Empirical speed models for cycling in the Oslo road network

Authors: Stefan Flügel, Nina Hulleberg, Aslak Fyhri, Christian Weber, Gretar Ævarsson

Institute of Transport Economics, Oslo (TØI)

Corresponding author: Stefan Flügel, Institute of Transport Economics Gaustadalleen 21, 0349 Oslo, sfl@toi.no.

Abstract:

Knowing the speed at which a cyclist travels is important in route and mode choice modelling. Empirical evidence suggests that it varies significantly in accordance with – among other things – infrastructure and topology. Despite this, in many network-based transport models cycling speed is constant, making travel distance the predominant variable of cycling behavior.

Motivated by the lack of a comprehensive speed model in the literature, we present models for bicycles and e-bikes estimated based on a large-scale collection of GPS data in the Oslo area. In the models, speed on a network link is described as a function of several characteristics of the infrastructure and topology, and differs by user segments such as gender, trip purpose and type of bicycle. Model parameters are estimated with regression models using data from close to 50 000 single cycling trips.

The data indicate that, on average, men cycle at a faster rate than women, although the difference is significantly less in the case of e-bikes. There is a non-linear and non-monotonic relationship between speed and gradient, with speed increasing up to a gradient of –6%, but decreasing thereafter most likely due to safety concerns. Notable is the fact that cycling speed is significantly higher on routes where cyclists and pedestrians have their own dedicated space.

Keyword: *bicycle, e-bikes, speed model, road network*

1. Introduction

The speed at which cyclists travel is a direct consequence of cycling behavior, which is in contrast to public transport, where travel speed is beyond the control of the traveller. As opposed to driving a car, where the upper and lower ranges of travel speeds are restricted by limits, cycling speeds are less regulated. Cycling speed is restricted primarily by the physical condition of the cyclist and varies with the effort put into it. The environment (e.g. weather, wind, slipstreams, road surface and topology) and to some extent bicycle technology are factors that influence the degree to which physical power translates to speed. In addition, safety motives in relation to road infrastructure and volume of traffic influence the speed at which cyclists choose to travel.

These considerations suggest that cycling speed varies greatly across individuals, regions, time of year, road sections and the bicycle itself. Variation in speed within and across studies is documented in Allen et al. (2008).

Cycling speeds influence travel times and therefore play an important role in transport model systems. Route choice models are typically based on the algorithm "choose the fastest available cycling route". Thus, knowledge about how speed varies across network links is of crucial information in route choice modelling. Cycling speed is also important knowledge in travel mode choice modelling, since travel time savings or losses of the alternative "cycle" influences the choice of whether to cycle or not. All other things equal, a transport model will predict a higher market share of cycling where it can take place at higher speeds.

Yet many transport models (such the Norwegian regional transport models; Madslien et al. 2005) impose a constant cycling speed when route choice is being modelled, thus making travel distance the sole determinant. Some models have route choice dependent on the characteristics of road infrastructure. For instance, Menghini et al. (2010) and Ehrgott et al. (2012) have established route choice models where qualitative information about the road network is among the generalized cost functions underlying the behavioral model. In this case, behavioral changes in route choice due to qualitative improvements in the road network (e.g. separating pedestrians from cyclists) can be predicted. Such route choice models circumvent the need for separated speed models by directly modelling the effect of network characteristics on cyclist behavior without making cycling speed explicit. However, speed models are needed when one wants to predict the effect of changes in the road network on cycling speeds and travel times savings. Information about the latter could be crucial for economic appraisal studies, where travel time changes enter the economic accounting of the user benefits of road infrastructure projects. Applications to economic appraisal, which has a central role in Norwegian transport planning, was a major motivation for the establishing of the speed models presented in this paper.

One reason why so few transport models vary cycling speed across network links is likely a scarcity of comprehensive speed models in the scientific literature. El-Geneidy et al. (2007) were probably the first to study a comprehensive model where

speed is described as a function of network link characteristics. Some significant results were identified (see later discussion in Section 6) despite the low sample size of the study. To our knowledge, El-Geneidy et al.'s speed model was never implemented in a transport model. In fact, it seems that most behavior cycle models in which network link information is used within a transport model system are route choice models that do not have an underlying speed model. Models by Menghini et al. (2010) or Ehrgott et al. (2012) use link length (not link speed or link travel time) as one of the explanatory variables in their route choice models.[1] An alternative approach, where link attributes are first translated into link speed by means of a speed model and then consecutively used in route choice modelling, can be found in Ziemke et al. (2017), whose speed model, however, is limited and not estimated on empirical data.

In this paper, we present speed models (one for bicycle and one for e-bike) describing speed on a network link as a function of several characteristics of infrastructure and topology, as well as the user segment defined by gender, trip purpose and type of bicycle. We regard our model as the most comprehensive cycling speed model in the literature.

Besides its scope, a strength of the model is that it is estimated empirically based on revealed preference data. Thus, the model mirrors real world behavior and is therefore an empirical model as opposed to (i) pure theoretical models that are based on physical laws (see, e.g., Parkin and Rotheram (2010, page 4) and (ii) speed models that use given/assumed parameter values (as in Ziemke et al., 2017).

Another contribution of the paper is the speed modelling of e-bikes utilizing a data set with over 12 000 registered e-bike trips. In only a few previous studies (Schleinitz et al., 2017; Dozza et al., 2016) have separate models for e-bikes been estimated.

Our data were collected in Oslo, which has about 650 000 inhabitants and a GDP per capita of about 100 000 euros. Oslo is one of the wealthiest cities in Europe.[2] The mode share of cycling for all trips in Oslo was 4% in 2012 (URBANET 2013), and although this has increased in recent years, it is still lower than in most other European capitals. Barriers to cycling are long and cold winters, the hilly topology of the city and a lack of separate cycle lanes and pathways. The latter has recently received much attention in the political debate, resulting in city authorities allocating significant resources to improve cycling infrastructure.

The paper is structured as follows: the data are described in Section 2, the statistical model in Section 3, parameter estimates in Section 4, and calibration and implementation of the model in Section 5. Section 6 is a discussion of (a) methodological issues and possible future improvements, (b) how our results compare with the earlier literature and what policy implications can be drawn. Section 7 concludes.

---

[1] Also, Sener et al. (2009), who established a route choice model based on stated preference data, do not use cycling speed as an exploratory variable for route choice

[2] https://www.oecd.org/gov/regional-policy/resilient-cities-oslo.pdf (last retrieved 17 April 2017)

2. Data

Data collection

Cycling speed, the dependent variable of the model, is measured on the basis of GPS observations. This method is gaining in popularity in cycling research (Menghini et al., 2009; Winters et al., 2010; Broach et al., 2012; Dill et al., 2014) and in studies with the primary objective of measuring cycling speed (El-Geneidy et al., 2007; Dozza et al., 2016; Schleinitz et al., 2017).

The GPS observations in this study were recorded by means of the commercial mobile application Sense.Dat (www.dat.nl) downloaded and taken into use by the respondents.

The original data collection was initiated by Fyhri et al. (2016), along with respondents recruited from two samples: (a) 1,000 persons who had applied to take part in an e-bike subvention program in Oslo[3] and (b) a sample of 10,000 cyclists from Oslo drawn from a bike insurance register. Data collection was carried out over three rounds. In the first (carried out in January 2016), respondents were asked if they wanted to use the app to record all their travel at a later point. Information about the program and login was sent via email. The second round was an intermediate questionnaire survey for those who opted out of using the app (conducted in April). The final round was carried out from May to June 2016, and consisted of the app data collection and a questionnaire survey (completed between 26 May and 3 June).

