CrossMark

ORIGINAL PAPER

Transferring urban traveling speed model fits across cities

Christian Heinze[1] · Maximilian Leodolter[2] · Hannes Koller[2] · Dietmar Bauer[1]

Abstract

Purpose Link traveling time models form the basis for route planning methods used in navigation devices as well as for logistic applications. These models are provided based on extensive real world data sets which are available to a differing degree in different cities as well as for different links within a given city. For smaller cities, where typically fewer data is available or less frequently measured links, it might be beneficial to transfer models from close-by cities or links from the same city with sufficient data basis. In this paper, the potential for transferring link traveling time model fits, that is, the estimated models, between cities and within a city is investigated. Methods that combine information typically contained in street maps with empirically derived features that are easily transferred are developed and tested with substantial real world data sets. This provides the basis for developing route planning methods in cities with insufficient real world data coverage to base accurate route traveling time predictions on.

Methods Link traveling time models are derived on the basis of an extensive floating taxi data set in Vienna, Austria. The models incorporate typical map information such as speed limits and functional road classification (*frc*). Estimation is performed using penalized least squares methods to control for overfitting. The expected accuracy for the model transfer is investigated both in terms of intracity transfer (from

modelled links to other links in the same city) and in terms of intercity transfer (from one city to another city). Data sets of different extent are used from the two Austrian cities of Vienna and Linz as well as for the French city of Lyon.

Results The models presented in this paper are demonstrated to lead to superior performance compared to the benchmark model of Leodolter et al. (2015). It is shown that transfer between cities in the same country (here using the Vienna model for Linz) may be beneficial in terms of prediction accuracy while the transfer between countries (here from Vienna to Lyon) decreases accuracy but not dramatically.

Conclusion These results demonstrate that the transfer of link traveling time models within a city or from one city to another city can provide acceptable prediction accuracy and thus can be used as the basis for navigation algorithms in case no good data basis is accessible for a city.

Keywords Traveling time estimation · Navigation · Routing · Floating car data

1 Introduction

Local traveling speed measurements and predictions provide the basis for many vehicle routing and traveling time prediction algorithms. The latter can be found in navigation devices used in private vehicles as well as in logistic applications. Predictions are especially relevant for congested networks [16]. While navigation devices increasingly use real time information, logistics applications typically used in the planning stage (pre-trip) rely on long term (in the sense of several hours or days ahead) route traveling time predictions. Such predictions may be derived from link traveling time predictions based on local (to the links in the network) traveling speed predictions.

✉ Christian Heinze
cheinze@wiwi.uni-bielefeld.de

[1] Bielefeld University, Bielefeld, Germany

[2] AIT Austrian Institute of Technology GmbH, Vienna, Austria

⚏ Springer

Traveling speed predictions have been derived from either local measurements, taken from stationary road side sensors (for a survey see e.g. [1]), or from floating sensors distributed to a fleet of vehicles (from a long list of contributions see e.g. [3, 4, 10]). There exists an extensive literature dealing with traveling time estimation; see the collection of survey papers in [2] or the literature reviews in [13] and [23] amongst others. All these approaches presuppose large databases in order to yield accurate predictions for all possible routes in a given city and for a given time of the day (see e.g. [11] and [18]).

Two main hurdles for providing such a large sample exist. First, only a subset of the road links can be equipped with road side sensors due to cost restrictions. Second, the data collecting fleet vehicles do not traverse all road links at all times. Especially those links located on the outskirts of a city are typically neglected. Consequently, many cities currently lack sufficiently large data sets required by the mentioned approaches. In order to fill the gaps in the data, one can implement an extensive measurement campaign to cover missing road links. An alternative option—a backup strategy—is to estimate free flow speeds based on static data. Moses and Mtoi [15] propose different models for estimating free flow speeds based on the speed limit, spacing between signalized intersections and vehicle type. Transportation Research Board [20] and [5] consider additional adjustment factors for structural road parameters and vehicle types. Tseng et al. [21] extend these methods to suburban highways. Graser et al. [8] and [9] suggest the usage of different road network centrality measures as predictors for link traveling speed. However, these authors acknowledge that some of the measures depend on the boundaries of the chosen map which introduces some arbitrariness to the approach.

Most of these basic approaches do not address the presence of daily variation of traveling speeds in sufficient detail, or even neglect it completely. Leodolter et al. [12] highlight the dramatic losses in prediction accuracy resulting from neglecting the daily variation.

In general, one should expect decreased traveling speeds in morning and evening rush hours. However, more precise information on the daily variation of local traveling speeds must be inferred from data. Nonetheless, it may be postulated that daily variation patterns are very similar for road links of similar usage (in a given city). Herein, road usage is operationalized as functional road categories. Moreover, these patterns may be similar for similar cities, which suggests the possibility of some form of transfer across cities. Leodolter et al. [12] illustrate this idea by transferring several models fitted on data from Vienna, Austria, to the nearby city Linz, Austria. In particular, this amounts to using the daily variation of one city as a proxy for the daily variation in a different city and is shown to improve the predictive power in

comparison to approaches not taking daily variation into account. Hence, this method couples the simplicity and low costs of the mentioned backup approaches with the accuracy gained from using local measurements.

In this paper, we refine and extend the method presented in [12] in several respects. Most notably, we provide a more careful modeling of the daily variation patterns by allowing these profiles to differ across road types. Earlier approaches neglect dependence between different time-of-the-day intervals and hence result in noisy daily profiles. This problem becomes more severe when increasing the model complexity by allowing daily profiles to change with road type (as we do). Here, we propose to use penalized least-squares methods, which are already popular in non-parametric (spline-based) estimation. Penalizing the roughness of the daily variation patterns lowers the estimation uncertainty and yields smooth estimated daily profiles. We include details on an efficient numerical implementation of our approach in the Appendix. In addition, we observe that the similarities of the daily profiles across cities concentrate in low dimensional subspaces. We enhance the cross-city prediction approach of [12] to exploit this finding. Using these new developments, we comprehensively evaluate the effectiveness of a wide range of predictive models with respect to on-site performance (predictive power for the city for which data is available) and cross-city transfer. For the on-site comparisons we use data from the two Austrian cities Vienna and Linz as well as the French city Lyon. The cross-city performance is evaluated for the transfer from Vienna to Linz, Austria, and from Vienna to Lyon, France. The choice of cities here is due to data availability; all methods can be applied without change to other cities/countries as long as adequate data is available.

The remainder of this paper is organized as follows: Section 2 provides background information on the data set used for accessing the predictive power in subsequent sections. Subsequently, Section 3 presents the models and estimation techniques as well as the formulas used for prediction. Section 4 details on the evaluation methodology. Section 5 discusses the evaluation results. Section 6 concludes the paper. The Appendix provides some details for practical implementation of our methods.

2 Data & descriptive analysis

The floating car data (*FCD*) used in our study is similar to that in [12]. More specifically, the *FCD* is collected by about 3500 registered taxis in the region of Vienna (Austria), about 300 taxis in the region of Linz (Austria), and about 400 taxis in the region of Lyon (France).

The *FCD* raw data consists of anonymized vehicle trajectories (time and GPS position) with a variable sampling
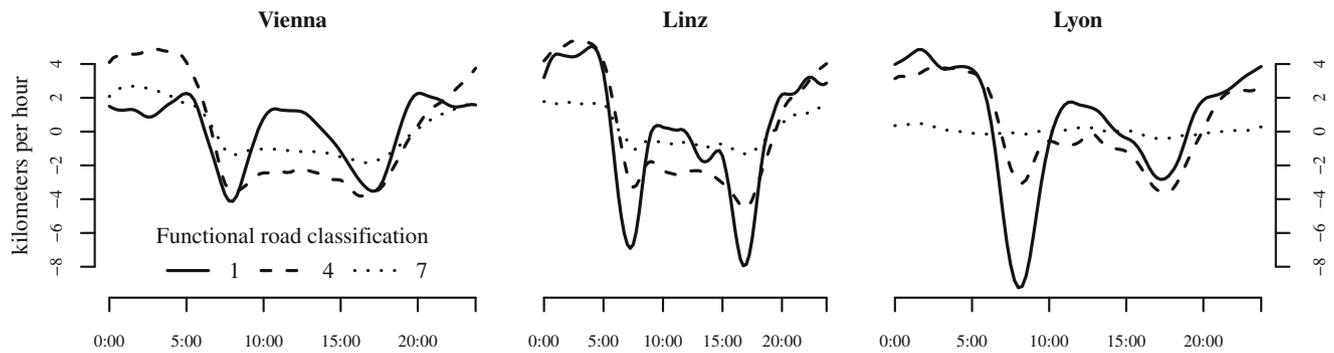
Fig. 1 The figure shows the (estimated) daily variation of prediction errors from a regression model for link traveling speed sp in km/h with road type (*frc*) specific intercepts and speed limit slope coefficients for three road types (1, 4, 7) and three cities (Vienna, Linz, and Lyon)

interval (between 10 and 60 seconds). In the *FCD* processing, the vehicle trajectories are first map-matched to an OpenStreetMap road network graph of the region. During map-matching, the most probable road link for every GPS position is identified (by taking into account the great-circle distance between GPS position and road geometry as well as the estimated heading of the vehicle). Positions without plausible map-matchings are discarded from the trajectory. Then, the covered road distance between two consecutive map-matched positions is determined by a shortest-path routing on the road network. The quotient of the covered distance and the time elapsed between two GPS measurements provides one speed observation. The latter is assigned to the respective links by linear interpolation. Finally, trajectories corresponding to implausible speed observations of more than 110% of the link speed limit are discarded.

