



Kalibrierung und Validierung von Mikroskopischen Verkehrsflussmodellen

Elmar Brockfeld, Peter Wagner

Institut für Verkehrsforschung, DLR, Berlin

Braunschweig, 08.01.2004



Inhalt

- **Verkehrsflussmodellierung (Strassenverkehr)**
 - Anwendungsgebiete von Modellen**
 - Modellklassifizierung**
 - Verkehrszustände und -phänomene**
 - Beispiel mikroskopisches Modell: Krauss-Modell**
 - State of the art: Mikroskopische Verkehrsfluss-Modellierung**
- **Modelltestbeispiel Teststrecke Hokkaido (Japan)**
 - Datenbeschreibung und Modellierungsansatz**
 - Ergebnisse Kalibrierung und Validierung**
- **Modelltestbeispiel San Pablo Dam Road (USA)**
 - Datensatzbeschreibung und Modellierungsansatz**
 - Ergebnisse Kalibrierung und Validierung**
- **Zusammenfassung und Ausblick**



Anwendungsgebiete von Verkehrsflussmodellen

Wesentliche Anwendungsgebiete von Verkehrsflussmodellen:

- ▶ Verkehrs- und Infrastrukturplanung
- ▶ Verkehrsflusssteuerung (Auffahrten, Ampelsteuerung, Routenempfehlungen)
- ▶ Kurzfristprognose auf Relationen oder in Netzen (bis ca. 30 oder 60 Minuten voraus)
- ▶ Entwicklung von Fahrerassistenzsystemen



Verkehrsflussmodellierung

Modell-Klassifizierung nach der räumlichen Auflösung:

Makroskopisch (1950er...): Einteilung des Straßennetzes in Abschnitte, auf denen sich auf Grundlage von Differential- bzw. Differenzgleichungen der Verkehrszustand mit der Zeit ändert.
(**linkbasiert**)

Makroskopische Eigenschaften werden über abstrakte Parameter eingestellt.

Beispiele: Lighthill-Whitham, Herman, Kerner-Kohnhäuser-Kühne, Cremer

Mikroskopisch (70er...):

Jedes einzelne Fahrzeug wird modelliert.

(**fahrzeugbasiert**)

Makroskopische Eigenschaften ergeben sich aus der Interaktion der Fahrzeuge untereinander.

Beispiele: Nagel-Schreckenberg (CA), VDR, PELOPS, SK, OVM, Wiedemann

Mesoskopisch (Ende 70er...):

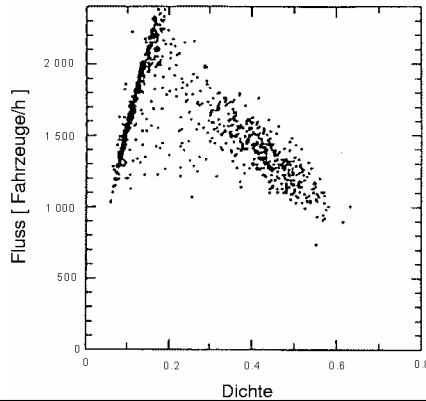
Vereinfachte mikroskopische Modellierung der einzelnen Fahrzeuge unter makroskopischen Randbedingungen.

(**Mischform**)

Beispiele: Animal, DYNEMO, FASTLANE

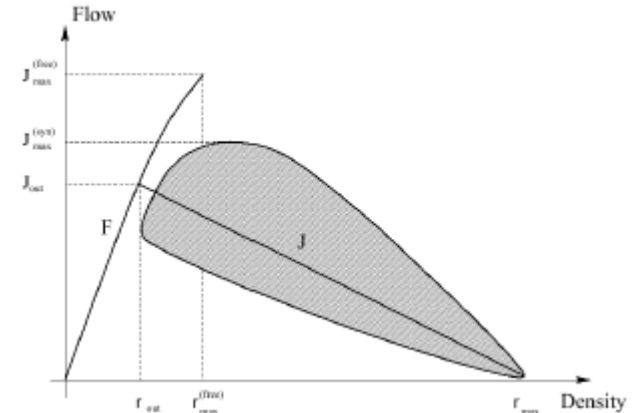


Verkehrszustände und -phänomene



Fundamentaldiagramme

empirisch schematisch



Freier Verkehr

Die Fahrzeugführer können ihre Geschwindigkeit unabhängig von vorausfahrenden Fahrzeugen wählen

Teilgebundener Verkehr

Die Fahrzeugführer sind zumindest zeitweise durch andere Fahrzeuge in ihrer Geschwindigkeitswahl eingeschränkt. Es bilden sich Fahrzeugpuls unterschiedlicher und wechselnder Länge.

Gebundener Verkehr

Alle Fahrzeugführer sind in ihrer Geschwindigkeitswahl mehr oder weniger durch vorausfahrende Fahrzeuge eingeschränkt.

Spontane Staubildung / "Stau aus dem Nichts"

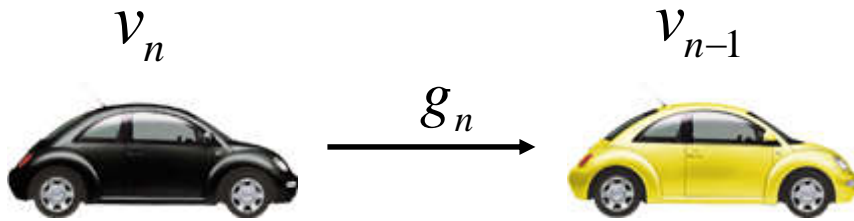
Staus, die - anders als bei Unfällen oder Baustellen - ohne offensichtlichen Anlass entstehen.

Moving Jams

große, relativ kompakte Staus (hohe Dichte)



Mikroskopisches Modell von Krauß (1996)



v_{\max}	Maximalgeschwindigkeit [m/s]
a	Beschleunigung [m/s ²]
b	Bremsbeschleunigung [m/s ²]
ε	Max. zusätzliche Bremsbeschleunigung [m/s ²]
τ	Reaktionszeit eines Fahrers [s]

Prüfung, ob Sicherheitsabstand eingehalten wird und errechnen der "sicheren" Geschwindigkeit v_{safe}

$$d(v_n) + \tau * v_n \leq d(v_{n-1}) + g_n \quad \Rightarrow \dots \quad v_{\text{safe}} = v_{n-1} + b * \tau \frac{g_n - v_{n-1}}{v_{n-1} + b * \tau}$$

Errechnen der Wunschgeschwindigkeit v_{des}

$$v_{\text{des}} = \min\{v(t) + a, v_{\text{safe}}, v_{\max}\}$$

Berechnung der neuen Geschwindigkeit und Fahrzeugbewegung

$$v(t+1) = \max\{v_{\text{des}} - \varepsilon * p, 0\} \quad , \text{ mit } \quad p \in [0..1]$$

$$x(t+1) = x(t) + v(t+1)$$



„State of the art“

Die Situation in der mikroskopischen Verkehrsflusssimulation heute:

- ▶ Es existiert eine sehr große Anzahl an Modellen, die den Verkehrsfluss beschreiben
- ▶ Es gibt keine etablierten Methoden, wie man mikroskopische Verkehrsmodelle testen sollte.
- ▶ Wenn sie getestet werden, geschieht dies zumeist alleine mit speziellen Datensätzen.
- ▶ Die exakte Qualität der Modelle im Vergleich zur Realität ist derzeit nicht klar.

→ Die Modelle sind quantitativ nicht vergleichbar



„State of the art“

Ideen

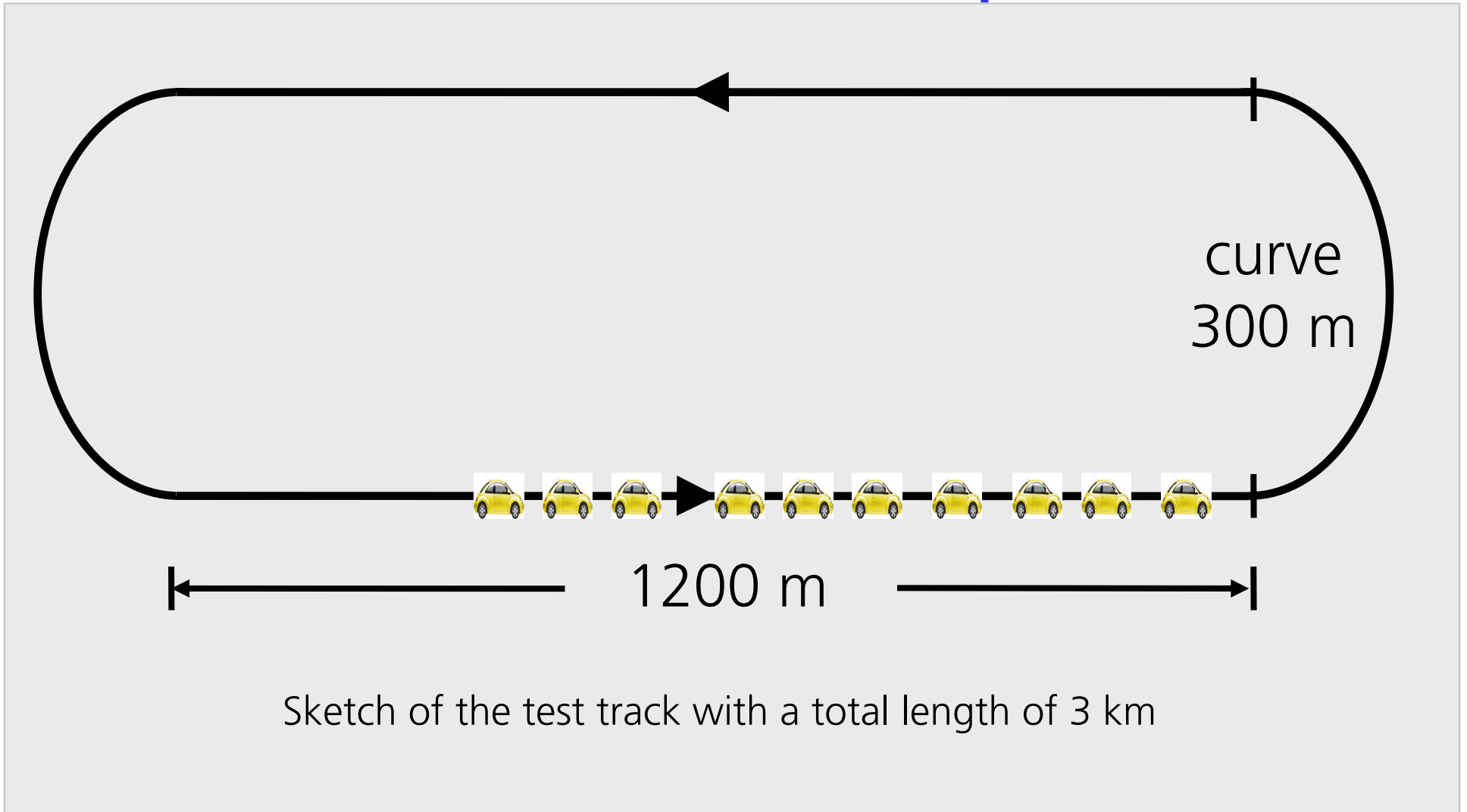
- Vergleiche mikroskopische Verkehrsflussmodelle, indem sie alle mit den gleichen Datensätzen kalibriert und validiert werden. (→ **quantitative Vergleichbarkeit, Benchmark möglich ?**)
- Kalibrierung und Validierung auf möglichst mikroskopische Weise, indem Zeitreihen, die durch einzelne Fahrzeuge erzeugt werden, untersucht werden.