A total of 3,132 participants (890 and 2,242 from the first and second sample, respectively) responded to the first survey. Of these, 1,376 participants (376 and 1,000) agreed to register their travel with the app, and were sent the information regarding its download and use, as well as a unique user ID for login. Not all those invited took the app into use, so the final sample of app users was 721 (161 vs. 560, respectively) participants.

Table 1 presents an overview of relevant background variables for the participating app users, and compares these with data from the total sample of cycle owners in Oslo, from which the current sample is partially drawn (Fyhri et al. 2016).

Table 1 Background characteristics of participating app users and the bicycle owners in Oslo (data from Fyhri et al 2016).

---

[3] The subvention program in Oslo amounted to 25% of cost of e-bike, max 500 €. The budget of the program was large enough to pay for 1000 e-bikes. The only conditions that had to be filled were that the applicant had to live in Oslo, the bike had to be an approved pedelec, it had to be registered in an insurance registry (to locate bicycles based on their frame number in the case of theft) and that they had to respond to a questionnaire prior to using the e-bike.

	App users	Bicycle owners
Mean age	45.2	47.4
% Female	41	44
% Employed	91	87
% Own/access to car	85	85
% Own/access to bicycle	97	98
% Used bicycle on interview day (June)	45	27
N	721	2242

Study participants differ in several respects from the broader population of bicycle owners: fewer of them are female and the employment rate among them is higher.

Registration of cycling information and geographical mapping
The app automatically recognized transport mode from the speed and motion pattern. According to the developers, the accuracy of recognition of transport mode is 90%. The app does not distinguish between bicycle and e-bike, however. Identification of the type of bicycle used was therefore based on the purchase date of the e-bike. All bike trips registered after the purchase date are considered e-bike trips.

The purpose of trips was given automatically by the app based on re-occurring places and trips. The user can state the purpose of the trip or correct the guess, in this way future trips being more accurate. However, the distribution of trip purposes in our data was substantially different from traditional survey data, indicating that the automatic registration of trip-purposes in Sense.Dat is not very reliable (see also section 6).

The app automatically registers when a trip is completed and stores a trip ID to single GPS observations. After data cleaning, the data set consists of the 48,633 cycling trips – 36,447 by bike and 12,186 e-bike trips – made by 709 persons identified by a user ID in the mobile application. The number of registered cycle trips per person in the data collection period (1 April to 31 June 2016) ranges from 1 to 411.

The mobile application maps raw GPS observation on Open Street-Map in an initial data cleaning process. This procedure removes noise from the GPS observations by forcing them onto the network links. For the scope of our study, we regard that possible effects/errors of this procedure to be uniform over the data set. We therefore decided to use the map-matched time and location stamps for our speed calculation rather than the one's from the raw GPS-observations.

For modelling building we used the official road network of the Public Roads administration of Norway (the national road database NVDB[4]). The reason for this is the need to have link information that is consistent with the network used in

---

[4] http://www.vegvesen.no/fag/teknologi/Nasjonal+vegdatabank/In+English (last retrieved 17 April 2017)

Norwegian transport models. Therefore, we projected GPS observations on the Oslo region NVDB network,[5] which consists of close to 50,000 links (see Figure 1).

Figure 1 The Oslo road network (NVDB).

Projections result in the attachment of link IDs onto single GPS observations,[6] which previously were marked with a user and trip ID as well as time and location stamps.

Calculation of speed at a link level

Based on this information, we calculate the average speed S_l on a network link l for a given cycling trip m. The following equation is used (indices for trips are suppressed).

---

[5] This was done by means of the spatial join approach in ArcMap (http://desktop.arcgis.com/en/arcmap/10.3/tools/analysis-toolbox/spatial-join.htm)

[6] In this connection, we excluded observations that where further than 8 meters from the actual link. The link represents the center line of the road, and since the roads we are interested in have one or two lanes in each direction, we consider 8 meters as a critical distance for an observation to take place on the given link.

$$(1) \qquad\qquad S_l = \frac{\sum_{i=1}^{n} d_{li}}{\sum_{i=1}^{n} t_{li}}$$

where

> i=1 is the first GPS observation for a given trip on link l,
>
> i=n is the last GPS observation for a given trip on link l,
>
> d_{li} is the distance (in meters) between observation i and $i-1$ identified by the related location stamps; we set $d_{l1} = 0$,
>
> t_{li} is the duration (in seconds) between observation i and $i-1$ identified by the related time stamps.[7]

Thus, only GPS observations mapped on a link are used to calculate speed on that link; speeds measured at link transitions are not included. Because speed may decrease at link transitions, due to braking and acceleration, we measure speed on a link level as higher than on a trip level. We therefore need to calibrate the speed model (see Section 5).

Alternatively, we could have included distance and duration between the first observation of link l and the last observation on link $l-1$ (or between the first observation on $l+1$ and the last of link l). This would have meant that the speed measurements were more continuous. However, it would also have led to inconsistency in the concept structures of the model, as the estimated effects of characteristics of link l would partly estimate speed associated with link $l-1$. This was deemed problematic, especially when l is short or when the last observation on $l-1$ is far from the link transition. Note that after projection on the applied network there are links without any (valid) GPS observation, which means that some speed measures would have been based on speed on $l-2$, $l-3$ etc.

Besides the need to calibrate the speed model, the approach chosen has implications for the total sample size of the estimation model. That is because trip-link pairs with only one valid GPS observation could not be used for estimation (see next section)

3. Statistical model

We specify the same statistical model for both bicycle and e-bike, but estimate two sets of coefficients by means of separate model runs.

The dependent variable in the model is average speed on link l calculated by Equation (1). Speed is specified by a function of the user segment, defined by gender g and trip purpose p, link characteristic $X_{k,l}$ and related coefficients β_k, as well as an error term ε.

$$(2) \qquad S_{l,g,p} = e^{\beta_o + \beta_{men} * D_{g=1} + \beta_{work} * D_{p=1} + \sum_k (\beta_k * X_{k,l}) + \varepsilon}$$

---

[7] Note that $t_{li} = d_{li}/s_{li}$ where s_{li} is the speed between observation i and i-1. Thus, Equation (1) corresponds to the weighted harmonic mean of speed measurements along link l.

where β_o is the constant term, $D_{g=1}$ a dummy variable that equals 1 if cyclist is male, $D_{p=1}$ a dummy that equals 1 if the cycle trip is registered as a work-related trip. $X_{k,l}$ is a set of variables describing characteristics of network link l. ε are independent and identically (IID) normally distributed error terms.

The main argument for modelling speed with an exponential function is that it allows calibration of the model at a later stage with respect to only a single parameter and without the need to scale the beta-coefficients. After log transformation, the linear regression model can be estimated using the least-squares method in standard statistics programs (here SPSS[8]).