Each of the three city-specific data sets covers a period of one year with about 3 million speed measurements per day in Vienna, about 700,000 measurements per day in Linz, and about 800,000 measurements per day in Lyon. Data from irregular days (public holidays, weekends, school holiday, etc.) is discarded as well as observations concerning links with a speed limit below 20 km/h. In particular, our analysis focuses exclusively on weekdays during the school period. Arguably, this time features the most severe congestion delays and is therefore most critical for traveling speed prediction.

Time is expressed in the form of 96 intervals of 15 minutes covering the 24 hours of a day. Speed measurements pertaining to a given road link, day, and time interval are averaged (using harmonic averages) to form a single observation. This leads to roughly 2.5 million (M) observations for Viennese road links, roughly 1.1 M. observations for road links in Linz, and roughly 1.25 M. observations for road links in Lyon. The associated speed limits and road classification information are taken from OpenStreetMap. Finally, every road link is assigned a functional road classification (*frc*) number on the basis of its OpenStreetMap highway tag. These numbers indicate the road type and

range from one (motorways) to eight (living streets). As a rough guide, the importance of road links decreases as the *frc* number increases.

For preliminary descriptive analysis, the regression model

$$\mathtt{sp}_i = c_{f(i)} + \beta_{f(i)}\mathtt{mxsp}_i + v_i$$

is fitted by least-squares to the first half of each city-specific data set. Herein, $\mathtt{sp}_i$, $f(i)$, and $\mathtt{mxsp}_i$ denote the traveling speed (average), road type (*frc*), and speed limit of the i-th observation, respectively. In particular, intercept and slope coefficients are allowed to change across road types (*frc*). Next, prediction errors $\hat{v}_i$ are calculated for each city based on the second half of the respective data set. Figure 1 shows the averages[1] of the prediction errors over each of the 96 time intervals and for *frc* numbers 1, 4, and 7. The shown daily variation patterns reflect the *frc* classifications: highways (*frc* = 1) exhibit strong signs of congestion during morning and evening peaks, which fade during midday; medium size roads (*frc* = 4) show a clear night/day divide, but of lesser extent than highways; finally, living streets (*frc* = 7) show only little daily variation.

Figure 1 also reveals considerable differences in the daily variation patterns across cities, which raises concerns about the approach of [12] of using the same pattern for all cities. Figure 2 further investigates this issue. To this end, we arrange the prediction error averages in matrices $\hat{\boldsymbol{\Gamma}}_q = [\hat{\gamma}_{t,j|q}] \in \mathbb{R}^{96 \times 8}$, wherein $\hat{\gamma}_{t,j|q}$ denotes the average pertaining to time interval t, road type j (*frc*), and city $q \in \{$Vienna, Linz, Lyon$\}$. Hence, the j-th column of $\hat{\boldsymbol{\Gamma}}_q$ embodies the estimated daily variation for the j-th road type (*frc*) in city q. Insufficient data on *frc* class 2 for Linz leads to several missing entries in the second column of $\hat{\boldsymbol{\Gamma}}_{\mathrm{Linz}}$. Therefore we nullify the second column of $\hat{\boldsymbol{\Gamma}}_q$ for all three cities to ensure comparability. Panels (a)–(d) of Fig. 2 show the first four left singular vectors $\hat{\mathbf{u}}_{1|q}, \ldots, \hat{\mathbf{u}}_{4|q}$

---

[1]These averages are calculated by regularized least-squares to obtain a smooth daily variation estimate; see Section 3 for details.
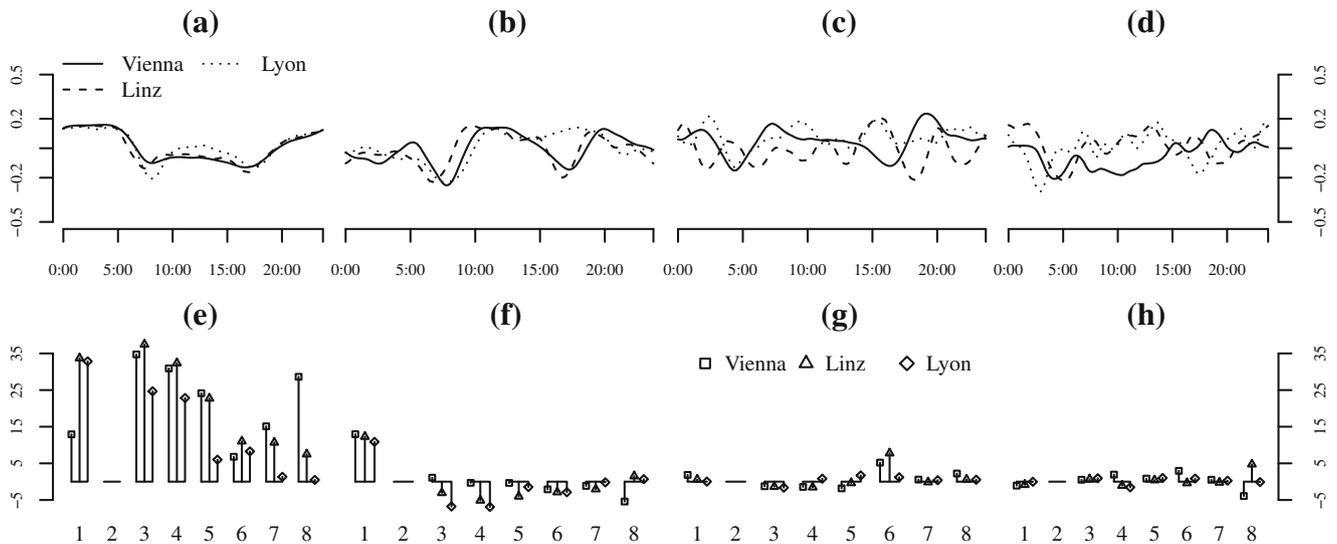
Fig. 2 Panels (**a**)–(**d**) show the first four left singular vectors $\hat{\mathbf{u}}_{1|q}, \ldots \hat{\mathbf{u}}_{4|q}$ of the estimated 96×8 daily variation matrices $\hat{\mathbf{\Gamma}}_q$ for three cities $q \in \{\text{Vienna, Linz, Lyon}\}$. Panels (**e**)–(**h**) show the corresponding loading vectors $\hat{\sigma}_{1|q}\hat{\mathbf{v}}_{1|q}, \ldots, \hat{\sigma}_{4|q}\hat{\mathbf{v}}_{4|q}$ calculated by multiplication of the respective singular value $\hat{\sigma}_{i|q}$ and right singular vector $\hat{\mathbf{v}}_{i|q}$. Roads of type (*frc*) 2 are excluded from the estimation due to lack of corresponding data. The zero loadings for road type 2 in panels (**e**)–(**h**) reflect this omission

of the 96×8 matrices[2] $\hat{\mathbf{\Gamma}}_q = \sum_{i \leq 7} \hat{\sigma}_{i|q}\hat{\mathbf{u}}_{i|q}\hat{\mathbf{v}}_{i|q}^{\mathsf{T}}$, wherein $\hat{\sigma}_{i|q}$ and $\hat{\mathbf{v}}_{i|q}$ denote the i-th singular value and i-th right singular vector of $\hat{\mathbf{\Gamma}}_q$, respectively.