Im Folgenden

- Kalibrierung und Validierung von zehn Modellen mit Datensätzen von einer japanischen Versuchsstrecke. Analyse über das Abstandsverhalten.
- Kalibrierung und Validierung von neun Modellen mit Datensätzen aus den USA. Analyse über Reisezeiten auf Strassenabschnitten.
- In beiden Fällen: Einspurige Strasse ohne Fahrbahnwechsel und somit auch ohne Überholvorgänge!



Test track Hokkaido, Japan



Sketch of the test track with a total length of 3 km



Test track Hokkaido, Japan

Impressions





Hokkaido, Japan – The data

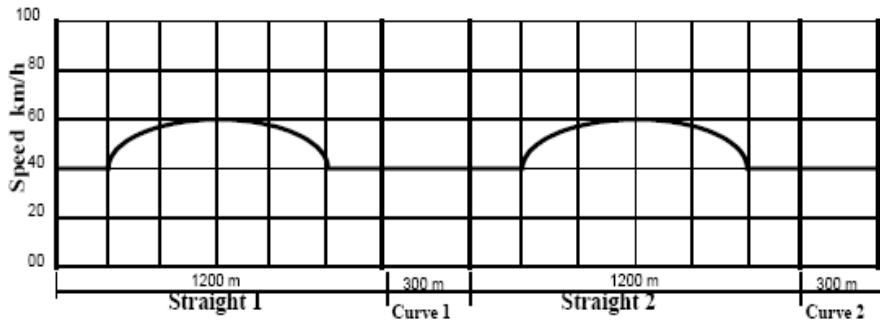
- » Data recorded by Nakatsuji in 2001 (see Gurusinghe et al., TRB 03-41371).
- » Data from ten GPS-equipped cars driving on a 3km test track in Hokkaido, Japan
- » GPS positions (accuracy 1 cm) in intervals of 0.1 second
- » Derived speeds with errors of less than 0.2 km/h
- » Seems to be adequate for the analysis of car following behavior
- » The data of four experiments, each with ten cars driving 15-30 minutes, are used for the analyses. Details:

experiment	Duration [min]	Full loops
„11“	26	6
„12“	25	7
„13“	18	6
„21“	14	4

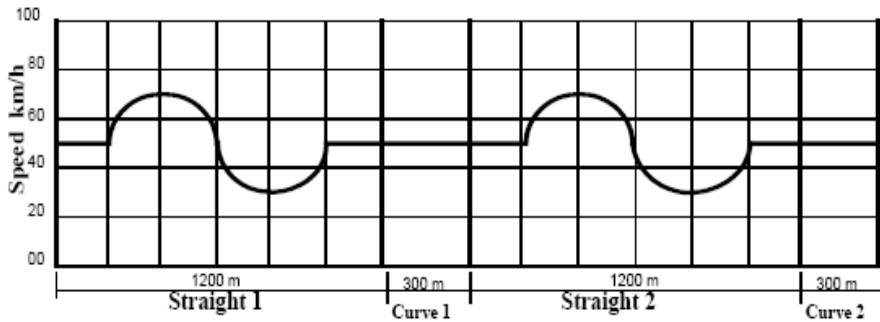
- » To minimize driver dependent correlations the drivers were exchanged between the cars after each experiment
- » First car performed certain “driving patterns” (see next slide) on the straight sections



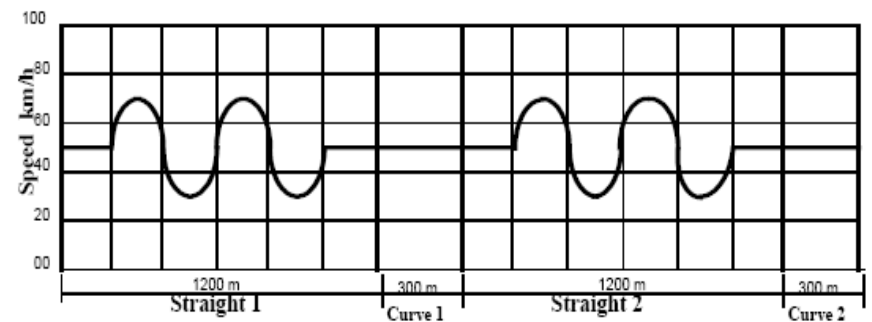
Hokkaido, Japan – Driving patterns



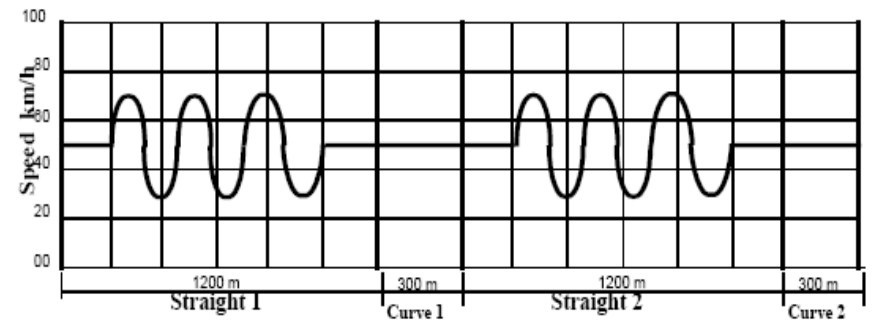
(a) Half wave



(b) One wave



(c) Two waves



(d) Three waves

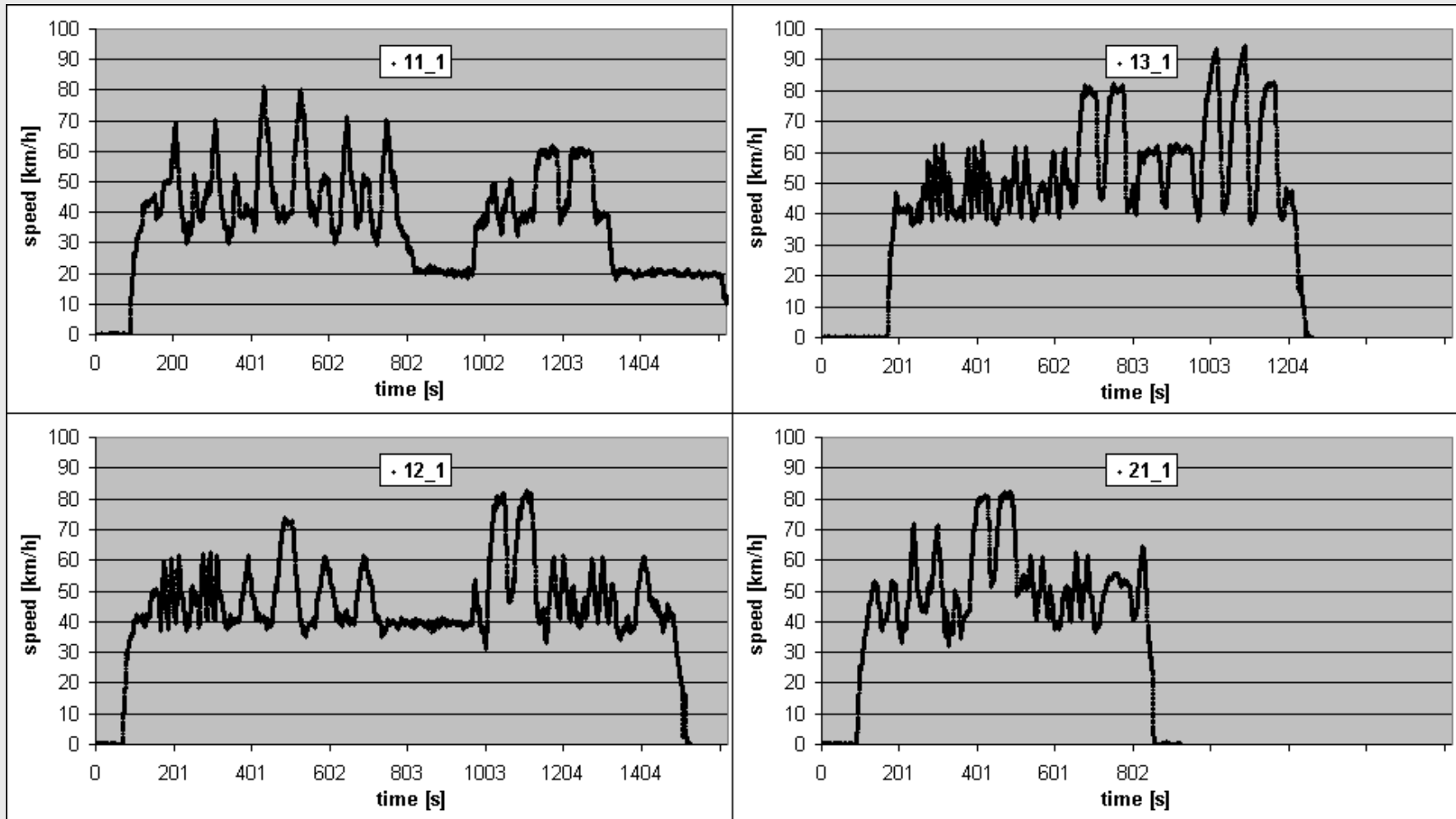
Source: Gurusinghe et al., TRB 03-41371

Driving patterns:

- » constant speeds of 20, 40, 60 and 80 km/h;
- » driving in waves varying from about 25 to 65 km/h (half, single, double and triple waves on a straight 1200 meter long section)



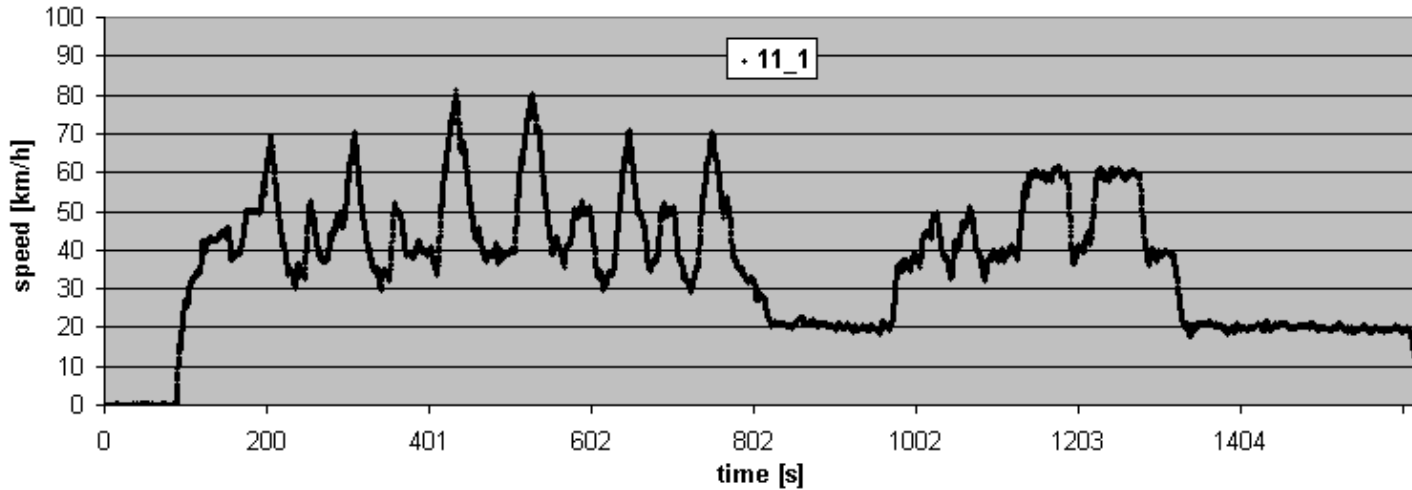
Hokkaido, Japan – Speed development



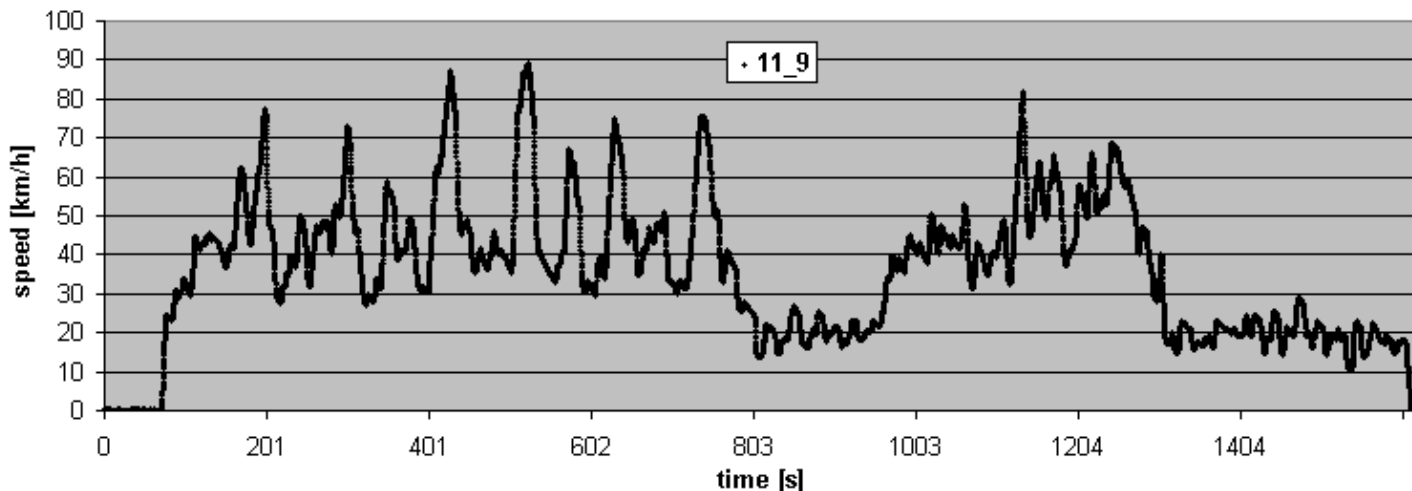
Speed development of the leading car in all four experiments



Hokkaido, Japan – Speed development



**leading car
(data set 11)**



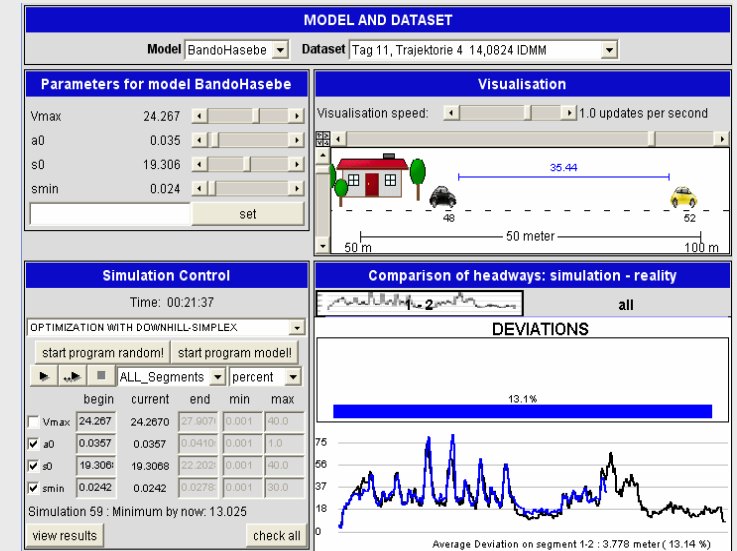
**last car
(data set 11)**



The models

The following models have been analysed:

- » **4 parameters**, CA0.1 („cellular automaton model“)
- » **4 p**, OVM (“Optimal Velocity Model“)
- » **6 p**, GIPPSLIKE (basic model by P.G. Gipps)
- » **6 p**, Aerde (model used in the software INTEGRATION)
- » **6 p**, IDM (“Intelligent Driver Model“)
- » **7 p**, IDMM (“Intelligent Driver Model with Memory“)
- » **7 p**, SK_STAR (model based on the model by S. Krauss)
- » **7 p**, Newell (CA-variant of the model with more variable acceleration and deceleration by G. Newell)
- » **13 p**, FRITZSCHE (model used in the british software PARAMICS; similar to what is used in the german software VISSIM by PTV)
- » **15 p**, MitSim (model used in the software MitSim)



Java Applet for testing the models



The model's parameters

Parameters used by all models:

V_{\max} Maximum velocity

l Vehicle length

All models except for the CA:

a acceleration

Most models:

b deceleration

τ reaction time

Some models:

a and b depend on the current speed or the headway

Random braking used by:

CA0.1, SK_STAR, MitSim



Some special model features

Special features for IDMM, MitSim and FRITZSCHE:

IDMM:

Memory effect: vehicles try to hold their speed according to a rollin horizon depending on the density ahead.

MitSim:

Three driving modes: very complicated if very close to leader. Depends on headway, own speed, speed-difference and current density. Needs the speed of the leader one time step before as a special feature.

FRITZSCHE:

Various driving modes: (dV, g) -plane is divided into regions of free driving, approaching, emergency brake and two other driving behaviors. Needs acceleration of follower and leader one timestep before. Used some kind of brake light of the leader by reacting on its deceleration.