We weight observations by link length (meters), putting higher weights on longer links. As a direct consequence, coefficient estimates are affected more by longer links where speed measures are presumably more robust and where the relationship between link characteristics and link speed is profound.[9] Weighting also makes sense from a model application point of view, as the speed model should predict average speed (travel times) for entire cycle trips. The model should therefore have a relatively better fit for long links that contribute more to the calculation of travel times on a trip level. To avoid bias in coefficient estimates for explanatory variables that are likely to depend on link length, we applied interaction effects. In final model specification, this was done for the effect of T and X crossings, for which the relative impact of braking and acceleration is likely to depend on link length. Technically, we have simply specified 12 dummy variables based on combinations of the dummy variables for type of crossing and dummy variables for the link length groups.

Having a large database at hand, we decided to exclude from the estimation observations with presumed low data quality. As already mentioned (in Section 3), the applied methodology of speed calculation at a link level required that we discarded link-trips pairs with only one GPS observation per link. Furthermore, we decided to drop all observations where we did not observe at least 75% of the link length with GPS observations. A few links in our network had an unreasonably high net gradient attached.[10] We excluded all gradients below –20% and above 20%, and also observations with unreasonably high (more than 60 km/h) and low speeds (below 5 km/h) on a link. The former was likely due to data issues or to a wrong registration of transport mode, while among the latter cases there were presumably many trips when the cycle was walked (rather than ridden). Finally, after some testing, we excluded very short links (below 10 meters) from the estimation, because the data quality, in particular the measurement of gradients, was regarded as inferior for short links.

The applied exclusion criteria are summarized in Table 2.

---

[8] https://www.ibm.com/analytics/us/en/technology/spss/
[9] For shorter links, speed will to a higher degree depend on the link characteristics of previous links. The goodness-of-fit increases considerably when using weighted least square regressions.
[10] The net gradient is identified by the difference in z-coordinates in link nodes and the length of the link. It is measured as percent points and represent the average gradient over the entire link.

Table 2 Applied criteria for exclusion of GPS-observations that could be mapped on the NVDB-network (original N=2848560).

Exclusion criteria	Excluded cases *
Only one GPS-observations on link	1278596
Less than 75% of the link length covered by GPS-observations	954974
Link gradient above 20% (link *i* and link *i-1*)	24273
Measured link speed below 5 km/h or above 60 km/h	204217
Measured average trip speed above 60 km/h	50
Link length short than 10 meters**	6546
Links that are not allowed for cycling (tunnels, motorways)	187
Remaining cases after data cleaning (%)	*379717 (13.3%)*

*) calculated from remaining cases after the previous exclusion rule
**) many observations on such short links are already removed by the first exclusion rule

These exclusions of unreasonable and "low quality" observations led to clear improvement in the goodness-of-fit in the estimation model. As t-statistics of the estimated parameters remained very high in general, the reduction in the number of observations was not regarded as a problem from a statistical perspective. Most importantly, the parameter estimates became more reasonable (comparing parameter sign and order with prior expectations) after the exclusion.

The following link characteristics are used in the model:

- Net gradient specified as 18 different dummy variables for ranges of gradients (the level "0% – 1% gradient" is normalized).
- Average net gradient of inbound (i.e. the preceding) links (continuous variable).[11]
- Horizontal curvature (continuous variable).
- Type of road specified by four dummy variables: cycling path, cycling lane, walk/cycle path and remaining roads (the latter is normalized).
- Type of crossing specified as 12 dummies depending on (1) T or X cross, (2) at start or end of link, (3) short, middle or long link.
- Main cycling route alongside major roads in and around the city (one dummy variable).

---

[11] This variable is motivated by the possibility that cyclists enter a given link at a higher (lower) speed if the preceding links have a downhill (uphill) slope. Compared to the gradient at the evaluated link (*i*), the average gradient of preceding links (*i*-1) is expected to be of minor importance for speed predictions on link *i*. We therefore model this variable not by a set of dummy variables (that can capture possible non-linearity) but as a simple continuous variable. This helps to limit the overall number of parameters in the model.

- Proxy for traffic density/safety concerns specified as four dummies depending on whether link is in city center and whether the road has a 30 km/h speed limit for cars. [12]

A more detailed overview of the explanatory variables – presented as descriptive statistics – and a short note about their technical coding is given in Appendix. Table A1 in the appendix also gives the a-prior expectation of the parameter signs.

4. Parameter estimation

Because of log transformation of the estimation model, the beta-coefficients (besides the constant term) are interpreted as percentage change given marginal change in the explanatory variable.

Tables 3 reports on the estimation results for the bicycle and the e-bike model.

Table 3 Estimation results.

Model		bicycle		e-bike	
Number of observations		*283104*		*96613*	
Adjusted R-square		*0.24*		*0.174*	
Coefficient	Measure of related variable	Value	95% CI	Value	95% CI
β_o	constant (unity)	3.008***	[3.002; 3.013]	3.109***	[3.1; 3.118]
male	Dummy	0.1298***	[0.1268; 0.1327]	0.0491***	[0.0448; 0.0535]
work-related trips	Dummy	0.1142***	[0.1112; 0.1173]	0.1071***	[0.1022; 0.112]
gradient below -9%	Dummy	0.0491***	[0.0291; 0.0692]	0.0518***	[0.0201; 0.0835]
gradient between. -9% and -7%	Dummy	0.1081***	[0.0934; 0.1227]	0.0617***	[0.0411; 0.0823]
gradient between -7% and -6%	Dummy	0.1357***	[0.1238; 0.1476]	0.1228***	[0.1025; 0.1432]
gradient between -6% and -5%	Dummy	0.1795***	[0.1703; 0.1888]	0.1861***	[0.1728; 0.1993]
gradient between -5% and -4%	Dummy	0.1802***	[0.1729; 0.1875]	0.1488***	[0.1373; 0.1603]
gradient between -4% and -3%	Dummy	0.1494***	[0.1423; 0.1565]	0.1196***	[0.1074; 0.1317]
gradient between -3% and -2%	Dummy	0.1124***	[0.1061; 0.1187]	0.0779***	[0.0682; 0.0876]
gradient between -2% and -1%	Dummy	0.0589***	[0.0533; 0.0645]	0.0312***	[0.0221; 0.0404]
gradient between -1% and 0%	Dummy	0.0412***	[0.0364; 0.046]	0.0196***	[0.0122; 0.027]
gradient between 0% and 1%	Dummy	normalized		normalized	
gradient between 1% and 2%	Dummy	-0.0973***	[-0.1028; -0.0918]	-0.0376***	[-0.0466; -0.0286]
gradient between 2% and 3%	Dummy	-0.1299	[-0.1362; -0.1236]	-0.0770***	[-0.087; -0.0671]
gradient between 3% and 4%	Dummy	-0.1951***	[-0.2019; -0.1884]	-0.0778***	[-0.0894; -0.0662]
gradient between 4% and 5%	Dummy	-0.2669***	[-0.2738; -0.26]	-0.1218***	[-0.1323; -0.1113]
gradient between 5% and 6%	Dummy	-0.3034***	[-0.3119; -0.2949]	-0.1448***	[-0.1566; -0.1329]
gradient between 6% and 7%	Dummy	-0.3854***	[-0.3956; -0.3753]	-0.1807***	[-0.1972; -0.1642]
gradient between 7% and 9%	Dummy	-0.3949***	[-0.4078; -0.382]	-0.2434***	[-0.2623; -0.2244]

---

[12] We considered road surface as an attribute, but decided not to include it in the models, partly because of missing/unreliable data information and partly because the vast majority of roads in Oslo are actually of concrete (almost no cobblestone as in many other cities centers). The omission of cycle's age is discussed in section 6.

