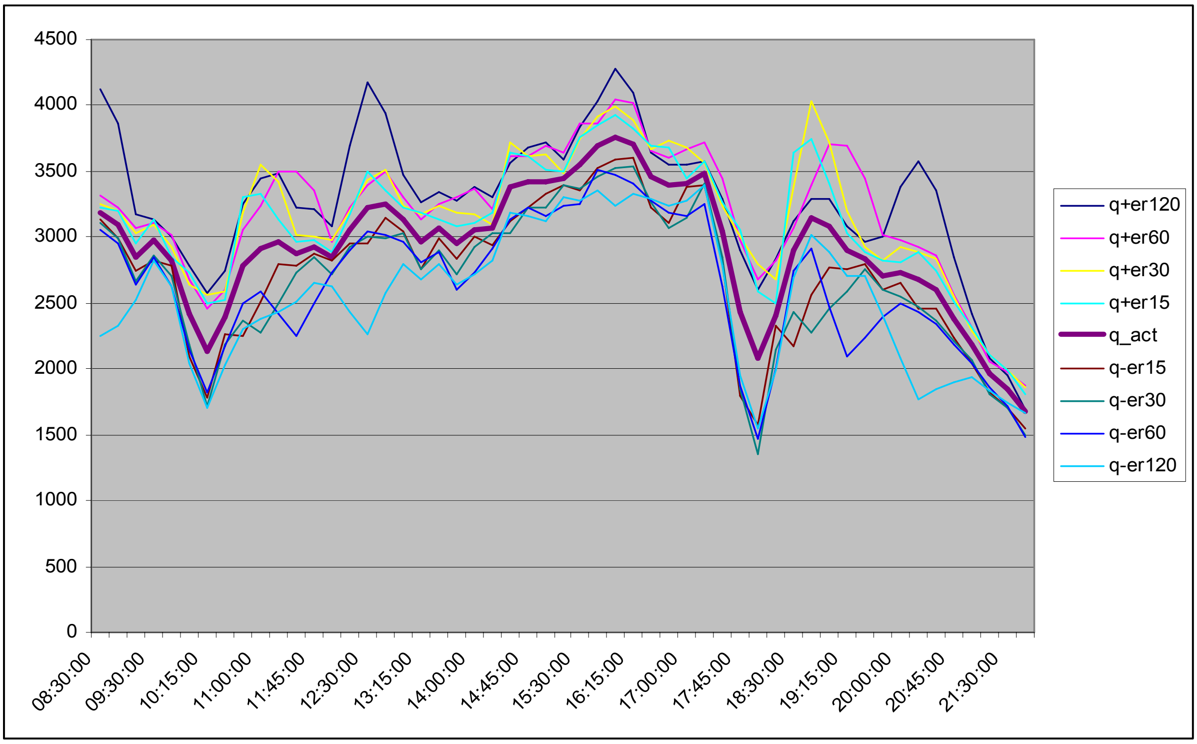

Fig. 31, Too many sharp rise and falls show the limits of the system

This figure shows the limits of the system. If there art too many sharp changes, the method will not be able to forecast the near future with accuracy. This is why forecasted values are only one side of the picture, the other being the accuracy of these values.