Hokkaido, Japan - Simulation setup



- » For each simulation run one vehicle pair is under consideration
- » The first car is updated with the speeds as recorded in the data
- » The second car is updated following the rules of the traffic model under consideration

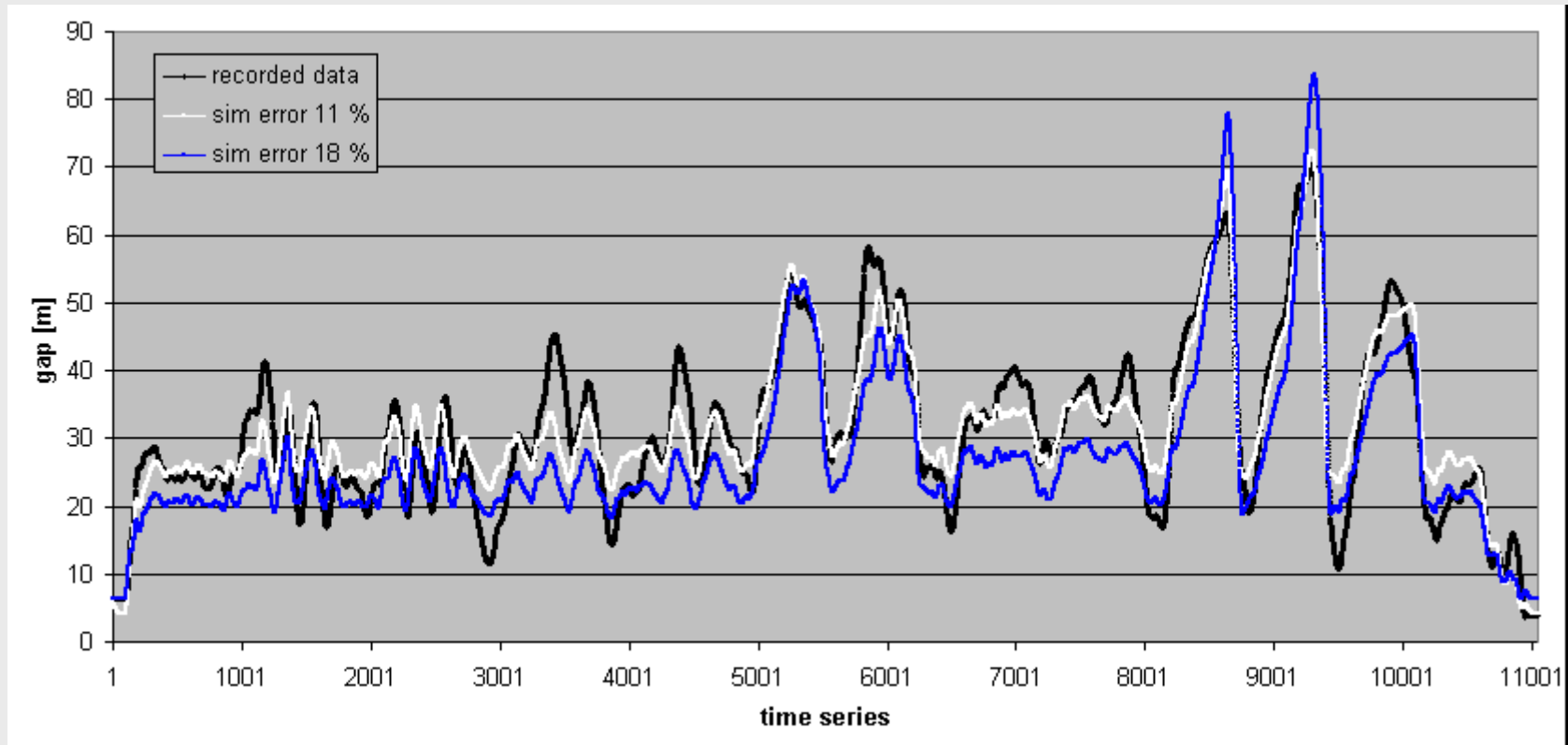
- » For error measurement the percentage error e is calculated by taking the absolute differences of the observed gaps g_{obs} to the simulated gaps g_{sim} and relating it to the mean gap in each of the observed data sets, respectively, over the complete simulation time T (with $t \in \{0; 0.1; \dots; T\}$).

$$e = \frac{\frac{1}{T} \sum_{t=1}^T |g^{(sim)}(t) - g^{(obs)}(t)|}{\frac{1}{T} \sum_{t=1}^T g^{(obs)}(t)}$$

- » To find the optimal parameters for the models a gradient free (direct search) optimisation technique is used named downhill simplex
- » The optimisation process is started with random initialisation several times to avoid local minima.

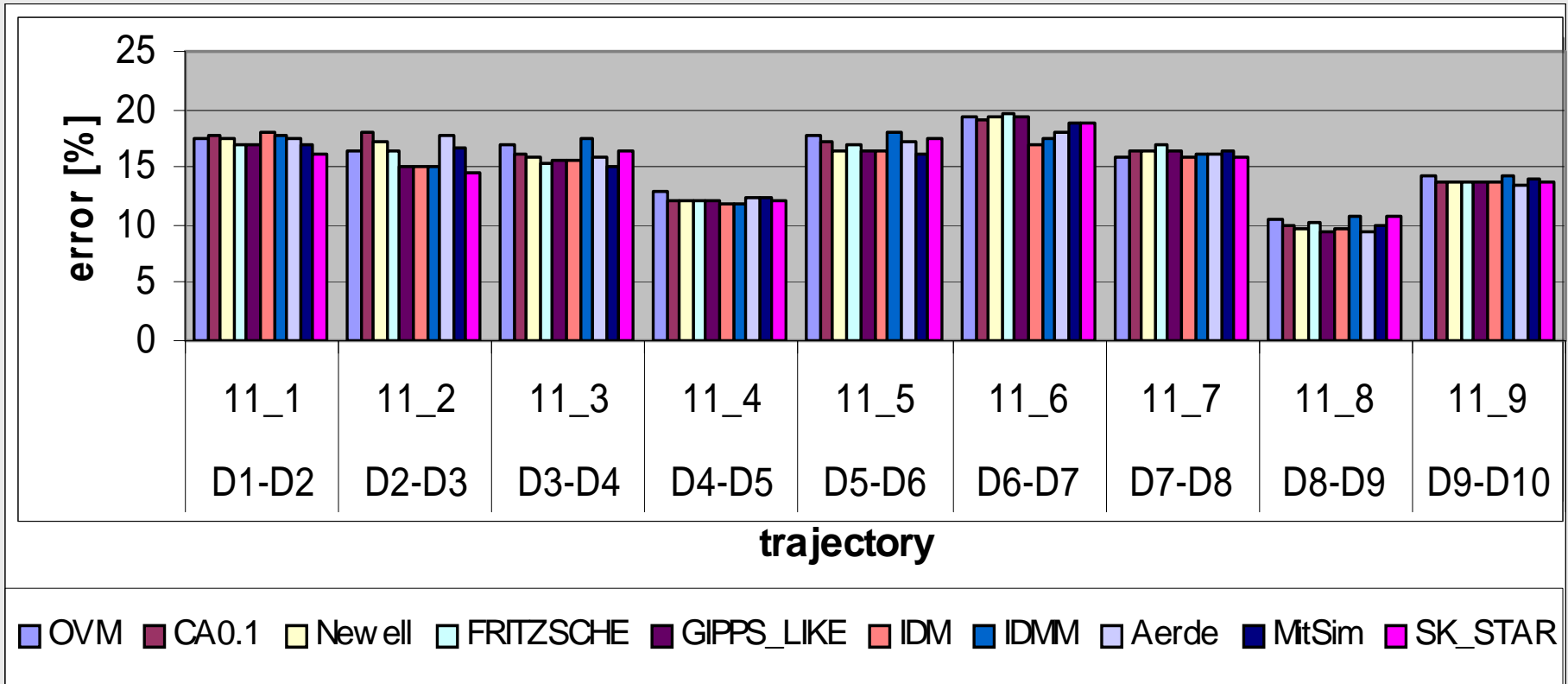


Hokkaido, Japan - Gap development





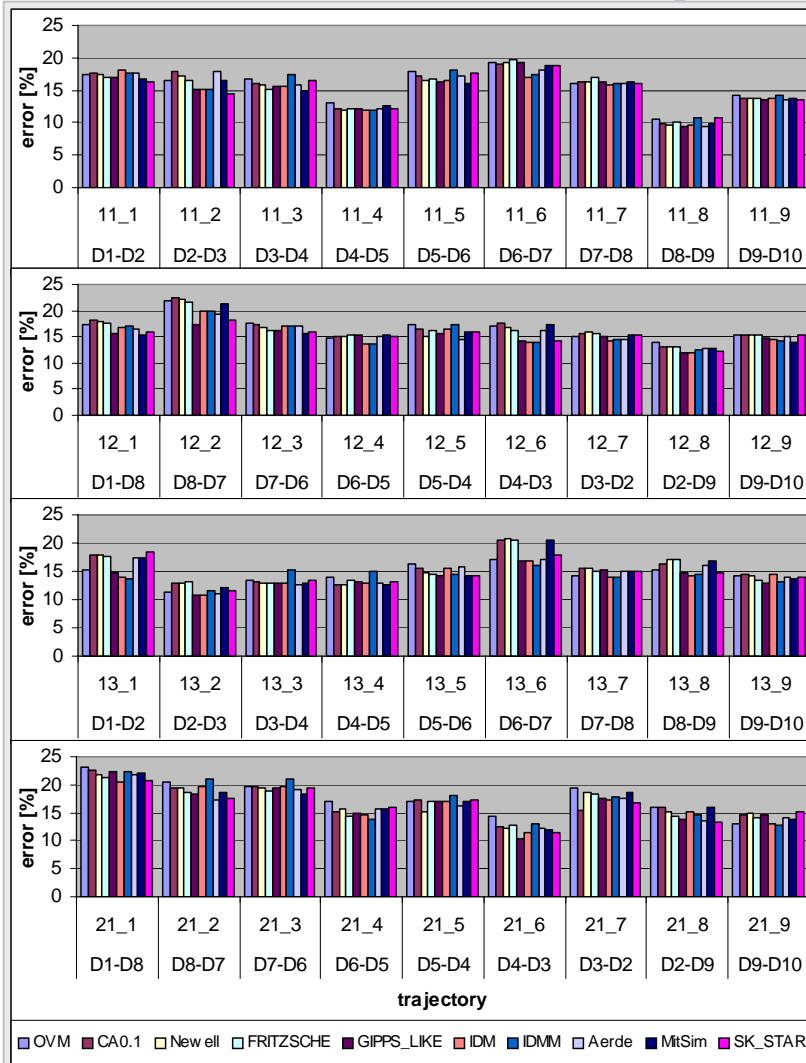
Hokkaido, Japan - Calibration Results (1/2)



- » Results for the first experiment ("11")
- » All models share the same problems with the same data sets
- » Errors between 9 and 19 %, mostly between 13 and 17 %



Hokkaido, Japan - Calibration Results (2/2)