We observe that the first two left singular vectors shown in panel (a) and (b) are quite similar across cities, but one distinct feature of Lyon stands out. These two signals mostly represent the decline of average traveling speed during the day in panel (a) as well as the morning and evening peak in panel (b), respectively. In particular, the second left singular vector for Lyon reflects the considerable difference in magnitude between the morning and evening peak shown in Fig. 1. In contrast, the two peaks are quite similar for Vienna and Linz, which manifests in the divergence of the three left singular vectors in panel (b) during this time. Panel (c) and (d) reveal considerable differences between the left singular vectors $\hat{\mathbf{u}}_{3|q}, \hat{\mathbf{u}}_{4|q}$ across cities; $\hat{\mathbf{u}}_{5|q}, \ldots, \hat{\mathbf{u}}_{7|q}$—not shown—exhibit qualitatively comparable differences across cities q as $\hat{\mathbf{u}}_{3|q}$ and $\hat{\mathbf{u}}_{4|q}$. Finally, it should be kept in mind that Figs. 1 and 2 show estimates, which are subject to sampling uncertainty.

Panel (e)–(h) of Fig. 2 show the coefficient vectors $\hat{\sigma}_{i|q}\hat{\mathbf{v}}_{i|q}$ corresponding to the first four left singular vectors $\hat{\mathbf{u}}_{i|q}$. More specifically, each column of $\hat{\mathbf{\Gamma}}_q$—representing the daily pattern for one road category in form of 96 estimates $\hat{\gamma}_{t,j|q}$, $t \leq 96$—equals a linear combination of the left singular vectors $\hat{\mathbf{u}}_{1|q}, \ldots, \hat{\mathbf{u}}_{7|q}$ (panel (a)–(d)).

The latter can therefore be interpreted as basic daily patterns. The coefficients $\hat{\sigma}_{i|q}\hat{v}_{j,i|q}$—called loadings herein—corresponding to the i-th basic pattern $\hat{\mathbf{u}}_{i|q}$ and the eight road types $j = 1, \ldots, 8$ gather in the vector $\hat{\sigma}_{i|q}\hat{\mathbf{v}}_{i|q}$. The zero loadings $\hat{\sigma}_{i|q}\hat{v}_{2,i|q}$ for road type (*frc*) 2 with respect to all basic patterns $i = 1, \ldots, 7$ reflect its omission in the estimation. The other loadings express how the respective basic pattern enters the daily variation of the corresponding road type. In case of a positive loading $\hat{\sigma}_{i|q}\hat{v}_{j,i|q}$, the basic pattern i enters in the form shown in panels (a)–(d) of Fig. 2. A negative loading implies that the basic pattern is turned upside down. The absolute value of the each loading governs the strength of the respective basic pattern in the daily variation of the corresponding road type. In this regard, we observe two notable features. Firstly, the signs of $\hat{\sigma}_{1|q}\hat{v}_{j,1|q}$ and $\hat{\sigma}_{2|q}\hat{v}_{j,2|q}$ are identical across cities. Thus, the first two basic patterns enter in the same form across cities q, which affirms the above interpretation of $\hat{\mathbf{u}}_{1|q}$ and $\hat{\mathbf{u}}_{2|q}$. However, the magnitude of the loadings differs considerably for some road types, that is, the basic patterns occur with different strength across cities. Secondly, the loadings decrease rapidly as i increases, however, are still of considerable size for some road types and $i > 2$; see panel (g) and (h).

Finally, we observe that the ratio $\|\hat{\mathbf{\Gamma}}_q - \tilde{\mathbf{\Gamma}}_{\text{Vienna}}\|_{\text{F}}/\|\hat{\mathbf{\Gamma}}_q - \hat{\mathbf{\Gamma}}_{\text{Vienna}}\|_{\text{F}}$, wherein $\tilde{\mathbf{\Gamma}}_{\text{Vienna}} = \sum_{i \leq 2} \hat{\sigma}_{i|\text{Vienna}}\hat{\mathbf{u}}_{i|\text{Vienna}}\hat{\mathbf{v}}_{i|\text{Vienna}}^{\mathsf{T}}$ and $\|A\|_{\text{F}}^2 = \sum_{i,j} a_{i,j}^2$ for any matrix $A = [a_{i,j}]$, equals 82.1% for $q = \text{Linz}$ and 98.4% for $q = \text{Lyon}$. Hence, the rank two approximation $\tilde{\mathbf{\Gamma}}_{\text{Vienna}}$ to the Viennese daily variation matrix $\hat{\mathbf{\Gamma}}_{\text{Vienna}}$ is a closer substitute for $\hat{\mathbf{\Gamma}}_{\text{Linz}}$ and $\hat{\mathbf{\Gamma}}_{\text{Lyon}}$ than the full estimate $\hat{\mathbf{\Gamma}}_{\text{Vienna}}$. This finding

---

[2]Nullification of the second column induces a zero singular value; thus, summation of seven components suffices.

motivates the alternative prediction strategy proposed in the following Section 3 and evaluated in Sections 4 and 5.

3 Models, estimation & prediction

This paper considers predictions of the traveling speed $\mathtt{sp} = y$ (in km/h) and predictions of the ratio $\mathtt{sp}/\mathtt{mxsp} = y'$ of traveling speed to speed limit (in km/h) for a given road type (frc) $f \in \{1, \ldots, 8\}$ and time $t \in \{1, \ldots, 96\}$. We start with two general formulations

$$y_i = \mathtt{sp}_i = c_{f(i)} + \gamma_{t(i),f(i)} + \beta_{f(i)}\mathtt{mxsp}_i + u_i \quad \text{and} \quad (1a)$$

$$y_i' = \frac{\mathtt{sp}_i}{\mathtt{mxsp}_i} = c'_{f(i)} + \gamma'_{t(i),f(i)} + \beta'_{f(i)}\mathtt{mxsp}_i + u_i' , \quad (1b)$$

wherein $t(i)$ denotes the time-of-the-day interval of observation i. The remainder terms u_i and u_i' are assumed to be zero mean. The superscript $'$ in Eq. 1b acknowledges the possibility of differences in parameter values between Eqs. 1a and 1b. The subsequent discussion is in terms of the former to circumvent superfluous replications. We tested the assumption of linearity in $\mathtt{mxsp}$ and found it to be appropriate; see the comment at the end of Section 5.

These general models lead to a number of variations by imposing different restrictions of the frc-specific intercept and slope coefficients c_f and β_f as well as the daily variations $\gamma_{t,f}$. For the slope coefficient β_f, we mostly focus on the unrestricted case. Section 5 also comments on the choice of a common slope coefficient $\beta_f = \beta$, $f = 1, ..., 8$. For the daily variation coefficients $\gamma_{t,f}$ we allow several (pre-specified) frc-based groups $F_0, F_1, ..., F_g$ (partitioning the set $\{1, ..., 8\}$) of identical daily variation. More specifically, we use

$$\gamma_{t,i} = \bar{\gamma}_{t,j} \text{ for all } i \in F_j \text{ with } \sum_{t=1}^{96} \bar{\gamma}_{t,j} = 0, \ j = 1, ..., g, \quad (2)$$

and $\bar{\gamma}_{t,0} = 0$. Notable special cases of Eq. 2 are the case of no daily variation ($F_0 = \{1, ..., 8\}$), a single daily variation pattern ($g = 1$, $F_0 = \emptyset$, $F_1 = \{1, \ldots, 8\}$), and the unrestricted case ($g = 8$, $F_0 = \emptyset$, $F_h = \{h\}$, $1 \leq h \leq 8$). In the latter case, no restrictions—except $\sum_{t \leq 96} \gamma_{t,f} = 0$—are imposed on $\gamma_{t,f}$. The parameters can be conveniently collected into the matrix $\mathbf{\Gamma} = [\gamma_{t,f}]$.

Additional constraints on $\mathbf{\Gamma}$ allow further reduction in model complexity, e.g., $\gamma_{t,f}$ being identical at night time, in order to optimize the variance-bias trade-off; we further comment on this in Section 5.

The models are fitted using regularized least-squares. The regularization aims at smoothing the daily variation by adding a penalty term $\lambda\|\mathbf{\Delta}_2\mathbf{\Gamma}\|_F^2$ to the least-squares objective. Here $\lambda(\geq 0)$ is the regularization constant, $\|\cdot\|_F$ denotes the Frobenius norm (square root of the sum of

the squares of the matrix entries), and $\mathbf{\Delta}_2$ the symmetric circulant matrix

$$\mathbf{\Delta}_2 = \begin{pmatrix} 2 & -1 & \ldots & \ldots & 0 & -1 \\ -1 & 2 & -1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -1 & 0 & 0 & \ldots & -1 & 2 \end{pmatrix} .$$

This penalty encourages a smooth daily variation estimate; see [6, sec. 4.2] amongst others. The regularization constant λ is either set to zero (no regularization) or chosen by the GCV-criterion [7]. The Appendix further details on an implementation strategy that easily adapts to large-scale applications.

The estimation is carried out for all three cities—Vienna, Linz, and Lyon, separately; an additional subscript on the parameter estimates indicates the respective data source, e.g. $\hat{\gamma}_{t,j|\text{Vienna}}$, in case the distinction is important.