gradient above 9%	Dummy	-0.4267***	[-0.4472; -0.4062]	-0.3232***	[-0.3537; -0.2926]
average gradient of all inbound links	continuous (%-gradient)	-0.3936***	[-0.4354; -0.3518]	-0.2946***	[-0.3583; -0.2309]
horizontal curvature	continuous (index [1])	-0.2230***	[-0.2357; -0.2103]	-0.1945***	[-0.2159; -0.1731]
cycling lane	Dummy	0.0815***	[0.077; 0.086]	0.0739***	[0.0667; 0.0811]
walk/cycle path	Dummy	0.0609***	[0.0565; 0.0654]	0.0852***	[0.0781; 0.0923]
cycle path	Dummy	0.1063***	[0.0994; 0.1132]	0.1234***	[0.1134; 0.1334]
other types of roads [2]	Dummy	normalized		normalized	
T-crossing at start of link; link < 30 meter	Dummy	-0.0928***	[-0.1131; -0.0725]	-0.1278***	[-0.1627; -0.0929]
T-crossing at end of link; link < 30 meter	Dummy	-0.0414***	[-0.0631; -0.0197]	-0.0041	[-0.0415; 0.0334]
X-crossing at start of link; link < 30 meter	Dummy	-0.1223***	[-0.1678; -0.0768]	-0.2438***	[-0.3269; -0.1606]
X-crossing at end of link; link < 30 meter	Dummy	-0.0908***	[-0.1402; -0.0414]	-0.0385	[-0.1153; 0.0384]
T-crossing at start of link; link >=30, < 100 meter	Dummy	-0.0490***	[-0.0573; -0.0406]	-0.0699***	[-0.0838; -0.0559]
T-crossing at end of link; link >=30, < 100 meter	Dummy	-0.0674***	[-0.0761; -0.0587]	-0.0710***	[-0.0857; -0.0563]
X-crossing at start of link; link >=30, < 100 meter	Dummy	-0.0351***	[-0.0409; -0.0293]	-0.0383***	[-0.0478; -0.0288]
X-crossing at end of link; link >=30, < 100 meter	Dummy	-0.0235***	[-0.0293; -0.0177]	-0.0375***	[-0.047; -0.0281]
T-crossing at start of link; link >= 100 meter	Dummy	0.0031	[-0.0006; 0.0068]	-0.0059**	[-0.0115; -0.0003]
T-crossing at end of link; link >= 100 meter	Dummy	-0.0187***	[-0.0224; -0.0151]	-0.0091***	[-0.0147; -0.0035]
X-crossing at start of link; link >= 100 meter	Dummy	-0.0054**	[-0.0105; -0.0004]	-0.0096**	[-0.0174; -0.0019]
X-crossing at end of link; link >= 100 meter	Dummy	-0.0326***	[-0.0376; -0.0276]	-0.0348***	[-0.0425; -0.0272]
Main cycling route	Dummy	0.1140***	[0.1094; 0.1186]	0.0953***	[0.0884; 0.1021]
city center road with car speed limit <= 30 km/h	Dummy	-0.2087***	[-0.2137; -0.2038]	-0.2147***	[-0.2227; -0.2067]
Non-city center road with car speed limit <= 30 km/h	Dummy	-0.1182***	[-0.1227; -0.1137]	-0.1166***	[-0.1238; -0.1094]
city center road with car speed limit >30 km/h	Dummy	-0.1252***	[-0.1295; -0.1209]	-0.1443***	[-0.1515; -0.137]
Non-city center road with car speed limit >30 km/h	Dummy	normalized		normalized	

1) Measured LL/AD – 1, where LL stands for link length and AD for air distance. A straight line (link) therefore has a curvature of 0. See Appendix A for details. 2) The location stamps and the projection did not make it possible to separate between cycling on roads and cycling on sideways (pedestrian path). ***,**,* significant different from zero at level 0.90, 0.95 and 0.99 respectively

The parameter estimate β_0 lets us calculate the average speeds (prior calibration) given that all explanatory variables are zero. This would be for the reference group of all dummy variables, i.e. Non-city center road with car speed limit >30 km/h, other types of road with gradient 0% to 1%. We can see that e-bikes have a higher speed and, looking at the confidence intervals, this is clearly significant.

All other things being equal, males cycle 13.0% faster compared to women, while the effect of gender on e-biking is just 4.9% (see Section 6).
Cycling on work-related trips seems to be at higher speeds. As mentioned in Section 2, there is uncertainty about the reliability of which trip purpose is identified in the

mobile application. However, work-related trips are highly significant; they are stable between bicycle and e-bike and follow a priori expectation. We are therefore reassured that the identification of work-related trips worked reasonably well.

The net gradient of the link is one of the most important factors in explaining variation in speed over links. For positive gradients (cycling uphill) there is, as expected, a monotone (and roughly linear) decrease in speed with increasing gradient. With a gradient above 9%, participants cycle 42.7% slower than on links with or without a marginal gradient (0–1%). As expected, the decrease is lower for e-bikes (32.3%).
The effect of the gradient for downhill cycling is non-monotonic. The highest speeds are on average estimated on gradients between 5 and 6%, while speed seems to decrease above 6%. This is probably due to braking motivated by safety concerns. Speed in relation to gradient for bicycle and e-bike is given in Figure 2, where the values are averages over segments and already calibrated to fit speed on trip level (see next section).
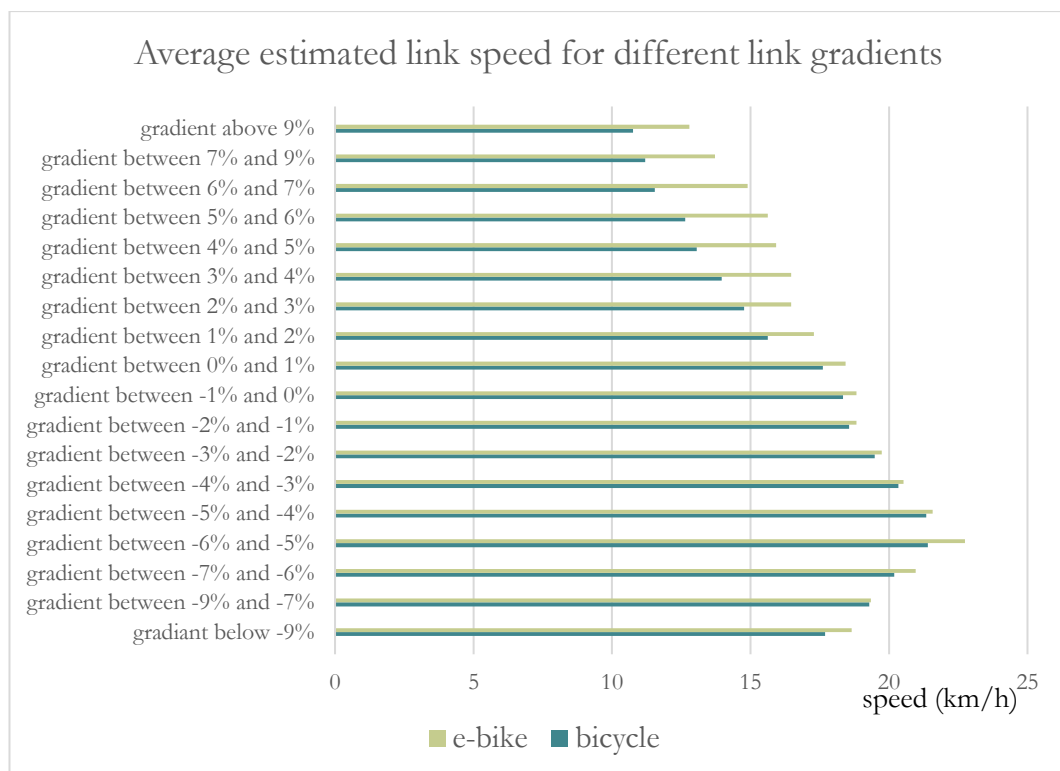

Figure 2 E-bike and bicycle speed given different gradients.