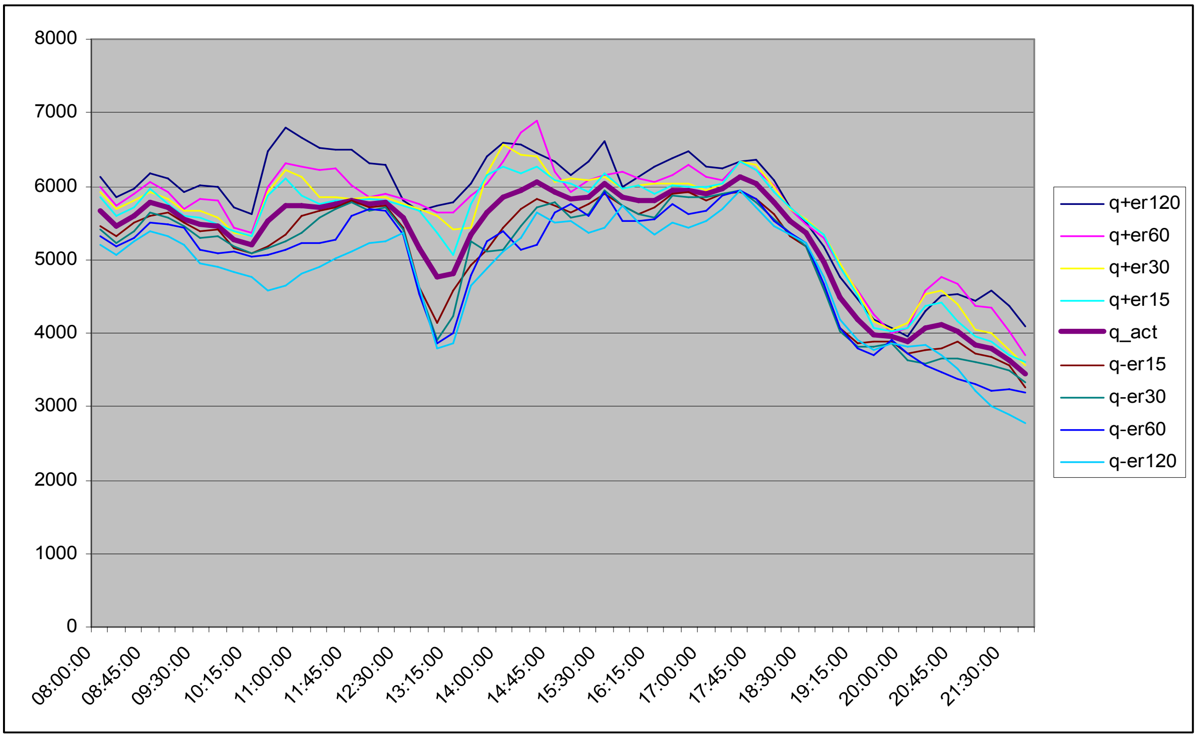

Fig. 32 The forecast is accurate for steady slopes: 17h - 20h

This is a mixed example with some large fluctuations but also a very quiet period from 17h to 20h. It confirms the fact that the forecast may be quite accurate, as long as local variations are small. This is independent of the absolute trend (raising or falling) of the actual traffic curve.