Predictions $\hat{y}$ either take the form

$$\hat{y}_i(p \,|\, q) = \hat{c}_{f(i)|q} + \hat{\gamma}_{t(i),f(i)|q} + \hat{\beta}_{f(i)|q}\mathtt{mxsp}_i \quad \text{or} \quad (3a)$$

$$\hat{y}_i(p \,|\, q) = \hat{c}_{f(i)|q} + \hat{\gamma}^r_{t(i),f(i)|q} + \hat{\beta}_{f(i)|q}\mathtt{mxsp}_i , \quad (3b)$$

wherein $p, q \in \{\text{Vienna, Linz, Lyon}\}$ indicate the target city for prediction (p) and the data source for estimation (q). A hat superscript identifies least-squares estimates. Moreover, the daily variation coefficient estimate $\hat{\gamma}^r_{t,f}$ in Eq. 3b amounts to the t, f-th entry of the rank r approximation

$$\tilde{\mathbf{\Gamma}}^r_q = \sum_{i \leq r} \hat{\sigma}_{i|q}\hat{\mathbf{u}}_{i|q}\hat{\mathbf{v}}^\mathsf{T}_{i|q}$$

of $\hat{\mathbf{\Gamma}}_q$. The latter is calculated based on the singular value decomposition $\hat{\mathbf{\Gamma}}_q = \sum_{i \leq 8} \hat{\sigma}_{i|q}\hat{\mathbf{u}}_{i|q}\hat{\mathbf{v}}^\mathsf{T}_{i|q}$ of the estimate $\hat{\mathbf{\Gamma}}_q$. Herein $r \leq 8$ is pre-specified. The reduction implemented by choosing $r < 8$ acknowledges the above observation of similarities across cities between the terms $\hat{\sigma}_i\hat{\mathbf{u}}_i\hat{\mathbf{v}}^\mathsf{T}_i$ for $i \leq 2$ and dissimilarities for $i > 2$.

4 Evaluation methodology

The various models are assessed by cross validation. Therein parameter estimation (training) is carried out on a data set from $q \in \{\text{Vienna, Linz, Lyon}\}$ of size $n_e = 500,000$. Prediction errors (evaluation) are calculated using a separate data set from $p \in \{\text{Vienna, Linz, Lyon}\}$ of size $n_v = 500,000$. That is, the cases $p = q$ and $p \neq q$ refer to on-site prediction and cross-city prediction, respectively. Training and evaluation data sets are randomly drawn from the respective full data sets in such a way that the two data sets are non-overlapping and each of the two has the same share of observations from every time interval as the

full data set. This stratification ensures that all time intervals receive appropriate attention. Moreover, the lack of overlap ensures that on-site comparisons rely on out-of-sample predictions, too.

We express the predictive performance in terms of the (estimated) mean absolute percentage error (mape) calculated as

$$\widehat{\text{mape}}(p \mid q) = \frac{1}{n_v} \sum_{i=1}^{n_v} \frac{|y_i - \hat{y}_i(p \mid q)|}{|y_i|} \quad \text{and} \tag{4a}$$

$$\widehat{\text{mape}'}(p \mid q) = \frac{1}{n_v} \sum_{i=1}^{n_v} \frac{|y_i' - \hat{y}_i'(p \mid q)| \, |\text{mxsp}_i|}{|y_i|} , \tag{4b}$$

respectively. Here, $y_i = \text{sp}_i$ and $y_i' = \text{sp}_i/\text{mxsp}_i$ denote the i-th response observation in an evaluation data set of size $n_v = 500,000$ from $p \in \{\text{Vienna, Linz, Lyon}\}$. In contrast to the root mean squared error used in [12], the criteria in Eqs. 4a and 4b acknowledge that prediction errors $y - \hat{y}(p \mid q)$ and $(y' - \hat{y}'(p \mid q))\text{mxsp}$, respectively, occur at different speed levels.

The predictions $\hat{y}(p \mid q)$ and $\hat{y}'(p \mid q)$ are based on either a special case of Eqs. 1a and 1b, respectively, or one of the additional benchmark strategies outlined below. In the former case, the calculation of predictions proceeds as in Eqs. 3a or 3b.

We repeat the calculation of Eqs. 4a and 4b for $M = 50$ different, but necessarily overlapping, pairs of a disjoint training sample and an evaluation sample and for $(p \mid q)$ equal to (Vienna | Vienna), (Linz | Linz), (Lyon | Lyon) (on-site) and (Linz | Vienna), (Lyon | Vienna) (cross-city). The replications reduce the effects of randomly drawing the estimation and prediction subsamples and generate information on the accuracy of the estimated performance measures. Table 1 reports averages and empirical standard deviations[3] (in parenthesis) of the mape estimates over the $M = 50$ replications and for the various prediction strategies. We considered several other variants but restrict the presentation to three benchmark procedures and some specializations of Eqs. 1a and 1b chosen to reflect the key lessons of our study. Section 5 comments on some extensions.

The three benchmark predictors for y comprise

a) the speed limit $\hat{y}_i = \text{mxsp}_i$,

b) the scaled speed limit $\hat{y}_i = \hat{\beta}\text{mxsp}_i$, wherein $\hat{\beta}$ symbolizes a least-squares estimate, and

c) a linear model prediction $\hat{y}_i = \hat{c} + \hat{\gamma}_{t(i),f(i)} + \hat{\beta}_{f(i)}\text{mxsp}_i$, wherein $t(i)$, and $f(i)$ once more indicate the respective time interval and the road type (frc). The least-squares estimates $\hat{c}$, $\hat{\gamma}_{t,f}$, and $\hat{\beta}_f$ represent an intercept, the daily variation, and the influence of the speed limit, respectively. The mxsp slope coefficient estimates $\hat{\beta}_f$ are only allowed to differ across the road type (frc) groups $\{1, 2\}$, $\{3\}$, $\{4, 5, 6\}$, and $\{7, 8\}$. The daily variation coefficients $\hat{\gamma}_{t,f} = \hat{\bar{\gamma}}_{t,1}$ are identical across all road types, which amounts to ($g = 1$, $F_0 = \emptyset$, $F_1 = \{1, \ldots, 8\}$). In addition, the estimation of the daily variation coefficients $\bar{\gamma}_{t,1}$ enforces identical coefficient estimates for (night) time intervals between 23:00 and 5:30.

The first two are used as benchmarks in [12]; the third method is the specification advocated therein. The second benchmark b) may be understood as a constant prediction of $y_i' = \text{sp}_i/\text{mxsp}_i$. Its acceptable performance—shown in Table 1—motivated the consideration of the refined ratio model in Eq. 1b.

In addition, we include several variants of the models Eqs. 1a and 1b. All of these specializations allow differences in the effect of the speed limit mxsp across road types (frc) via unrestricted[4] slope coefficients β_f; Section 5 comments on the restriction $\beta_f = \beta$, $f = 1, ..., 8$. More specifically, we consider non-regularized fitting ($\lambda = 0$) of

d) Eq. 1a with no daily variation ($g = 0$, $F_0 = \{1, .., 8\}$),

e) Eq. 1a with a single daily variation ($g = 1$, $F_0 = \emptyset$, $F_1 = \{1, \ldots, 8\}$),

f) Eq. 1a with three groups of identical daily variation ($g = 3$, $F_0 = \emptyset$, $F_1 = \{1, 2\}$, $F_2 = \{3, 4, 5, 8\}$, $F_3 = \{6, 7\}$),

g) Eq. 1a with unrestricted[5] daily variation matrix $\boldsymbol{\Gamma}$ ($g = 8$, $F_0 = \emptyset$, $F_j = \{j\}$, $j \leq 8$), and

h) Eq. 1b with unrestricted[6] daily variation matrix $\boldsymbol{\Gamma}$.

The latter two cases g) and h) are fitted with and without regularization. Both estimation strategies are combined with prediction as in Eq. 3a as well as Eq. 3b with ranks $r \in \{1, 2\}$.

A few words on the selection of model variants are in order. The first two variants d) and e) act as further benchmarks reflecting the simplest variants of Eq. 1a. The

---

[3]More specifically, $\overline{\widehat{\text{mape}}}(p \mid q) = \frac{1}{M} \sum_{m=1}^{M} \widehat{\text{mape}}_m(p \mid q)$ and

$$\left[\frac{1}{M-1} \sum_{m=1}^{M} \left(\widehat{\text{mape}}_m(p \mid q) - \overline{\widehat{\text{mape}}}(p \mid q) \right)^2 \right]^{1/2} , \tag{**}$$

wherein the additional subscript indicates the m-th of the M replications. If the M sample pairs were non-overlapping, then (**)$/\sqrt{M}$ should provide a good estimate of the variability of $\overline{\widehat{\text{mape}}}(p \mid q)$ (over hypothetical replications of the present study). The overlap potentially induces a downward bias of this estimate. A worst case (complete overlap) estimate is given by (**).