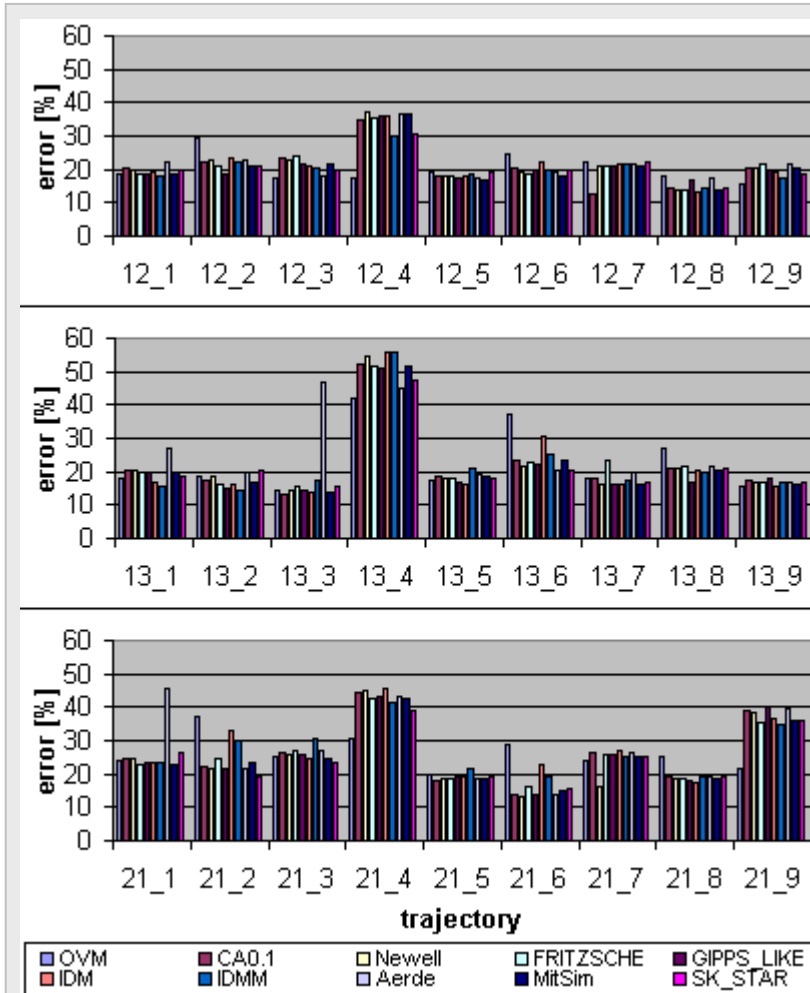
Calibration of 10 models with 36 data sets (four experiments „11“, „12“, „13“ and „21“, each with 9 driver pairs; „D“ = Driver; leading driver is always „D1“):

- » Error rates from 9% to 24 %, mostly between 12% and 17 %
- » No model appears to outperform the others regularly; the average error the best model produces is 15.14 %, that of the worst model 16.20 %
- » Diversity in the driver behaviour (5%-15%) is bigger than diversity of the models (average difference per data set is 2.5 %)
- » All models share the same problems with the same data sets
- » For these data sets models with more parameters do not produce better results



Hokkaido, Japan -Validation Results

driver-independent



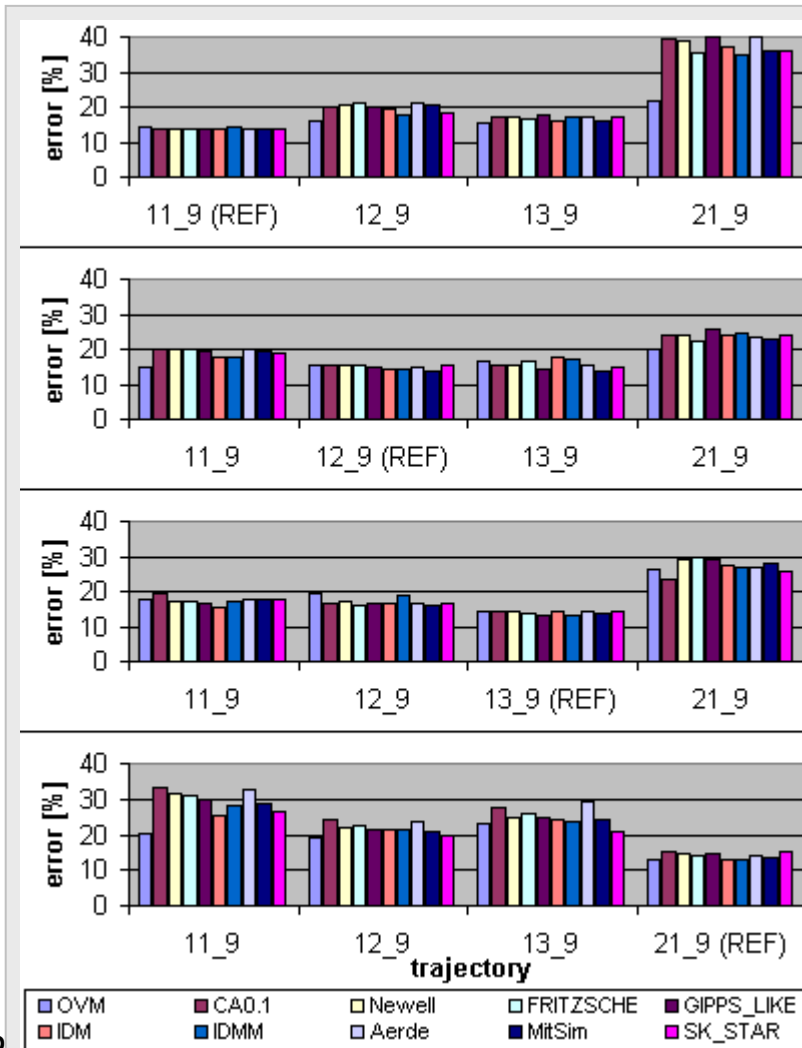
The resulting 9 parameter sets for each model of experiment „11“ are validated with the other three „12“, „13“ and „21“.
 (11_1 -> 12_1, 13_1, 21_1; 11_2 -> 12_2, 13_2, 21_2...):

- » Error rates are mostly between 17% and 22%
- » Additional error in comparison to calibration is about 3.2% to 5.5 % for most models
- » “Overfitting” of data set 11_4 produces high errors up to 55% for data sets 12_4, 13_4 and 21_4
- » Special driver behaviour (driver pairs 11_4 and 21_9) because of exceptional bad validation results.
- » Small problems with the models OVM (6.6% worse than calibration) and Aerde (7.6% worse) in some cases and singular high errors for special data sets



Hokkaido, Japan - Validation results

driver-special



Validation only with driver pair D9-D10, which drove consecutively in each of the four experiments.

- » Special driver behaviour in data set 21_9 is confirmed, because
 - » the resulting optimal parameter sets for 21_9 in the fourth experiment cannot reproduce the data in the other experiments
 - » the 21_9 can not be reproduced by the other optimal parameter sets
- » Additional validation error only about 1.41 (MitSim) up to 2.43 (Aerde) percentage points. For OVM even only 0.41 additional percentage points.
- » -> Validation with data of the same drivers seems to produce much better results. But difficult concerning transferability.



Hokkaido, Japan - Conclusions

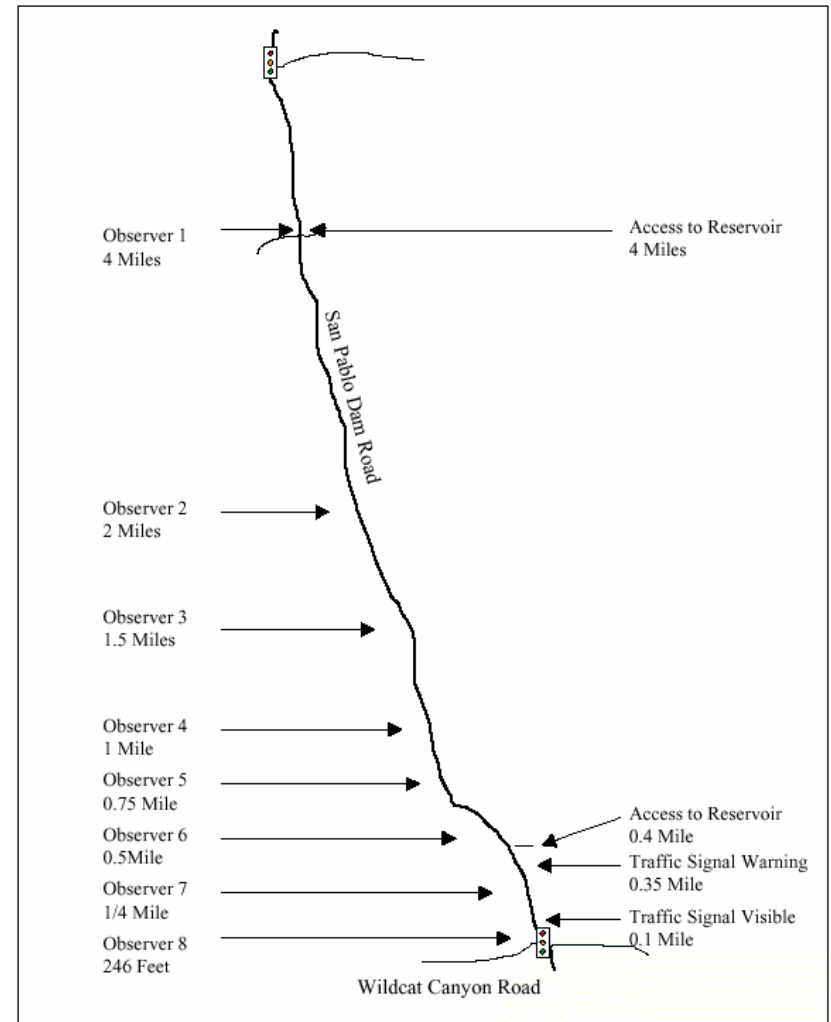
Main results analysing gaps:

1. Calibration error: 12-17 % (total range: 9-24 %).
2. Validation error: 17-22 % (about 3.2-5.5 % in addition to calibration error).
3. Validation error (special driver-pair): 1.5-2.5 % in addition to calibration error. (Perhaps a lower limit for validation errors?)
4. Average errors of the models: 15.1-16.2 %. -> No model can be denoted to be the best.
5. Using only a few data for calibration and validation may cause "overfitting" (here: up to 60%), thus other data sets can not be reproduced well.