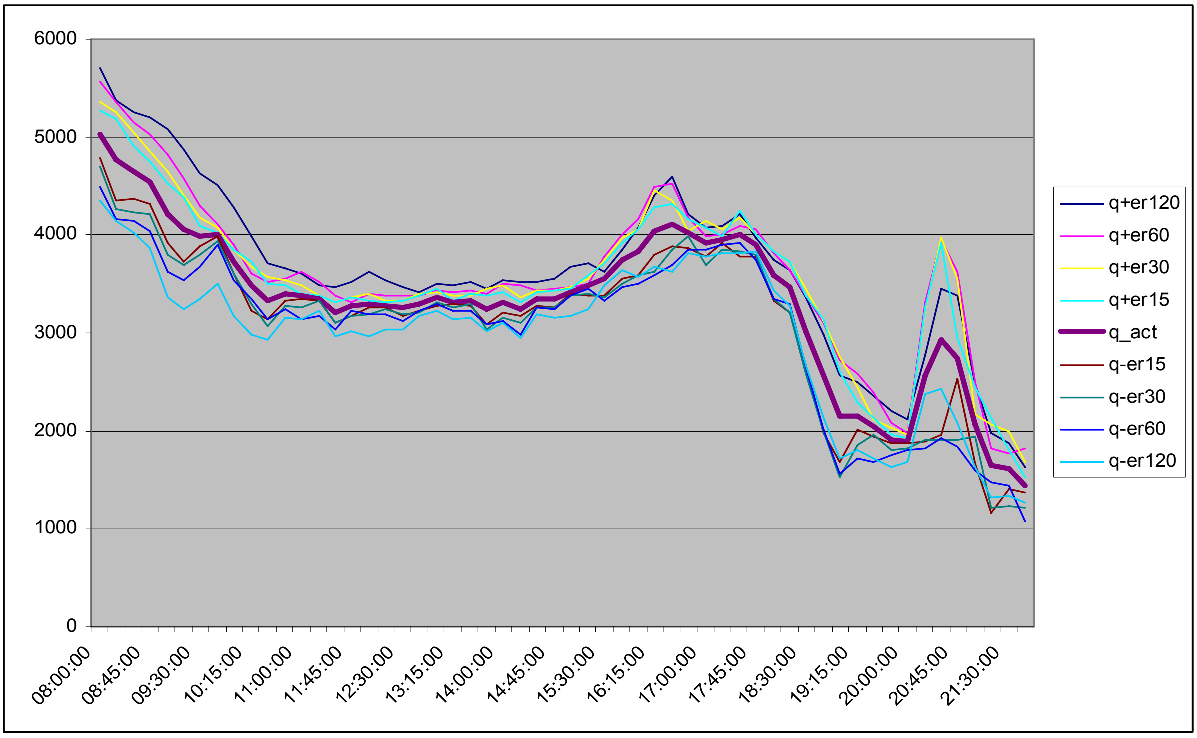

Fig. 33 Asymmetry of forecasts for up- and downward trends.

The forecast seems to be more comfortable with the downward trend (17:30-19:30) than the upward peak (after 20h). The reason for this is that many DVC follow a downward trend in the evening but probably very few, if any to follow the upward peak.