The coefficient continuously measured gradient of inbound links is not readily comparable to coefficients for the dummy variables on gradients on the actual link.[13] To get the marginal effect of a percentage point increase in gradient of the inbound

[13] In alternative models where both types of gradients are measured continuously (not reported here) the effects of the gradient of inbound links was about 1/6 of the effect of the gradient on the actual link.

link, the coefficient estimates of –0.3936 and –0.2946 are divided by 100. Interpretation is that an increase in average gradient of inbound links lowers speed for bicycle and e-bike by 0.39% and 0.29%, respectively. Thus, for gradient on the actual link, the effect on speed is lower for e-bike. This highlights how additional power from e-bikes helps to maintain speed when cycling uphill.

Horizontal curvature is measured by an index variable that equals 0 if the link is a straight line and 1 when link length is twice the air distance. The curvature effect on speed going from the latter to the former is estimated at –22.3% bicycle and –19.5% e-bike.

Estimates for type of cycling infrastructure are relative to the category "other type of roads", which entails streets not facilitated for cycling. We can see that highest speeds are estimated for cycling paths, where the cyclist is separated from cars and pedestrians. The relative effects compared to the category "other type of roads" are 10.6% (bicycle) and 12.3% (e-bike).
By comparison, there is a significantly lower effect for cycle lanes marked on roads otherwise used by cars. Here, speeds are estimated to be (just) 8.2% and 7.4% high. On paths shared by pedestrians and cyclists, cycling speed is 6.1% increased for bicycle compared to category "other type of roads". Interestingly, the effect is significantly higher (8.5%) for e-bikes.

For estimates relating to crossings, there are three general effects: (1) speeds are lower on trips with crossings, (2) speed decreases for X-crossings are higher than for T-crossings, and (3) the effect of crossings is greater for short links and small for long trips. All findings follow a priori expectation. There are no clear effects related to a crossing at the start versus end of a link. Also, the differences between bicycle and e-bike are ambiguous. As discussed in Section 6, estimates for crossings may be underestimated due to the way average trip speeds are measured at link level.

The variable "main cycling route" encompasses three main cycling routes into and around the inner city of Oslo. These routes are often used for longer distance cycling. The infrastructure is mixed in terms of types of road (separated cycling path, cycling lanes and walk/cycle path), but in general of good standard. The estimated relative effects on speed are 11.4% (bicycle) and 9.5% (e-bike).

The last set of explanatory variables (in Table 3) is of dummy variables for city center and car speed limit on the link. Car traffic is often regulated to 30 km/h and below for streets with a high density of pedestrians and/or cyclists. Furthermore, since the density of cyclists and pedestrians is generally higher in the city center, these variables may therefore function as a rough proxy for traffic density and related safety motives. Studying the estimated relative effects, we can see that both the city center and the reduced speed limit for cars have a significantly decreasing effect on cycling speed.

The goodness-of-fit of the model and other variables that might explain the variation in cycling speed are discussed in Section 6.

5. Calibration and implementation

As mentioned in Section 2, the method applied to measure average speed on link level (Equation 1) is an overestimation compared to speeds measured on a trip level. While calibration of the overall level of speed may not be necessary for route choice models that are based purely on travel times (as the fastest route remains the fastest route independently of the overall level of speed), a correct calibration of the model is needed for mode choice modelling and economic appraisal.

In order to calibrate the models for predicting – on average – correct travel times on a trip level, we introduce a calibration factor $C_{g,p}$ for each segment identified by gender g and trip purpose p. This is done separately for bicycle and e-bike with the following equation:

$$(3) \qquad C_{g,p} = \frac{\frac{1}{M_{g,p}}\sum_{m=1}^{M_{g,p}} S_{g,p,m}^{trip}}{\frac{1}{L_{p,g}}\sum_{l=1}^{L_{p,g}} \hat{S}_{g,p,l}^{link}} = \frac{\overline{S_{g,p}}^{trip}}{\hat{S}_{g,p}^{link}}$$

where $\overline{S_{g,p}}^{Trip}$ is the average speed measured over all trips $\left(M_{g,p}\right)$ of segment g,p [14] and $\overline{\hat{S}_{g,p}}^{Link}$ are the model predictions of average speed over links $L_{p,g}$ that we have data for.

For implementing the model, we calculate speed for all segments on all links in the network with the following equation

$$(4) \qquad S_{l,g,p} = \left(e^{\beta_0 + \beta_{men}*D_{g=1} + \beta_{work}*D_{p=1} + \sum_k(\beta_k * X_{k,l})}\right) * C_{g,p}$$

Average speeds at trip level and calibration factors for each segment are presented in Table 4.

---

[14] $S_{g,p,m}^{trip}$ is simply derived from the time and locations stamps between the first and last GPS observation of each trip.

Table 4 Average speed at trip level and calibration factors.

Segment			Observation on trip level $(M_{g,p})$	Average speed in km/h $(\overline{S_{g,p}^{Trip}})$	S.E.	5% Percentile	95% Percentile	Calibration factor $(C_{g,p})$
bicycle	female	Non-work	10961	15.16	.05	8.28	23.40	0.874
		work	2861	17.63	.08	10.44	25.20	0.852
	male	Non-work	17261	16.55	.04	8.28	25.56	0.840
		work	5364	20.82	.07	11.88	28.08	0.870
e-bike	female	Non-work	4201	16.60	.08	8.28	24.12	0.838
		work	1258	20.38	.13	12.24	26.64	0.876
	male	Non-work	5422	17.82	.08	8.64	25.92	0.857
		work	1305	21.80	.15	11.52	29.16	0.890