Modelltestbeispiel San Pablo Dam Road (California, USA)

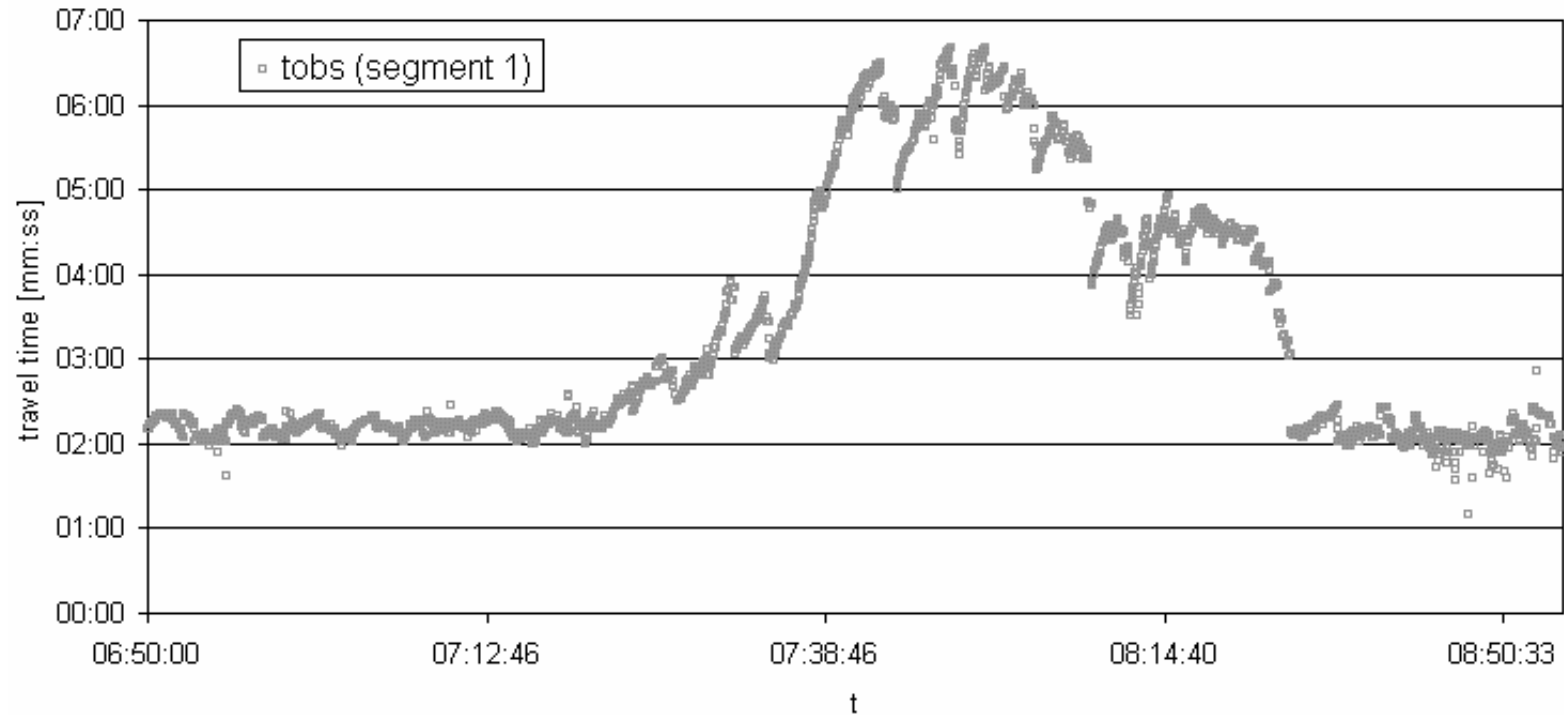
- ▶ Erhoben durch Daganzo et. al., Universität Berkeley
- ▶ Straßenlänge ca. 4 Meilen (6,1 km)
- ▶ Einspurige Landstraße
- ▶ Tempolimit 50 mph (80 km/h)
- ▶ Ampel am Ende der Straße verursacht Stauungen
- ▶ Zwei Tage in 1997, jeweils 7 - 9 Uhr.
- ▶ Passierzeiten von 2300 Fahrzeugen an acht Observerpositionen (nur Richtung Wildcat Canyon Road)
- ▶ Fahrzeugtypen wurden registriert, aber keine Geschwindigkeitsinformation





San Pablo Dam Road - Die Daten

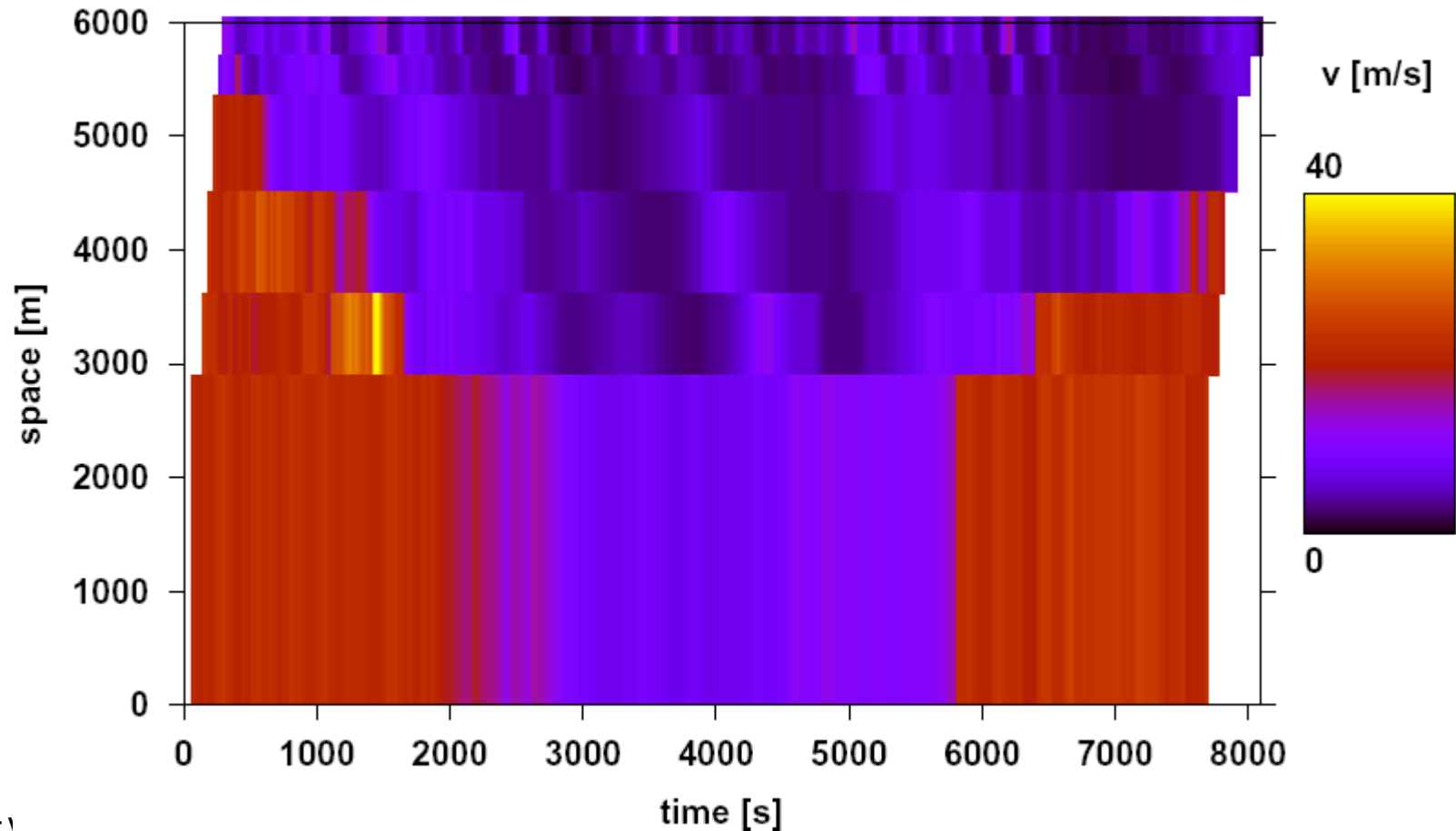
Entwicklung der Reisezeiten auf dem ersten Segment (Observer 1 - 2), Tag 1





San Pablo Dam Road - Die Daten

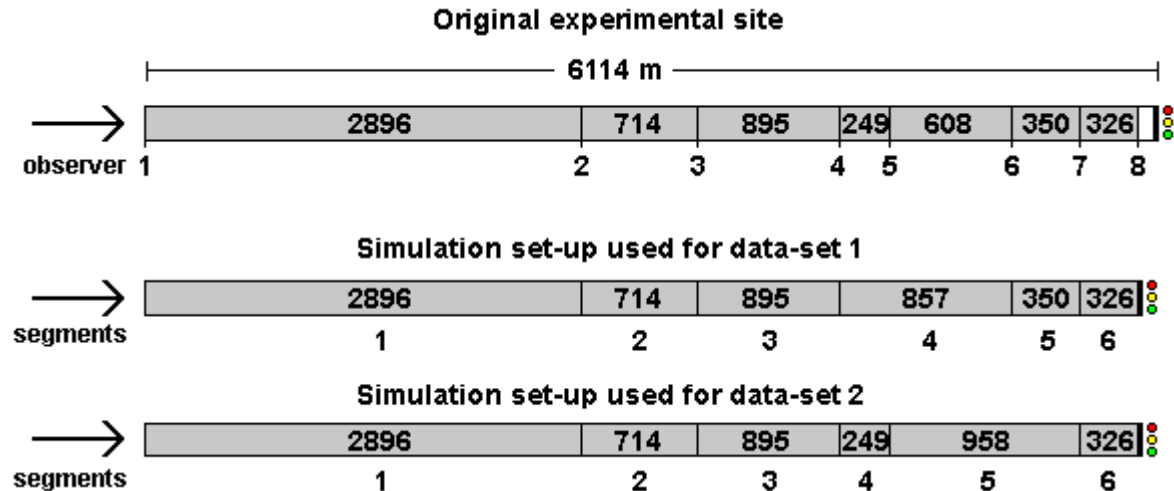
Entwicklung der Geschwindigkeiten im
Untersuchungsausschnitt, Tag 1





San Pablo Dam Road - Modellierung

Realisierung der Simulation:



- ▶ **Segmentierung** der Straße entsprechend der Observerpositionen (Daten von Observer 5 an Tag 1 und von Observer 6 an Tag 2 ignoriert aufgrund unvollständiger Daten).
- ▶ **Zufluß:** Entsprechend der am ersten Observer registrierten Zeiten werden die Fahrzeuge mit Freiflussgeschwindigkeit auf die Straße gelassen.
- ▶ **Abfluß:** Keine Information über die Ampelschaltung -> Implementierung einer "virtuellen" Ampel, welche die Fahrzeuge ungefähr zu den am letzten Observer registrierten Zeitpunkten aus dem System fahren lässt.