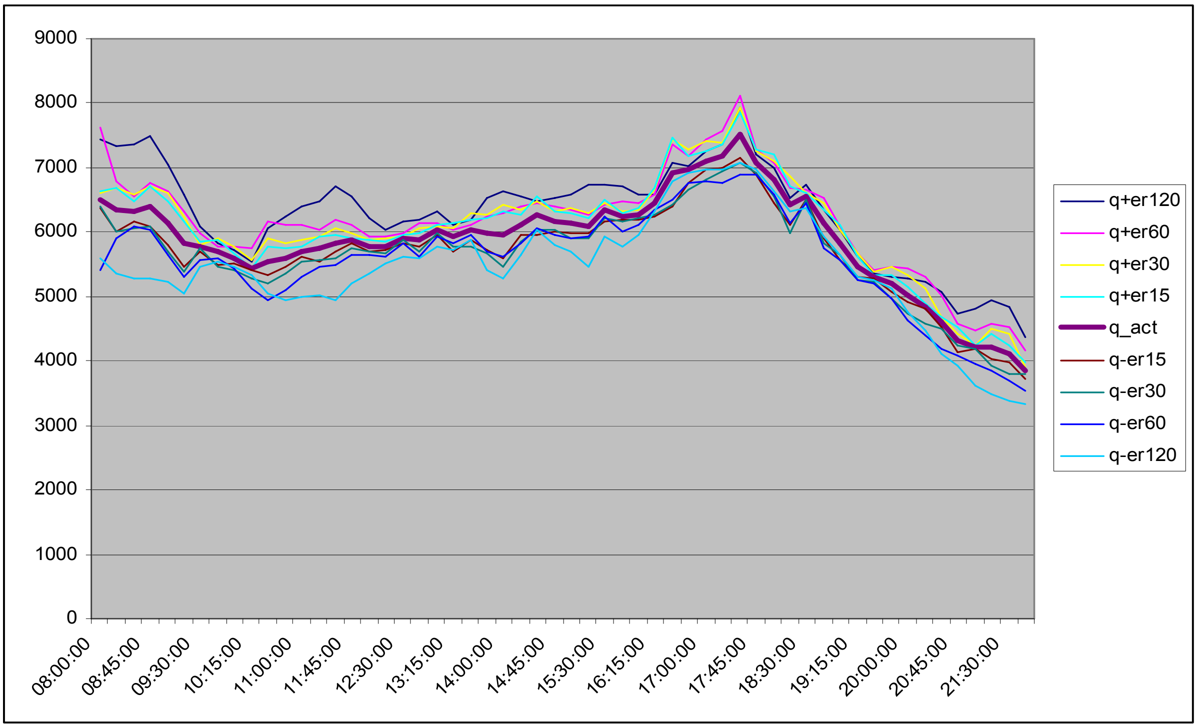

Fig. 34 Normal cases should look like this one!

6 Error Propagation for Quality Management

The integration of explicit error propagation into a transport modeling software opens new possibilities for online quality management, quality assessment, quality control, etc.

The idea is simple: if the quality (= errors) of the input is known and if the software is able to propagate this quality (= errors) throughout the whole calculation process, the output quality (= errors) will be known too.

Of course, the quality assessment of simulation software is not new. The traditional method uses the so-called "brute force" approach with repeated simulations.

Input	Process	Output
X	Y = f(X)	Y
X_i distributed with σ_x	$Y_i = f(X_i)$, means	Y_i distributed with σ_y

This methodology has many disadvantages, it takes effort, time and costs to setup the experiment, to create software for the handling of input and output data, to run the needed number of experiments and eventually to analyze the results with multivariate statistical tools. The consequence is that this technique is rarely applied and will never be widely spread.

The possibility of separating and following error propagation calculations in parallel to the normal processes simplifies many aspects:

Input	Process	Output
X	Y = f(X)	Y
σ_x	$\sigma_y = f(\sigma_x)$	σ_y

Where $\sigma_y = f(\sigma_x)$ are all the rules governing error propagation. The advantages are clear:

- Minimal software extension, thanks to operator overloading (like in ADA or C++),
- Marginal increase of runtime, no need to run many cases,
- No need for statistical tools and analysis, the results are directly available.

The application for online traffic modeling is straightforward. The input errors are the σ_{qm} of the counted flows Qm and the σ_{vm} of the measured speeds Vm on links m equipped with sensors (loops, cameras, etc.). The output errors are the σ_{qa} of the calculated flow Qa and the σ_{va} of the calculated speeds Va on all links of the network.

Input, on "measured" links m Actual values of Q and V	Process dynamic traffic model	Output, on all links a estimated situation
Qm	Qa = f(Qm, Vm)	Qa
σ_{qm}	$\sigma_{qa} = f(\sigma_{qm}, \sigma_{vm})$	σ_{qa}
Vm	Va = f(Qm, Vm)	Va ➔ Ta ➔ T_travel
σ_{vm}	$\sigma_{va} = f(\sigma_{qm}, \sigma_{vm})$	σ_{va} ➔ σ_{ta} ➔ σ_{t_travel}

The advantages are numerous:

- Continuous quality control of input, process and output,
- Input: targeted elimination of weak points in the measuring infrastructure,
- Process: choice of algorithms that minimize the effects of error propagation,
- Output: guaranteed quality of the traffic situation, information and travel times.

7 Conclusions

The goal of this study was to identify the best possible methods and their parameterization to apply short-term forecast to the online traffic model of VM-CH. Many methods were left aside from the beginning, mostly because they were too complex and/or could not guarantee better results.

A General remark: The quality of short-term forecasts for traffic is only limited by the inherent fluctuations of the traffic itself. This sets a lower limit to the possible precision of the forecasts. Measured in terms of mean relative error (MRE), this limit is around 10% for a time horizon of two hours. As long as random fluctuations of traffic lie in the same range, there is no hope to obtain better results.