[4]The data set for Linz contains too little data on road type (frc) 2 for unrestricted estimation of $\boldsymbol{\Gamma}$. Therefore, "unrestricted daily variation" refers to $F_0 = \emptyset$, $F_1 = \{1, 2\}$, $F_{j-1} = \{j\}$ for $3 \leq j \leq 8$ together with restricted slopes $\beta_1 = \beta_2$ (and likewise for Eq. 1b if $q = \text{Linz}$).

[5]See footnote 4.

[6]See footnote 4.

Table 1 Mean Absolute Percentage Error Estimates (Overview)

	$p \mid q$	benchmarks			special cases of Eqs. 1a and 1b		
		a)	b)	c)	d)	e)	f)
on-site	Vienna | Vienna	56.27 (0.18)	28.89 (0.13)	23.68 (0.13)	24.69 (0.13)	23.29 (0.13)	**23.14** (0.13)
	Linz | Linz	43.72 (0.16)	26.69 (0.12)	25.56 (0.12)	26.21 (0.12)	25.47 (0.12)	**25.27** (0.12)
	Lyon | Lyon	62.15 (0.055)	22.84 (0.028)	22.51 (0.030)	21.86 (0.028)	21.65 (0.028)	**21.61** (0.028)
cross city	Linz | Vienna	43.72 (0.16)	26.21 (0.11)	**25.41** (0.11)	26.15 (0.11)	25.56 (0.11)	25.42 (0.11)
	Lyon | Vienna	62.15 (0.055)	25.79 (0.038)	**25.29** (0.037)	25.67 (0.082)	25.37 (0.07)	25.32 (0.067)

	$p \mid q$	g)	g) $r = 1$	g) $r = 2$	g) regl.	g) $r = 1$, regl.	g) $r = 2$, regl.
on-site	Vienna | Vienna	23.13 (0.13)	24.08 (0.14)	23.23 (0.13)	**22.67** (0.13)	23.56 (0.13)	22.72 (0.13)
	Linz | Linz	25.24 (0.12)	25.71 (0.12)	25.29 (0.12)	**25.23** (0.12)	25.71 (0.12)	25.28 (0.12)
	Lyon | Lyon	**21.45** (0.028)	21.82 (0.028)	21.53 (0.029)	**21.45** (0.028)	21.82 (0.028)	21.51 (0.029)
cross city	Linz | Vienna	25.41 (0.11)	25.82 (0.11)	25.52 (0.11)	**24.90** (0.11)	25.25 (0.11)	24.96 (0.11)
	Lyon | Vienna	25.38 (0.075)	25.91 (0.35)	25.30 (0.36)	24.81 (0.072)	25.41 (0.08)	**24.80** (0.074)

	$p \mid q$	h)	h) $r = 1$	h) $r = 2$	h) regl.	h) $r = 1$, regl.	h) $r = 2$, regl.
on-site	Vienna | Vienna	23.25 (0.13)	23.62 (0.13)	23.28 (0.13)	**22.78** (0.13)	23.14 (0.13)	22.79 (0.13)
	Linz | Linz	**25.23** (0.12)	25.39 (0.12)	25.27 (0.12)	25.26 (0.12)	25.44 (0.12)	25.32 (0.12)
	Lyon | Lyon	**21.45** (0.028)	21.73 (0.028)	21.46 (0.028)	21.48 (0.028)	21.69 (0.028)	21.52 (0.029)
cross city	Linz | Vienna	25.52 (0.11)	25.63 (0.11)	25.53 (0.11)	**25.00** (0.11)	25.10 (0.11)	25.01 (0.11)
	Lyon | Vienna	24.95 (0.064)	24.84 (0.34)	24.86 (0.35)	24.38 (0.055)	24.38 (0.054)	**24.37** (0.056)

grouping in f) showed the best performance among different configurations for the daily variation groups on the Viennese data (on-site). Its success reflects the similar loadings—shown in panel (e) and (f) of Fig. 2—of these groups with the first two signals—shown in panel (a) and (b) of Fig. 2. The most flexible configurations g) and h) show the best overall performance when fitted with regularization. The non-regularized cases help to judge the value of regularization.

5 Results

Table 1 summarizes the main results of our study. In particular, the upper third of Table 1 presents averages of mean absolute percentage error (MAPE) estimates from $M = 50$ replications of the same out-of-sample prediction exercises for six benchmark procedures. The parenthesized numbers are empirical standard deviations of these mean absolute percentage error estimates.

Table 2 Bias Estimates (Selected Models)

$p \mid q$	a)	b)	c)	g) regl.	h) regl.
V. | V.	−12.9 (0.01)	−0.48 (0.02)	**−0.004** (0.01)	**−0.004** (0.01)	−0.005 (0.01)
Li. | Li.	−9.32 (0.009)	0.159 (0.02)	0.0012 (0.02)	**0.0006** (0.02)	0.137 (0.02)
Ly. | Ly.	−17.1 (0.01)	0.115 (0.02)	0.0044 (0.02)	**0.0038** (0.01)	0.232 (0.02)
Li. | V.	−9.32 (0.009)	2.13 (0.02)	2.58 (0.01)	**2.06** (0.01)	2.2 (0.01)
Ly. | V.	−17.1 (0.01)	−3.36 (0.02)	**−1.04** (0.02)	−1.61 (0.03)	−1.51 (0.03)

Here predictions are calculated by Eq. 3a; the use of Eq. 3b is indicated by adding the choice for the rank r to the respective column title. Similarly, an additional label *regl.* signals regularized least-squares fitting. All numbers are multiplied by 100 to reflect percentages and rounded to at most 4 significant digits. Boldface indicates the best result within the respective row.

The supplementary Tables 2, 3, 4 and 5 follow the same layout and typographical conventions as Table 1. Therein, bias and root mean squared error estimates are averages as in Eq. 4a (or Eq. 4b) but over powers of the prediction errors $(y - \hat{y}(p \mid q))^s$ (or $(y' - \hat{y}'(p \mid q))^s \mathtt{mxsp}^s$) with $s = 1$ (bias) and $s = 2$ (root mean squared error), respectively. Again, the reported numbers are means and standard deviations over $M = 50$ replications. The abbreviations 'V.', 'Li.', and 'Ly.' replace 'Vienna'. 'Linz', and 'Lyon', respectively.

The table clearly shows that using the speed limit as the prediction of the actual traveling speed leads to a decisively worse performance than all its competitors. The latter is due to its ignorance of congestion, which is already clearly visible in Fig. 1. This is also seen in Table 2 which contains (averaged) bias estimates for the three benchmark procedures a), b), and c) alongside g) and h) fitted with regularization and using Eq. 3a for prediction. The latter two as well as the method c) advocated by [12] are unbiased for on-site prediction and feature a moderate bias when used for cross-city prediction. The speed limit a) is highly biased in both cases; simply scaling the speed limit b) removes this deficiency to a large extent. However, this simplest method a) exhibits the unique selling point of not requiring any estimation and thus no actual speed measurements. The identical on-site and cross-city performance of a) in Linz and Lyon is an obvious consequence.

The specialization f) of Eq. 1a performs best among the benchmarks shown in the upper third of Table 1. The less flexible c) seemingly outperforms f) in cross-city prediction; however, its superiority is small and lies within the sampling uncertainty. The performance of b), c), d), and e) ranges between that of a) and f). Regarding these methods, note that the cross-city prediction in Linz using b), c), and d) exceeds the respective on-site performance.

Table 3 shows (averaged) root mean squared error estimates for on-site and cross-city prediction in Linz using these methods. The latter measures are calculated from the same prediction errors as used for Table 1 to counter

Table 3 Root Mean Squared Error Estimates (Selected Models)

$p \mid q$	b)	c)	d)	e)
Li. | Li.	8.11 (0.009)	7.45 (0.007)	7.64 (0.007)	**7.38** (0.006)
Li. | V.	8.38 (0.009)	8.26 (0.009)	8.32 (0.009)	**8.14** (0.009)

Table 4 Mean Absolute Percentage Error Estimates (Selected Models with Single `mxsp` Slope Coefficient)

$p \mid q$	g) regl.	g) $r = 2$, regl.	h) regl.	h) $r = 2$, regl.
Li. | V.	**24.95** (0.11)	**24.95** (0.11)	25.21 (0.11)	25.21 (0.11)
Ly. | V.	24.04 (0.064)	23.96 (0.065)	23.76 (0.056)	**23.72** (0.058)

doubts regarding the results shown therein. More specifically, least-squares estimation implicitly optimizes the root mean squared error, and the on-site performance of b), c), and d) in Linz shown in Table 3 surpasses the respective cross-city result.