Figures 3a–d show how the average speed on a link calculated with Equation 3 varies over the network for uphill and downhill links and for bicycle and e-bike, respectively. Results are shown for the biggest segment (bicycle, male, non-work).
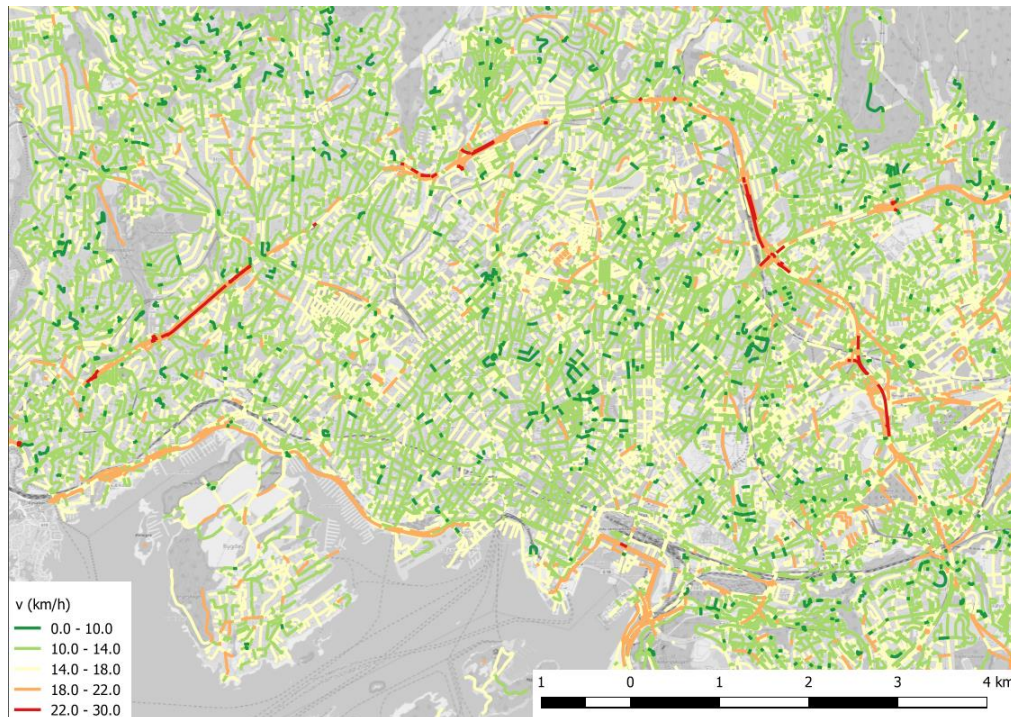

v (km/h)
- 0.0 - 10.0
- 10.0 - 14.0
- 14.0 - 18.0
- 18.0 - 22.0
- 22.0 - 30.0

Figure 3a Predicted speed (km/h) for **bicycle** *in the Oslo network for segment male/non-work:* **uphill links.**
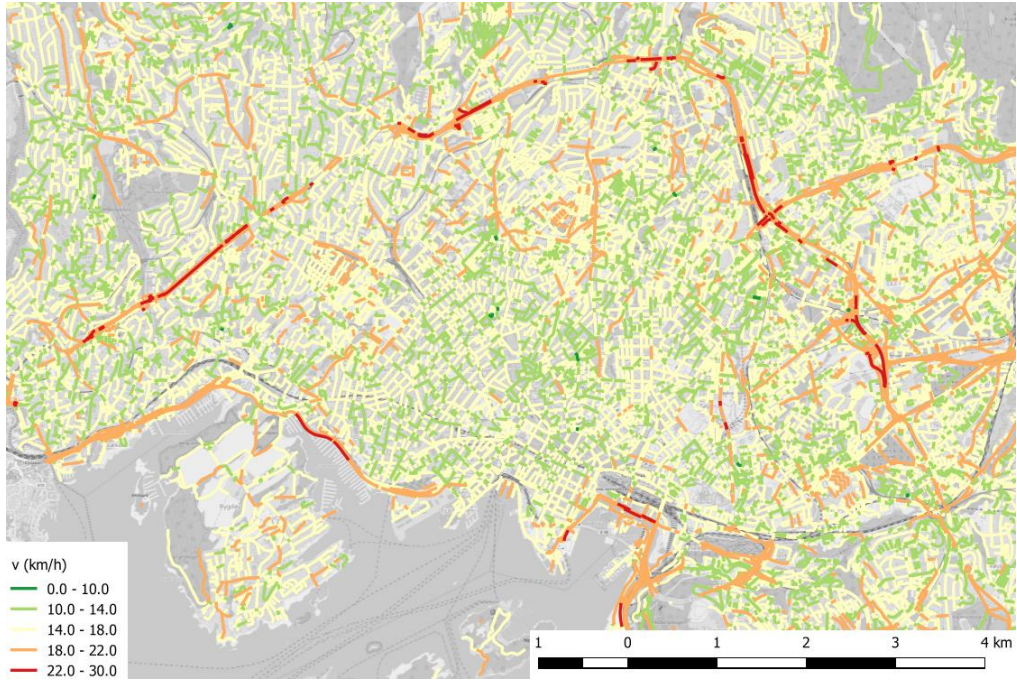
*Figure 3b Predicted speed for **e-bike** in the Oslo network for segment male/non-work: **uphill** links*
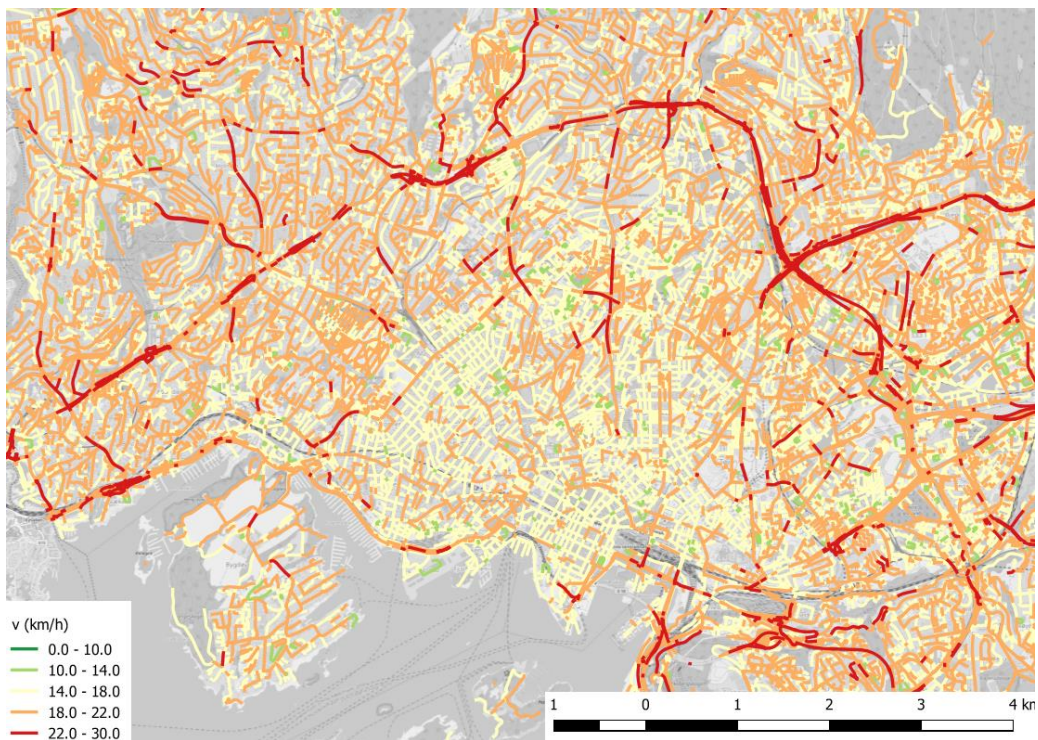