The conclusions that follow concern the same four topics mentioned after Fig 1:

a) Exponential smoothing to clean measured data from local random fluctuations and to estimate the error term of these measures: The smoothing parameter should be monitored by a feedback loop. The analysis of the residuals of a simple autocorrelation suits well to increase or decrease the α in steps of 0.1.

b) Measures of fit to compare DVCs with the past two hours and choose the most probable DVC to be applied to the next two hours: The quality of forecast will be increased if one chooses not only one best fitting curve, but about 8 of them, and takes their mean value as reference for the next two hours. In addition to that, this curve should be adjusted at t0 with the mean of the 3 last periods (t-30', t-15', t0) instead of only t0.

c) As to cluster analysis to produce typical DVCs out of many individual ones by successive aggregation around reciprocal pairs: This method produces an equilibrate set of clusters where every type is fairly represented. The assumption that it is good to give a great number of clusters (thousands or curves) to the system was wrong. It is better to reduce the number of curves to say 50 to 100 and to smooth them with a simple mean-value algorithm. The additional advantage is that the run time will not pose any problem, even for large applications with tens of thousands of links, which are to be forecasted at every time interval.

d) Cluster analysis to add new curves to the set of existing typical curves: The method used for this step should be the same as for the reduction of the initial set of DVC. Every day, the newly measured DVC are added to the existing data set. This data set is reduced by cluster analysis until its dimension reaches the lower threshold of 50 to 100. This self-learning procedure guarantees stability. Either the new clusters are different from the old ones and will be the seed of new typical curves, or the differences are small and the new curves will find similar existing clusters to merge with.

8 Future Research

Here are some ideas and suggestions for further research.

Robustness of the method: This analysis was done with different data sets from different studies, but did not (and was not intended to) include long-term behavior of short-term forecast methods. Some questions remain about the robustness of the results during longer periods of time. What is the influence of the self adapting procedure with local daily variation curves in the long term?

Public transport: The effect of short-term forecast on systems depending on timetables and other time constraints does not seem to have been analyzed so far. Which methods and procedures could be suitable for reliable short-term forecasts in timetable based systems?

Floating Car Data: It is only a question of a few years until FCD plays a major role in incident detection, traffic monitoring, etc. Are short-term forecast methods based on traffic counts also suitable for calibration systems based mainly on FCD?

Transition to longer forecasting periods: The scope of this study was limited to methods applied to forecast horizons of two hours. What is their validity for horizons of say 4, 6 or 12 hours? When should it be recommended to include information about the type of the day, about special events, weather forecasts, etc.?

Solving for the morning / evening asymmetry (ch. 4.8): Morning peaks may be poorly forecasted on roads towards the city centre if the daily variation curve originates from outward roads, even if the historical 2 hours period from say 4 to 6 o'clock fits well. This problem can be solved by keeping an individual set of DVC per link. It was not possible to add this capability and test it if the project had to be terminated on time by the end of 2007. However, this elegant solution should be kept in mind and whenever possible added to the software used by VM-CH.

Pilot application for VM-CH: The traffic management center VM-CH will start its operations in 2008. **The research team strongly recommends to initiate a pilot project to implement these methods and monitor the results during about six months.** This time will be divided in three test cycles, each of them with observation and data collection, followed by analysis of results and corrective action.

9 Glossaries

Institutions and associations

ARE	Bundesamt für Raumentwicklung
ASTRA	Bundesamt für Strassen
BAV	Bundesamt für Verkehr
EPFL	Ecole Polytechnique Fédérale de Lausanne
ETHZ	Eidgenössische Technische Hochschule Zürich
IVT	Institut für Verkehrsplanung und Transporttechnik (ETH)
SVI	Vereinigung Schweizerischer Verkehrsingenieure
TBA	Tiefbauamt
UVEK	Eidg. Departement für Umwelt, Verkehr, Energie und Kommunikation
VM-CH	Verkehrsmanagement Schweiz
VSS	Schweizerischer Verband der Strassen- und Verkehrsfachleute

Other abbreviations:

CA	cluster analysis
CVJ	Courbe de variation journalière
DVC	Daily variation curve
FCD	Floating Car Data
MAE	Mean absolute error
MRE	Mean relative error
OD	Origin Destination (matrix)
QZ	Quell Ziel (Matrix)
RMSE	Root mean square error
RSQ	R2, correlation coefficient
SQRT	square root
TGL	Tagesganglinie

10 Short-term Forecasts in Switzerland

A small survey about short-term forecasts applications in Switzerland was conducted in May 2007 among 20 public and private instances. Here is the text, which was sent by the secretariat of the SVI (in German):

Sehr geehrte Kollegen,

Ziel dieses Projektes ist, die neue Zentrale VM-CH (Betrieb ab Frühjahr 2008) mit bestmöglichen Methoden für Kurzfristprognosen (bis 2h) auszurüsten. Einerseits werden bekannte Methoden auf Grund historischer Ganglinien der ASTRA-Zähler bezüglich ihrer prädiktiven Leistung verglichen. Anderseits wird im SVI diese Mini-Umfrage gestartet.

Das ASTRA möchte aus Erfahrungen mit Kurzfristprognosen in der Schweiz lernen. Darf ich Sie bitten, diese 7 Fragen im Text bis zum 15. Mai 2007 per Mail an die Geschäftsstelle SVI (info@svi.ch) zu beantworten. Ich werde mir erlauben Sie nach dem 15. Mai 2007 persönlich zu kontaktieren.

Besten Dank für Ihre Mitarbeit

Q1: Entwickeln / nutzen Sie Kurzfristprognose-Methoden? Wenn ja, sind Sie bereit, Ihr Wissen zu teilen? (wenn nein STOP, danke dass Sie bis hier gelesen haben!)

Q2: Welchen Input nutzen Sie? Typ der Daten, Menge, Frequenz, usw.

Q3: Welche Methoden nutzen Sie? Wie ist ihre prädikative Leistung?

Q4: Welchen Output erhalten Sie? Wozu nutzen Sie diesen Output?

Q5: Wenden Sie Qualitätskontrolle an? Wenn ja, wie?

Q6: Was sind Ihre Erfahrungen / Kommentare (pos./neg.!) / Empfehlungen an das ASTRA?

Q7: Kontaktperson? Name, Tel.Nr., EMail?

The results were somewhat disappointing:

11 answers:	no forecasts applied
6 answers:	no answer
2 answers:	forecasts, but one for “days/weeks” and the other for “years”
1 answer:	yes short-term forecast in use
20 answers total	

Description of the only one answer from Mr Marc Laube, IVT-ETH:

A1: Ja
A2: Einzelfahrzeugdaten analog der Zählstellen ASTRA
A3: Prognosemodell mit gleitender Korrektur. Dabei werden mit Fünf-Minuten-Mittelwerten gerechnet (die Einzelfahrzeugdaten in Fünf-Minuten-Intervalle zusammenfassen).
A4: Belastung pro Fahrstreifen [Fz/h] aus Steuerungs-Algorithmus und LKW - Anteil zum Steuern einer dynamischen Signalisation eines LKW – Überholverbots
A5: Meldung bei fehlenden Daten, Meldung bei Fehler bei Einzelfahrzeugerfassung
A6: Das System ist seit rund einem Jahr in Betrieb. Die Steuerung arbeitet zuverlässig und der gewählte Algorithmus bildet den Verkehrsablauf sehr genau ab.

11 Acknowledgements

The authors would like to express their gratitude to the Swiss Federal Office of Roads, which sponsored this study from March to December 2007.

They thank the president and members of the steering committee who gave active support by mentioning literature concerning the subject, by adding valuable remarks during the meetings, and by taking the necessary time to review and comment the successive drafts.

Our thanks also go to two German institutions, the HLSV (Hessisches Landesamt für Strassen- und Verkehrswesen) and the LST (Landesstelle für Strassentechnik of Land Baden-Württemberg). The HLSV contributed by giving the permission to use a large set of daily variation curves and the LST by giving the permission to use the data from an actual study concerning the motorway traffic on the rectangle A5 / A6 / A8 and A81.