The second and third part of Table 1 concern the procedures g) and h). In summary, these procedures outperform the simpler alternatives and yield optimal results for both on-site and cross-city prediction when fitted with regularization and using Eq. 3a. Figure 3 shows how the gains in cross-city prediction are distributed over time. Panel (a) visualizes the differences between the (averaged) cross-city mean absolute percentage error estimates for g) together with Eq. 3a and fitted with regularization and the corresponding estimates for c). Panel (b) replaces g) with h). Notable improvements in cross-city performance are spread across the off-peak hours. In contrast, prediction quality remains largely unaltered during peaks.

Table 4 delivers (averaged) mean absolute percentage error estimates for cross-city prediction using two variants of g) and h) with restricted slope coefficients. It shows that the cross-city results of g) and h) can be improved in case of Lyon by enforcing an identical speed limit slope coefficient for all road types via $\beta_f = \beta$, $f = 1, ..., 8$. Cross-city prediction for Linz degrades under this constraint.

Similarly, the record of Eq. 3b is mixed. Tables 1 and 4 certify a slightly improved cross-city performance for Lyon and losses in prediction accuracy for Linz. In light of Table 3, one may suspect that gains are hidden by using the mean absolute percentage error. However, the corresponding root mean squared error estimates in Table 5 disprove this suspicion.

Table 5 Root Mean Squared Error Estimates (Selected Models)

$p \mid q$	g)	g) $r = 1$	g) $r = 2$	g) regl.	g) $r = 1$, regl.	g) $r = 2$, regl.
Li. | V.	8.09 (0.009)	8.21 (0.01)	8.12 (0.01)	**7.92** (0.009)	8.03 (0.009)	7.94 (0.01)
Ly. | V.	9.15 (0.03)	9.15 (0.2)	9.11 (0.2)	**8.93** (0.03)	8.97 (0.03)	**8.93** (0.03)
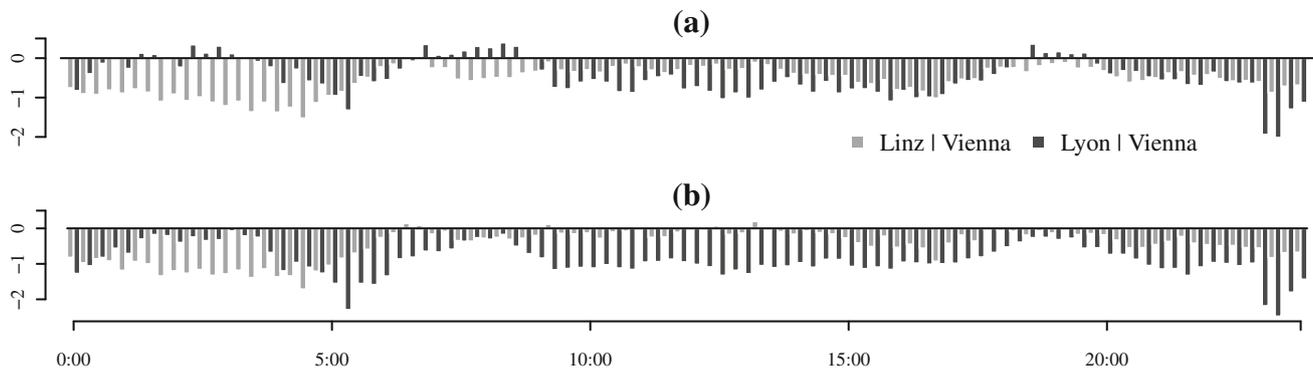
Fig. 3 Panel (**a**) shows the difference between the (averaged) mean absolute percentage error estimates for procedure g) using Eq. 3a and fitted with regularization and those for c) separately for the 96 time intervals. Panel (**b**) compares h) (using Eq. 3a and regularized fitting) with c) in the same way

A comparison of our main results (averaged MAPEs in Table 1) with competing studies in literature is in order. Moghaddam and Hellinga [14] provide models for predicting freeway travel times based on Bluetooth data. Thus, these authors consider an arguably simpler setting and data of higher quality. They find MAPEs of 13-18 %. Stathopoulos and Karlaftis [19] develop multivariate ARIMA and state space models in a setting similar to ours and obtain MAPEs between 12-20 %. Tulic et al. [22] investigate the Vienna taxi *FCD* in detail. They also build multivariate autoregressive models for a smaller number of links. In addition, these authors find that the MAPE for a given link varies strongly with the number of measurements on the link based on only a single observed taxi. Their (MAPE) results range from roughly 10 % (for links with almost no measurements based only on one taxi) up to 40 % (for links with almost all measurements being based on only one taxi) and with averages over all links ranging from 20 % to 30 % depending on the time of the day. In summary, the averaged MAPEs in Table 1 (ranging from 21 % to 25 %) are of similar magnitude as those found in related studies.

Finally, partial residual plots [6, Section 3.1.3] derived from the linear formulations showed no indication of a nonlinear effect of the speed limit `mxsp`. Nonetheless, we experimented with more flexible modeling of the effect of the speed limit `mxsp`, in particular, polynomial terms as well as more general non-parametric approaches, and orthogonality constraints on the columns of $\mathbf{\Gamma}$ to enforce zero variation in the mean at night time. We found no improvements in the prediction performance when using these extensions and therefore refrain from a detailed discussion.

6 Conclusion & outlook

This paper formulates several models expressing traveling speeds on a road link at a given time in terms of its road type (*frc*), its speed limit, and day time; see Section 3. Restricting the choice of covariates to time and static map information allows prediction when no measurement data is available. More specifically, we use the model fits derived from Viennese data to obtain predictions for the Austrian city Linz and the French city Lyon; see Section 4. We evaluate these transfer predictions using actual data for both cities and show that using the Viennese fit as a surrogate for a city-specific fit entails merely a moderate loss in prediction accuracy; see Section 5. We conclude that this transfer is a reasonable means to obtain traveling time predictions for cities which lack actual measurement data.

Concerning the model choice, we find that good on-site performance coincides with good transfer performance. We therefore suggest selecting a model based on its on-site prediction performance. In our study, a flexible model allowing distinct daily variation patterns for different road types (*frc*) together with regularized least-squares fitting dominates all its competitors. The Appendix provides instructions for a numerically efficient implementation of this procedure. In addition, we note that slight reductions in model complexity may further improve the transfer performance. Specifically, we consider equality constraints on the speed limit slope coefficients and a rank reduction of the daily variation coefficient matrix.

Finally, we list four possible directions for further research. Firstly, estimates for cross-city prediction can be obtained from a pooled data set consisting of data from two (or more) cities. Intuitively, these estimates should reflect peculiarities of individual cities to a lower degree than estimates based on data from a single city and therefore be better suited for cross-city prediction. A semi-transfer provides another alternative—applicable when too little data is available for a given city. More specifically, data fusion techniques can be used to "update" the transferred fit based on the available data prior to prediction. Secondly, our study focuses on three cities. Investigating whether our findings

generalize to a broader context is clearly important for the intended application. Thirdly, one may use different regularization constant λ_j for the daily variation pattern $\boldsymbol{\gamma}_j$ pertaining to different road types j to reflect the different levels of smoothness of the daily variation patterns shown in Fig. 1. Finally, a rank constraint could be added to the estimation of $\boldsymbol{\Gamma}$ to obtain an even better bias/variance trade-off. Therein, the appropriate rank may either be enforced as a hard constraint or estimated by penalization as in [17].

In conclusion, we can state that for the data sets examined in this paper out-of-sample prediction accuracy amounts to broadly 25 %. The transferal of model fits did not decrease the accuracy for very similar regions with Linz showing 25 % for on-site and cross-city evaluation. The penalty is higher for more distant cities: Lyon shows on-site errors of less than 22 % with transferal accuracy more than 24 %, thus, adding roughly 2.3 percentage points. It remains to be seen whether the more sophisticated models alluded to in the last paragraph can change this picture.

Appendix

A Implementation details

This appendix supplies the details on the calculation of the penalized least squares estimator used in the main text. In particular, efficient calculation of the estimates for a range of values of λ is explained, which facilitates the choice of an optimal λ by cross-validation. As an alternative, this section derives an explicit expression of the corresponding GCV-criterion of [7].