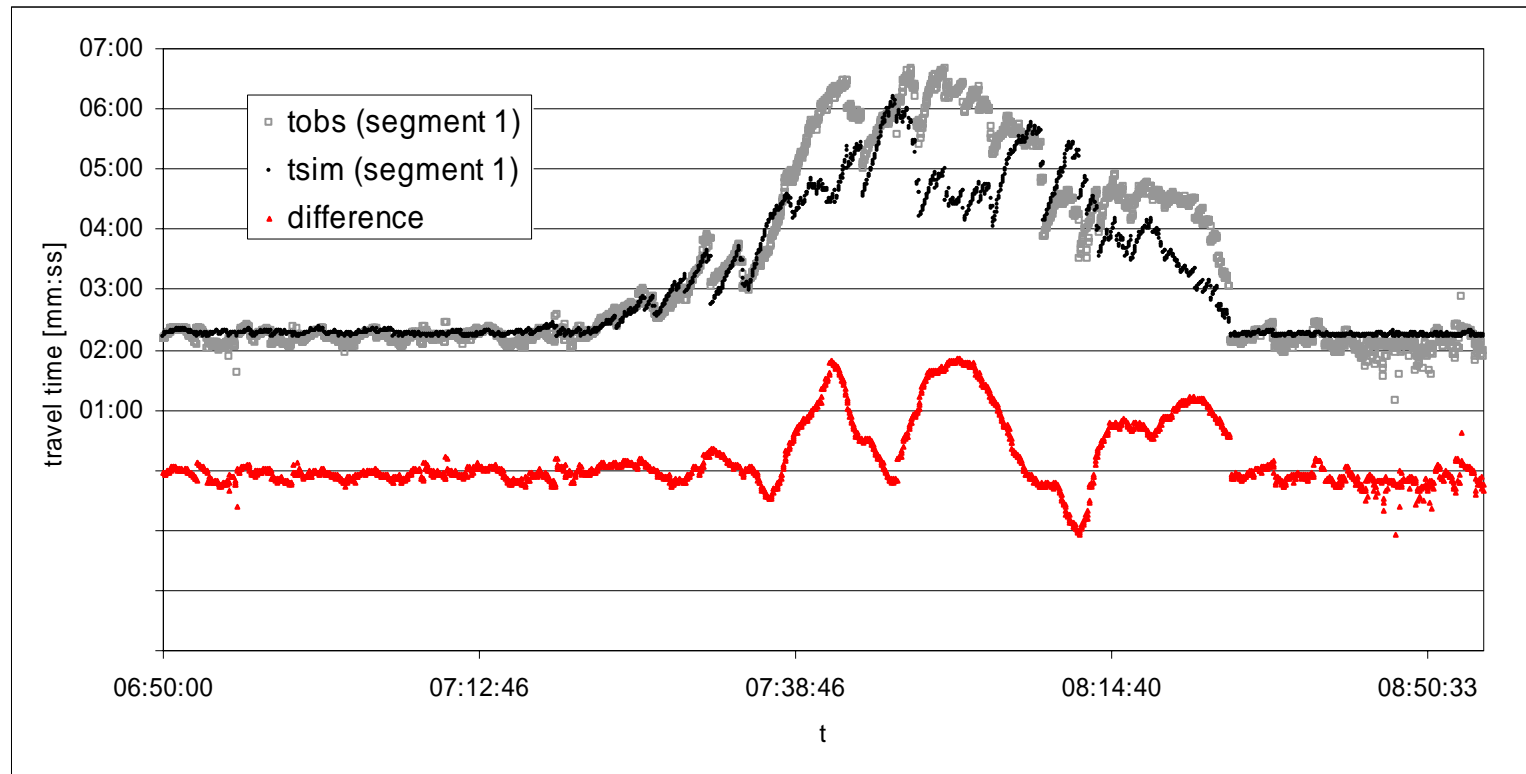
Testing the models ...

- ▶ Bis jetzt 9 Modelle getestet: CA, VDR, VDR++, CA SYNC, FRITZSCHE, GIPPS, IDM, OVM, SK.
- ▶ Jedes Modell ist durch Differenzgleichungen bzw. Regeln für die Fahrzeugbewegung definiert und hat einen spezifischen Satz von Parametern (Maximalgeschw., Beschleunigung, Bremsbeschleunigung, Zufälliges Bremsen, ...).
- ▶ Ziel ist es, für jedes Modell die "besten" Parametersätze zu finden.



San Pablo Dam Road - Reisezeitvergleich

Reisezeitdifferenzen auf dem ersten Abschnitt (Observer 1 - 2)





Fehlermaße

Vergleich der gemessenen mit den simulierten **Reisezeiten** auf jedem Abschnitt zwischen den Observern.

Absolute Abweichung der Reisezeiten auf jedem Abschnitt s

$$f_{mean}^s = \frac{1}{N} \sum_{i=1}^N |tobs_i^s - tsim_i^s|$$

Abweichung der Reisezeiten **relativ** zur durchschnittlichen Reisezeit auf jedem Abschnitt

$$f_{percent}^s = \frac{f_{mean}^s}{tobs_{mean}^s} * 100 \quad , \text{ with } tobs_{mean}^s = \frac{1}{N} \sum_{i=1}^N tobs_i^s$$

Relative Abweichung über den gesamten Untersuchungsausschnitt ("overall error")

$$f_{percent} = \frac{1}{S} \sum_{s=1}^S f_{percent}^s$$

Dann Finden der optimalen Parameterkonstellationen mit "shotgun-approach", strukturierter Analyse oder Optimierungsverfahren wie z.B. "Nelder Mead" ...



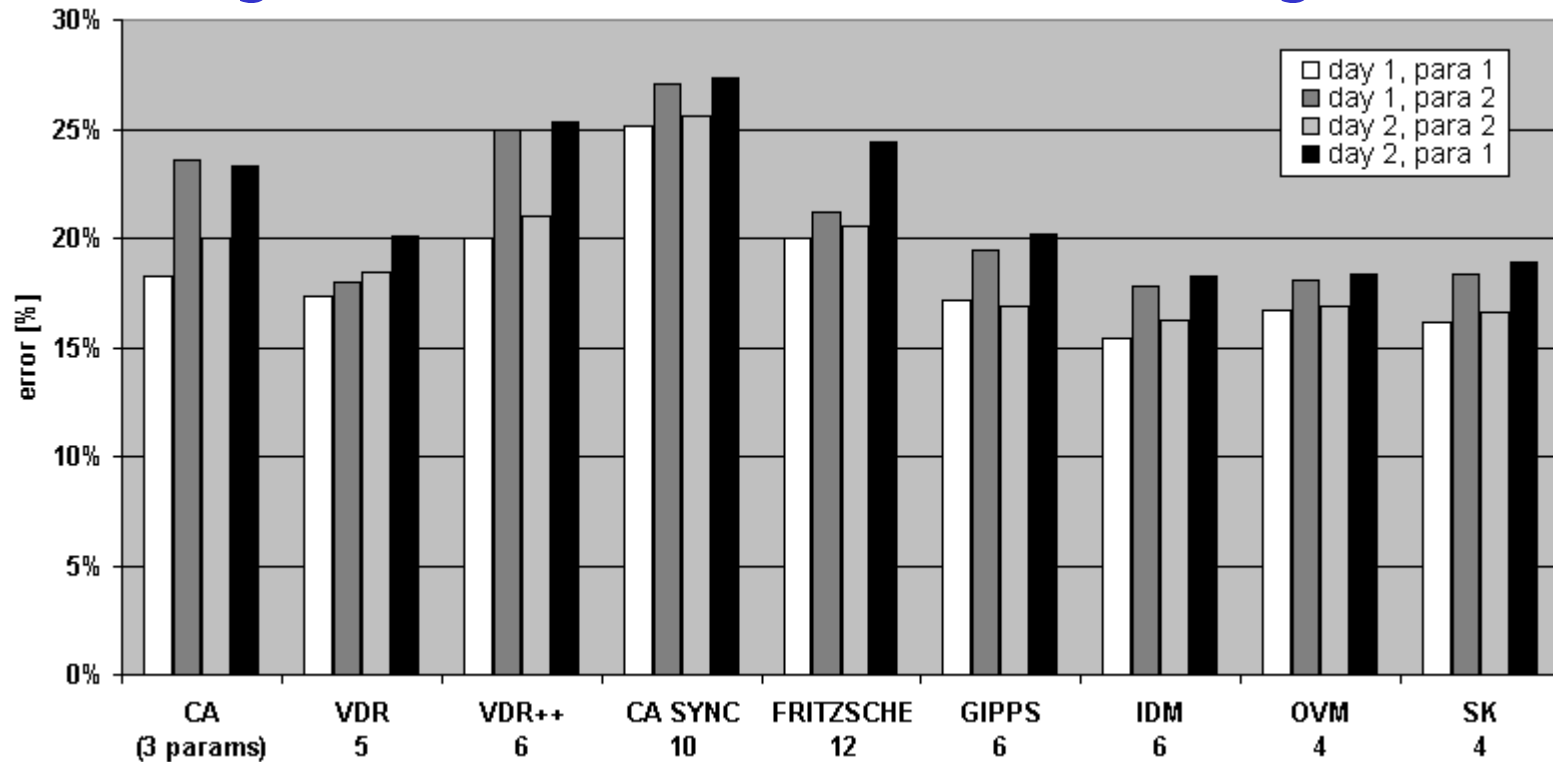
Analyse

Durchgeführte Analysen:

1. Kalibrierung jedes Modells mit beiden Datensätzen einzeln, um die jeweils beste Parameterkonstellation zu erhalten. Zur Validierung dann Einsetzen der Parameter "über kreuz", um Robustheit zu prüfen.
2. Vergleich der Ergebnisse bei Optimierung über alle Abschnitte zusammen im Gegensatz zur getrennten Optimierung jedes einzelnen Abschnittes (nur für SK-Modell!).