The authors are also thankful to Mrs. Michelle Sisto and Mrs Martine de Montmollin, both Professors at the International University of Monaco. Mrs. Sisto gave many and highly helpful advices during the statistical analysis and Mrs. de Montmollin took the time to do the final proof reading of the text.

12 References

[Bevington] Philip R. Bevington and D. Keith Robinson, Data Reduction and Error Analysis for the Physical Sciences, 3rd ed., McGraw Hill 2003

[FGSV 365] Hinweise zur Schätzung von Verkehrsbeziehungen mit Hilfe von Querschnittszählungen, FGSV, AG Verkehrsführung und Verkehrssicherheit, Köln 1995

[Halpern] Joseph Y. Halpern, Reasoning about Uncertainty, The MIT Press 2003

[de Jong] G. de Jong, A. Daly, M. Pieters, S. Miller, R. Plasmeijer, F. Hofman, Uncertainty in Traffic Forecasts: Literature Review and New Results for the Netherlands. European Transport Conference, Strasbourg 2005

[Kirschfink] H. Kirschfink, C. Chadenas, Traffic Situation Prediction Applying Pattern Matching and Fuzzy Classification, Traffic Situation Prediction Applying Pattern Matching and Fuzzy Classification, European Symposium on Intelligent Techniques, June 3-4, 1999 Orthodox Academy of Crete, Greece

[Kirschfink 94] H. Kirschfink, U. Uerlings, The Prediction System within the SOCRATES Information Centre, Proceeding of the 2nd DRIVE II workshop on Short-Term Road Traffic Forecasting, Delft, Holland 1994

[Peters, 05] J.-C. Peters, Qualitätsüberwachung und Mustererkennung verkehrstechnischer Zeitreihendaten, Tagungsbericht Heureka'05, Karlsruhe, 03/2005

[Ortuzar, Willumsen] J. de D. Ortuzar, L.G. Willumsen, Modelling Transport, Wiley 1990

[Rabinovich] Seymon G. Rabinovich, Measurement Errors and Uncertainties, Theory and Practice, 2nd ed., Springer 2000, ISBN 0-387-98835-1

[de Rham 80] C. de Rham, La classification hiérarchique ascendante selon la méthode des voisins réciproques, Les Cahiers de l'Analyse des Données, Vol V, 1980, no 2, p135-144.

[de Rham 86] C. de Rham, Herleitung von Verkehrsbeziehungen des öffentlichen Verkehrs aus Verkehrszählungen in Zügen. Auftrag Nr. 5-A111 Stab für Gesamtverkehrsfragen 1986

[de Rham 87] C. de Rham, Umlegung, Überprüfung und Korrektur von Wunschlinien des Personenverkehrs anhand von Querschnittszählungen. Auftrag Nr. 5-A133 Stab für Gesamtverkehrsfragen 1987

[de Rham 03] C. de Rham, Incident Detection with Floating Car Data, paper #4069, 10TH World Congress on Intelligent Transport Systems 2003 Madrid

[de Rham 04] C. de Rham and Michelle Sisto, How to Quantify the Quality of Matrix Estimation with Traffic Counts? paper #2799, Congress on Intelligent Transport Systems in Europe 2004 Budapest

[de Rham 06] C. de Rham, R. Schwarz, W. Schaufelberger, ERROR PROPAGATION IN MACRO TRANSPORT MODELS, Research report UVEK/ASTRA No 1157, June 2006

[Roughgarden] Tim Roughgarden, Selfish Routing and the Price of Anarchy, The MIT Press 2005

[von der Ruhren] Stefan von der Ruhren, Kurzfristprognosen von Verkehrszuständen auf Basis von Verfahren der Mustererkennung und von dynamischen Routensuch- und Umlegungsverfahren, Dissertation an der TH Aachen, 2006

[Taylor] John R. Taylor, An Introduction to Error Analysis, 2nd ed., University Science Books 1997