A change in notation simplifies the following presentation. The restrictions $\sum_{t=1}^{96} \bar{\gamma}_{t,j} = 0$, $1 \leq j \leq g$, in Eqs. 1a and 1b ensure that the intercepts c_f, $f \in \{1, \ldots, 8\}$, are identified. Alternatively, one may set $c_f = 0$ for one (reference) road type in each daily variation group F_j, $1 \leq j \leq g$. This alternative standardization produces the

same predictions[7]—irrespective of the choice of reference road types—and is used throughout this appendix. Then the model (1a) may be re-written as

$$y_i = \mathbf{x}_i^\mathsf{T} \begin{pmatrix} \bar{\boldsymbol{\beta}} \\ \text{vec}\bar{\boldsymbol{\Gamma}} \end{pmatrix} + u_i , \quad \text{wherein } \mathbf{x}_i^\mathsf{T} = \left(\mathbf{x}_{i,1}^\mathsf{T} \ \mathbf{x}_{i,2}^\mathsf{T} \right) \text{ with}$$

$$\mathbf{x}_{i,1}^\mathsf{T} = \left(\mathtt{mxsp}_i \ \mathbf{e}_{f(i)|8}^\mathsf{T} \ \mathtt{int}_{i,1} \ \ldots \ \mathtt{int}_{i,8-g} \right) \quad \text{and}$$

$$\mathbf{x}_{i,2}^\mathsf{T} = \mathbf{e}_{g(i)|g}^\mathsf{T} \otimes \mathbf{e}_{t(i)|96}^\mathsf{T} ,$$

wherein $g(i)$, $f(i)$, and $t(i)$ denote the frc-based daily variation group—out of the g possible groups, the road type (frc), and time of the i-th observation, respectively. The vector $\mathbf{e}_{p|q}$ denotes the p-th element of the standard basis of $\mathbb{R}^q$ and exemplifies the general use of lower case bold-face letters to represent vectors; upper case boldface letters symbolize matrices. The symbols $\otimes$, vec, and $^\mathsf{T}$ indicate Kronecker multiplication, the vec operator, and transposition, respectively. The indicator variable $\mathtt{int}_{i,j}$ equals one if the i-th observation comes from the j-th non-reference road type; $\bar{c}_j$ denotes the corresponding (possibly nonzero) intercept c_j. Hence, $\bar{\boldsymbol{\beta}}$ gathers the non-regularized[8] coefficients and is given by $(\boldsymbol{\beta}^\mathsf{T}, \bar{c}_1, \ldots, \bar{c}_{8-g})^\mathsf{T}$. The vector $\mathbf{x}_{i,2}$ and the matrix of regularized coefficients $\bar{\boldsymbol{\Gamma}} = (\bar{\boldsymbol{\gamma}}_1, \ldots, \bar{\boldsymbol{\gamma}}_g)$ represent the daily variation and are present only if $g > 0$. The discussion of model (1b) is essentially identical, and hence this appendix focuses exclusively on Eq. 1a.

To this end, denote the vector of response observations by $\mathbf{y}^\mathsf{T} = (y_1, \ldots, y_n)$ and the design matrix by $\mathbf{X} = \left(\mathbf{X}_1 \ \mathbf{X}_2 \right)$ with blocks $\mathbf{X}_1 = (\mathbf{x}_{1,1}, \ldots \mathbf{x}_{n,1})^\mathsf{T}$ and $\mathbf{X}_2 = (\mathbf{x}_{1,2}, \ldots, \mathbf{x}_{n,2})^\mathsf{T}$.

In this notation, the least-squares objective becomes

$$\left\| \mathbf{y} - \left(\mathbf{X}_1 \ \mathbf{X}_2 \right) \begin{pmatrix} \bar{\boldsymbol{\beta}} \\ \text{vec}(\bar{\boldsymbol{\Gamma}}) \end{pmatrix} \right\|^2 + \lambda \| [\mathbf{M} \otimes \boldsymbol{\Delta}_2] \text{vec}(\bar{\boldsymbol{\Gamma}}) \|^2$$

$$= \left\| \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} - \begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 \\ \mathbf{0} & \sqrt{\lambda}\mathbf{M} \otimes \boldsymbol{\Delta}_2 \end{pmatrix} \begin{pmatrix} \bar{\boldsymbol{\beta}} \\ \text{vec}(\bar{\boldsymbol{\Gamma}}) \end{pmatrix} \right\|^2 , \quad (5)$$

wherein the matrix $\mathbf{M}$ is a diagonal matrix with j-th diagonal element equal to the number of frc classes in daily variation group j. Finally, $\mathbf{0}$ symbolizes a matrix or vector of appropriate size and with all entries equal to zero.

In the unconstrained case $\lambda = 0$, least-squares solutions can be obtained as usual. Otherwise, a comparable multistage procedure becomes relevant as the choice of λ increases the numerical burden. Most of the latter relies on the assumption of a full column rank of the data part $(\mathbf{X}_1, \mathbf{X}_2)$ of the design matrix. This may be ensured by proper grouping of the frc classes.

Firstly, note that the matrices in Eq. 5 are tall in the sense that they contain many rows and comparatively few columns. Therefore, the dimension of the problem can be

---

[7]Note in this regard that $\boldsymbol{\Delta}\mathbf{1}$ with $\mathbf{1} = (1, \ldots, 1)^\mathsf{T}$ is zero.
[8]See footnote 7.

reduced significantly by virtue of a QR-decomposition $\mathbf{QR}$ of $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{y})$. The latter leads to

$$\left\| \begin{pmatrix} \mathbf{R}_1 & \mathbf{R}_{1,2} & & \mathbf{r}_{1,3} \\ \mathbf{0} & \mathbf{R}_2 & & \mathbf{r}_{2,3} \\ \mathbf{0} & \mathbf{0} & & r_3 \\ \mathbf{0} & \sqrt{\lambda}\mathbf{M}\otimes\boldsymbol{\Delta}_2 & & \mathbf{0} \end{pmatrix} \begin{pmatrix} \bar{\boldsymbol{\beta}} \\ \mathrm{vec}(\bar{\boldsymbol{\Gamma}}) \\ -1 \end{pmatrix} \right\|^2 .$$

In addition, the triangular form of the upper part—given by $\mathbf{R}$—allows to first focus on $\hat{\bar{\boldsymbol{\Gamma}}}$. Subsequently, the optimal $\hat{\bar{\boldsymbol{\beta}}}$ follows from (back)solving the triangular system

$$\mathbf{R}_1\hat{\bar{\boldsymbol{\beta}}} = \mathbf{r}_{1,3} - \mathbf{R}_{1,2}\mathrm{vec}(\hat{\bar{\boldsymbol{\Gamma}}}) .$$

Solving the least-squares problem

$$\left\| \begin{pmatrix} \mathbf{R}_2 \\ \sqrt{\lambda}\mathbf{M}\otimes\boldsymbol{\Delta}_2 \end{pmatrix} \mathrm{vec}(\bar{\boldsymbol{\Gamma}}) - \begin{pmatrix} \mathbf{r}_{2,3} \\ \mathbf{0} \end{pmatrix} \right\|^2 .$$

yields the daily variation matrix estimate $\hat{\bar{\boldsymbol{\Gamma}}}$, which therefore equals

$$\begin{aligned} \mathrm{vec}(\hat{\bar{\boldsymbol{\Gamma}}}) &= \left[\mathbf{R}_2^\mathsf{T}\mathbf{R}_2 + \lambda\mathbf{M}^2\otimes\boldsymbol{\Delta}_2^2\right]^{-1}\mathbf{R}_2^\mathsf{T}\mathbf{r}_{2,3} \\ &= \mathbf{R}_2^{-1}\left[\mathbf{I} + \lambda\mathbf{S}\mathbf{S}^\mathsf{T}\right]^{-1}\mathbf{r}_{2,3} , \\ \mathbf{S} &= (\mathbf{R}_2^\mathsf{T})^{-1}(\mathbf{M}\otimes\boldsymbol{\Delta}_2) . \end{aligned} \tag{6}$$

In Eq. 6—and subsequently, $\mathbf{I}$ denotes the appropriately sized identity matrix. In addition, a full singular value decomposition (SVD) $\mathbf{S} = \mathbf{U}_S\mathbf{D}_S\mathbf{V}_S^\mathsf{T}$ of $\mathbf{S}$ aids the consideration of a range of values for λ. Herein $\mathbf{U}_S$ and $\mathbf{V}_S$ are orthogonal matrices and $\mathbf{D}_S$ is diagonal with nonnegative entries. In fact, this decomposition leads to the representation

$$\mathrm{vec}(\hat{\bar{\boldsymbol{\Gamma}}}) = (\mathbf{R}_2^{-1}\mathbf{U}_S)\left[\mathbf{I} + \lambda\mathbf{D}_S^2\right]^{-1}(\mathbf{U}_S^\mathsf{T}\mathbf{r}_{2,3}) , \tag{7}$$

which reduces the repeated calculation of $\mathrm{vec}(\hat{\bar{\boldsymbol{\Gamma}}})$ to a sequence of weighted matrix-vector multiplications.