*Figure 3c Predicted speed for **bicycle** in the Oslo network for segment male/non-work: **downhill links.***
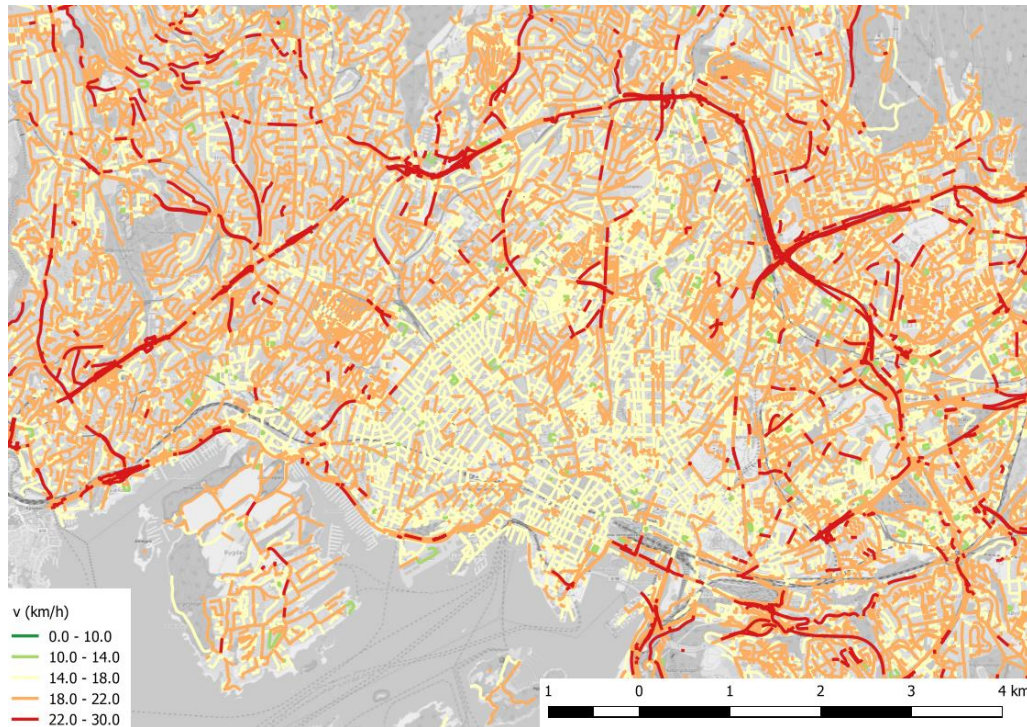
Figure 3d Predicted speed for **e-bike** *in the Oslo network for segment male/non-work:* **downhill links.**

The figures show the lower speed in Oslo city center. This relates to the flat geography, but is directly connected with the dummy variables for inner Oslo that were included in the model.

Figure 4 (a and b) shows the overall variation in predicted cycling speed over all user segments and network links for bicycle and e-bike, respectively.
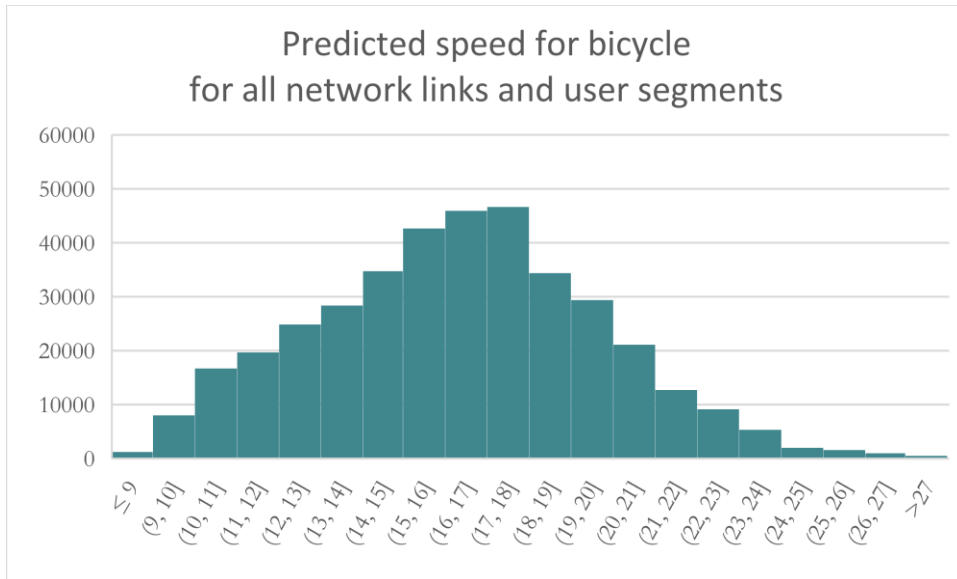
Figure 4a Histogram of predicted speed (km/h) for bicycle (all network links and population segments included).
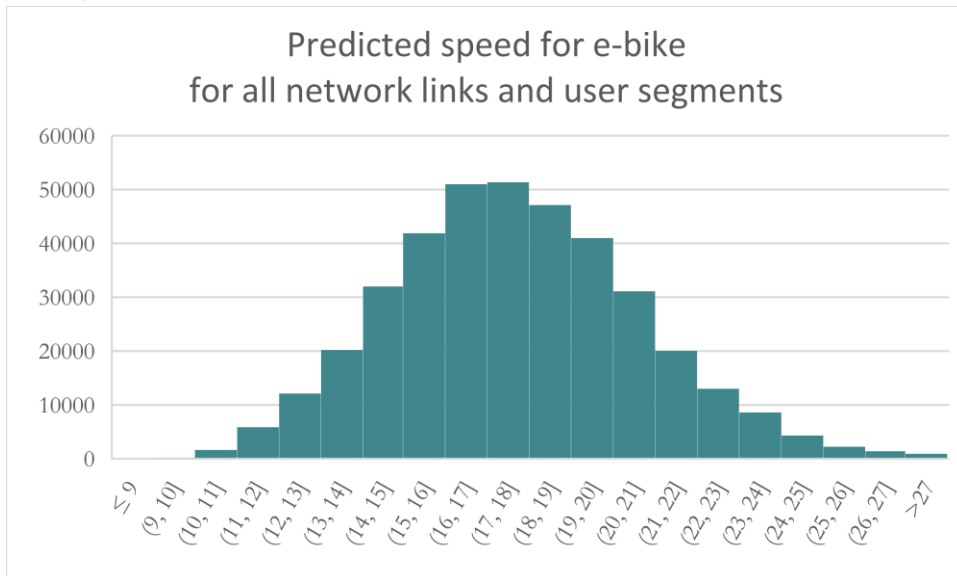

Figure 4b Histogram of predicted speed (km/h) for e-bike (all network links and population segments included).

Note that average values of the distribution depicted in Figure 4 (a and b) (16.3 km/h and 17.7 km/h, respectively) differ from averages measured at trip level (compare Table 4) as calculations at trip level are based only on links where cycling is actually taking place in our data collection.

6. Discussion

Methodological challenges

We start the discussion section by briefly elaborating the methodological issues of our study and possible improvements for future studies.

It is not easy to recruit people for studies based on GPS tracking and there is an obvious danger of sample selection bias, i.e. that those engaged in cycling and/or tend to cycle fast are more inclined to participate in the study. In addition, they might change behavior, e.g. cycle faster or slower due to social desirability effects (Podsakoff et al., 2003) when they know they are being tracked. This danger applies in our study as well; however, it could be argued that the problem is less pronounced here as we are using an app that tracks all movements/transport modes and is not a typical fitness or dedicated cycle app. Previous survey data of participants from the same sample population as used here have also indicated that their cycling behavior does not differ significantly from that of the general public in terms of self-reported speed choice or risky cycling (Fyhri et al., 2012).