Ergebnisse (1/3): relative Abweichungen

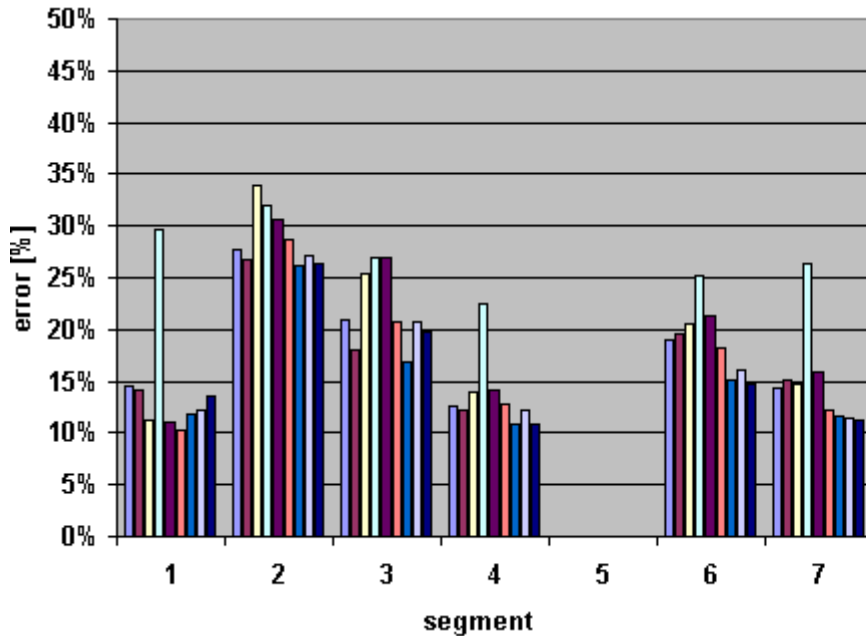


- ▶ Die besten Modelle produzieren ca. 16 % Kalibrierungsfehler.
- ▶ Modelle mit mehr Parametern schneiden nicht besser ab als einfache Modelle
- ▶ Validierung ("Über-Kreuz-Einsetzen") der besten Parametersätze verursacht 2-5% Fehler.

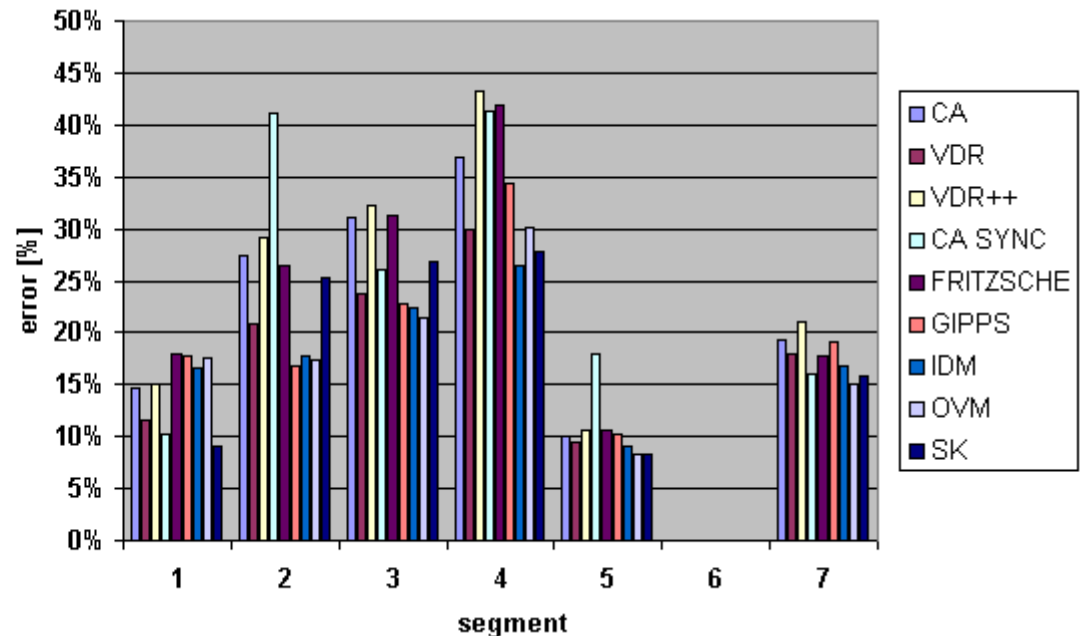


Ergebnisse (2/3): Relative Abweichungen pro Abschnitt

Tag 1



Tag 2



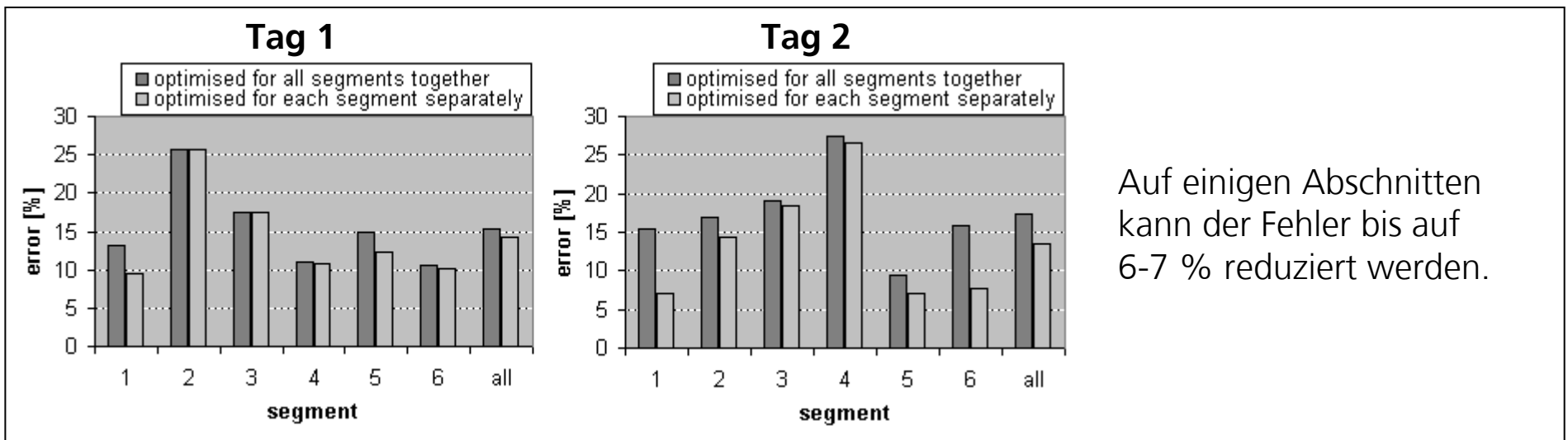
- ▶ Sehr unterschiedliche Fehlerraten auf den einzelnen Abschnitten
- ▶ Abschnitt 1 am besten mit zumeist 8-17% Fehlerrate.



Ergebnisse (3/3): Optimierung der einzelnen Abschnitte

SK-Modell

Vergleich der relativen Abweichungen auf den einzelnen Abschnitten



Auf einigen Abschnitten kann der Fehler bis auf 6-7 % reduziert werden.



San Pablo Dam Road - Zusammenfassung

Wesentliche Ergebnisse bei der Analyse von Reisezeiten:

1. Kalibrierungsfehler 15-20 %. (range: 15-25 % ; auf kleinräumigen Abschnitten 6-10 % möglich)
2. Validierungsfehler 17-27 % (bei jedem Modell 2-5 % mehr als Kalibrierungsfehler).
3. Modelle mit sehr vielen Parametern produzieren genauso große Fehler wie wesentlich einfachere Modelle. -> "Take the simplest one!"



Overall Conclusions

	Hokkaido Test Track (gaps)	San Pablo Dam Road (travel times)
Kalibrierungsfehler	12-17 % (range 9-24 %)	15-20 % (range 15-25 %)
Validierungsfehler	17-22 % (+ 3-5.5 %)	17-27 % (+ 2-5 %)
Durchschnittlicher Modellfehler	15.1-16.2 %	15-20 %

- » Bei der beschriebenen Vorgehensweise liegt der Fehler heutiger Modelle auf mikroskopischer Ebene bei ca. 15-25 %.
- » Die untersuchten Modelle scheinen sich in den Ergebnissen nicht sehr stark zu unterscheiden.
- » Die Unterschiedlichkeit von Fahrerverhaltensweisen scheint wesentlich größer zu sein als die Unterschiedlichkeit der Modelle.
- » Kalibrierung und Validierung mit nur wenigen Datensätzen kann zu Parametersätzen führen, die nur schwer übertragbar sind ("overfitting").



Perspectives

Perspectives and future research

- ▶ Perform sensitivity analyses of the parameters (robustness of the models)
- ▶ Testing more traffic flow models
- ▶ Check the models against other data sets (own measurements at the DLR-IVF with specially equipped vehicles)
- ▶ Confirm the minimum calibration errors of about 15-20% and validation errors of additionally 3-5% by testing with more data sets.
- ▶ Developing of better models
- ▶ Finally development of a benchmark for microscopic traffic flow models.