The representation (7) also facilitates the calculation of the corresponding GCV criterion for a range of values for λ. Specifically, the numerator of this criterion—given by the sum of squared residuals—becomes

$$\underbrace{\|(\mathbf{I} - \mathbf{U}_S[\mathbf{I} + \lambda\mathbf{D}_S^2]^{-1}\mathbf{U}_S^\mathsf{T})\mathbf{r}_{2,3}\|^2}_{\|\lambda\mathbf{D}_S^2[\mathbf{I}+\lambda\mathbf{D}_S^2]^{-1}\mathbf{U}_S^\mathsf{T}\mathbf{r}_{2,3}\|^2=\sum_j\left(\frac{\lambda d_j^2}{1+\lambda d_j^2}\right)^2\bar{r}_j^2} + r_3^2 ,$$

wherein d_j and $\bar{r}_j$ represent the j-th diagonal element of $\mathbf{D}_S$ and the j-th entry of $\mathbf{U}_S^\mathsf{T}\mathbf{r}_{2,3}$, respectively. Furthermore, the corresponding denominator is given by the trace of

$$\mathbf{I} - \begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{pmatrix} \begin{pmatrix} \mathbf{X}_1^\mathsf{T}\mathbf{X}_1 & \mathbf{X}_1^\mathsf{T}\mathbf{X}_2 \\ \mathbf{X}_2^\mathsf{T}\mathbf{X}_1 & \mathbf{X}_2^\mathsf{T}\mathbf{X}_2 + \lambda\mathbf{M}^2\otimes\boldsymbol{\Delta}_2^2 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}_1^\mathsf{T} \\ \mathbf{X}_2^\mathsf{T} \end{pmatrix} .$$

Linearity and the cyclic property of the trace yields

$$n - \mathrm{trace}\left[\begin{pmatrix} \mathbf{X}_1^\mathsf{T}\mathbf{X}_1 & \mathbf{X}_1^\mathsf{T}\mathbf{X}_2 \\ \mathbf{X}_2^\mathsf{T}\mathbf{X}_1 & \mathbf{X}_2^\mathsf{T}\mathbf{X}_2 + \lambda\mathbf{M}^2\otimes\boldsymbol{\Delta}_2^2 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}_1^\mathsf{T}\mathbf{X}_1 & \mathbf{X}_1^\mathsf{T}\mathbf{X}_2 \\ \mathbf{X}_2^\mathsf{T}\mathbf{X}_1 & \mathbf{X}_2^\mathsf{T}\mathbf{X}_2 \end{pmatrix} \right] ,$$

wherein n denotes the sample size. The first column block of the latter matrix follows from appending sufficient zero rows to the bottom of the $k_1 \times k_1$ identity matrix with k_1 being the number of columns of $\mathbf{X}_1$. The second column block amounts to the solution $(\mathbf{A}_1^\mathsf{T}, \mathbf{A}_2^\mathsf{T})^\mathsf{T}$ of

$$\begin{pmatrix} \mathbf{X}_1^\mathsf{T}\mathbf{X}_1 & \mathbf{X}_1^\mathsf{T}\mathbf{X}_2 \\ \mathbf{0} & \tilde{\mathbf{X}}_2^\mathsf{T}\tilde{\mathbf{X}}_2 + \lambda\mathbf{M}^2\otimes\boldsymbol{\Delta}_2^2 \end{pmatrix} \begin{pmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1^\mathsf{T}\mathbf{X}_2 \\ \tilde{\mathbf{X}}_2^\mathsf{T}\tilde{\mathbf{X}}_2 \end{pmatrix} ,$$

wherein only $\mathbf{A}_2$ is needed for the present application, and the columns of $\tilde{mxx}_2$ equal the residuals from projecting the columns of mxx_2 onto the column space of mxx_1. The 731 equality $\tilde{\mathbf{X}}_2^\mathsf{T}\tilde{\mathbf{X}}_2 = \mathbf{R}_2^\mathsf{T}\mathbf{R}_2$ identifies the trace of $\mathbf{A}_2$ as

$$\mathrm{trace}\left[\mathbf{R}_2\left[\mathbf{R}_2^\mathsf{T}\mathbf{R}_2 + \lambda\mathbf{M}^2\otimes\boldsymbol{\Delta}_2^2\right]^{-1}\mathbf{R}_2^\mathsf{T}\right] = \mathrm{trace}\left[\mathbf{I} + \lambda\mathbf{D}_S^2\right]^{-1} .$$

In summary, the GCV criterion can be obtained as

$$\mathrm{GCV}(\lambda) = \frac{\sum_j\left(\frac{\lambda d_j^2}{1+\lambda d_j^2}\right)^2\bar{r}_j^2 + r_3^2}{n - k_1 - \sum_j\frac{1}{1+\lambda d_j^2}} ,$$

which is easily evaluated for a wide range of values for λ.

References

1. Antoniou C, Balakrishna R, Koutsopoulos H (2011) A synthesis of emerging data collection technologies and their impact on traffic management applications. Eur Transp Res Rev 3(0):139–148
2. Bovy P, Thijs R (2000) Estimators of travel time for road networks: new developments, evaluation results, and applications. Delft University Press
3. Castro PS, Zhang D, Li S (2012) Urban traffic modelling and prediction using large scale taxi gps traces. Pervasive Computing 57–72
4. Cetin M, List GF, Zhou Y (2005) Factors affecting minimum number of probes required for reliable estimation of travel time. Transp Res Rec.: J Transp Res Board 1917:37–44
5. Dixon K, Wu CH, Sarasua W, Daniel J (1999) Estimating free-flow speeds for rural multilane highways. Transp Res Rec.: J Transp Res Board 1678:73–82
6. Fahrmeir L, Kneib T, Lang S, Marx B (2013) Regression: models, methods and applications Springer Science & Business Media
7. Golub GH, Heath M, Wahba G (1979) Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics 21(2):215–223
8. Graser A, Leodolter M, Koller H (2015) Towards better urban travel time estimates using street network centrality. In: 1St ICA european symposium on cartography, Vienna, November 10-20, 2015
9. Graser A, Leodolter M, Koller H, Braendle N (2016) Improving vehicle speed estimates using street network centrality (under review) (nd)
10. Jenelius E, Koutsopoulos H (2013) Travel time estimation for urban road networks using low frequency probe vehicle data. Transp Res B 53:64–81
11. Jones M, Geng Y, Nikovski D, Hirata T (2013) Predicting link travel times from floating car data. In: 16th International IEEE Conference on Intelligent Transportation Systems, pp 1756–1763

12. Leodolter M, Koller H, Straub M (2015) Estimating travel times from static map attributes. In: Models and technologies for intelligent transportation systems (MT-ITS), 2015, pp 121–126
13. Liu H (2008) Travel time prediction for urban networks. Ph.D. thesis, TU Delft
14. Moghaddam S, Hellinga B (2013) Quantifying measurement error in arterial travel times measured by bluetooth detectors. Transp Res Rec.: J Transp Res Board 2395:111–122
15. Moses R, Mtoi E (2013) Evaluation of Free Flow Speeds on Interrupted Flow Facilities. Tech. rep., Department of Civil Engineering, FAMU-FSU College of Engineering., Tallahassee U.S
16. Musolino G, Polimeni A, Rindone C, Vitetta A (2013) Travel time forecasting and dynamic routes design for emergency vehicles. Procedia-Soc Behav Sci 87:193–202
17. Negahban S, Wainwright MJ (2011) Estimation of (near) low-rank matrices with noise and high-dimensional scaling. The Annals of Statistics 1069–1097
18. Simroth A, Zhle H (2011) Travel time prediction using floating car data applied to logistics planning. IEEE Trans Intell Transp Syst 12(1):243–253
19. Stathopoulos A, Karlaftis MG (2003) A multivariate state space approach for urban traffic flow modeling and prediction. Transp Res C Emerg Technol 11(2):121–135
20. (2010) Transportation Research Board: Highway capacity manual. Tech. rep., Washington D.C. http://hcm.trb.org
21. Tseng PY, Lin FB, Shieh SL (2005) Estimation of free-flow speeds for multilane rural and suburban highways. Journal of the Eastern Asia Society for Transportation Studies (6), 1484–1495
22. Tulic M, Bauer D, Scherrer W (2014) Link and route travel time prediction including the corresponding reliability in an urban network based on taxi floating car data. Transportation Research Record: Journal of the Transportation Research Board 2442
23. Zheng F (2011) Modeling Urban Travel Times. Ph.D thesis TU Delft