The app's automatic registration of transport mode and trip purpose has measurement errors that are difficult to quantify. We attempted to circumvent wrongly registered cycling trips by excluding "unrealistic" travel speeds (below 5 and above 60 km/h). As mentioned in Section 2, the distinction between bicycle and e-bike trip is based on whether the respondents owned an e-bike at the time of the cycle trip, but this indirect measurement is likely to produce incorrect registrations. Given that more registered e-bike trips are actually by (regular) bicycle, the true difference in speed between e-bike and bicycle is potentially higher than estimated in our study. The same applies for the difference between work and non-work trips. Work trips are underreported in our study and if this is due to incorrect registration, the speed gap between work and non-work trips may be underestimated.

The method applied for measuring average link speed (see Section 2) is likely to underestimate the effect of crossings. Furthermore, the applied network had a poor coding of traffic lights, so we left this information out of the model. This is a clear way of improvement for future studies.

A variable we have information about but which was omitted from the models is cyclist's age. The relationship between age and speed was found to be non-linear with our data (Flügel et al., 2016), where the highest speeds were on average observed for the middle-age user group (35–55 years). To have included age in the model would therefore have required at least three new dummy variables and would have increased the number of different speed measures in the implemented model from 16 ($2^{\wedge 4}$

combinations of gender, trip purpose, type of cycle and direction) to 48. This was discarded from a practical point of view.[15]

The models presented in section 4 explain around 24% (bicycle) and 17% (e-bike) of the observed variation in cycle speed. Besides missing variables (such as traffic lights and age), a major reason for this relatively low level of explanatory power is likely related to measurement errors in the mapping of the GPS-data to our network. Figure 5 shows a plot of predicted average speed and measured average speed on a link level (here including data from both bicycle and e-bike).
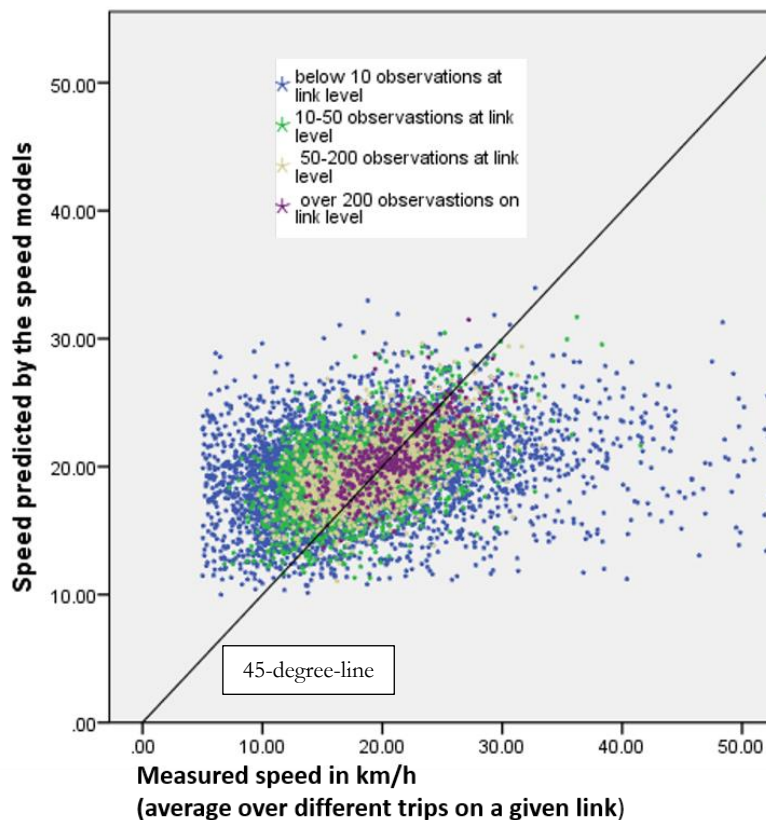

Figure 5 Measured versus predicted speed on a network link level (bicycle and e-bike combined)

From the color codes in Figure 5 one can see that network links with few observations (the blue ones) are pretty widely scattered, most likely because of measurement errors, while predicted values stay in the reasonable range of 10-30 km/h. For those cases, the model fit seems poor while in fact the model does a good job in "regressing out" the measurement errors in the empirical data.

---

[15] With additional models, we found the relative effects on speed for respondents under 35 years (compared to persons between 35-55 years) to be -1.5% for bicycle and 0.07% for e-bike. The relative effects for group over 55 years (again to persons between 35-55 years) are -6.1% and -7.2%. The adjusted R-square in the models included age were slightly higher as in the model in section 4 (0.242 for bicycle and 0.179 for e-bike).

Lastly, improvements to the statistical modelling are possible. The IID error terms applied are clearly restrictive, as we have repeated observations from identical trips and persons. As a test, we have performed models with fixed effects for respondents, accounting for the correlation in observations within a given persons and aiming to control for some of the unobserved variation on a person-level.[16] The adjusted R square statistics improved from 0.24 to 0.30 (bicycle) and from 0.17 to 0.26 (e-bike). These improvements may seem moderate and might again indicate that measurement errors are the most prominent source of the error term. The estimated beta-coefficients in these additional models are very close to the model presented in section 4, except for the dummy variable for gender. This parameter was only identifiable – as gender is constant for a given person – after removing some fixed effects. The estimated value for gender was furthermore deemed unreliable.[17] The personal fixed effects model was therefore discarded for model implementation.

Discussion of result and implications

Despite the methodological weaknesses mentioned, we are confident that our model is a sound representation of how cycling speed varies with different factors, primarily because sign and order of coefficient estimates follow a priori expectations and seem to fit well with findings in the literature.

It is not surprising that men cycle faster than women – a fact found consistently in the literature (El-Geneidy et al., 2007; Lin et al., 2008; Parkin & Rotheram, 2010). While the average speed is significantly different, our data show a high "gender-internal" variation in speed, and 25% of the faster female cyclists have higher speeds than the median male cyclist (Flügel et al., 2016). Our data also show that the gender gap in speed (on average) is clearly reduced for e-bike (around 5% compared to 13% for regular bicycle). Public subsidies for e-bikes (as in Oslo; Fyhri et al. (2016)) may therefore be interesting from the perspective of equity and may be seen as a move to get more female travellers into cycling.

The fact that e-bikes go faster than regular bicycles has been documented earlier in the literature (Schleinitz et al., 2017), and also our finding that they do particularly well when cycling uphill. Our modelling of downhill and uphill gradients showed a non-monotone and non-linear relationship which – to our knowledge – has not been measured to this degree of detail. The "braking effect" identified may have interesting implications from a road safety perspective, but needs to be verified in further studies.

Interestingly, evidence in the literature regarding the effects of cycle infrastructure is mixed. In one study, cycling in a separate infrastructure resulted in higher speeds (El-Geneidy et al., 2007), whereas in other studies there is no mention of this (Bernardi

---

[16] We considered also fixed effects for single trips to control for unobserved variation given the situation context of the cycling trips but did not performed such models as the number of fixed effects would have been impractically high with our data.

[17] After the necessary removing of the person fixed effects for implementation the model would have predicted that male persons cycle up to 50% faster than female persons. This contradicts our empirical evidence (see Table 4).

& Rupi, 2015; Schleinitz et al., 2017). The reason for the disparity is probably the presence of other road users on the infrastructure – pedestrians in particular. Our study shows that speeds are highest on roads where cyclists are kept separate from cars and pedestrians. Cycling speed is on average 20.6 km/h (bicycle) and 19.0 km/h (e-bike) on dedicated cycling paths, and lowest when cycling is not facilitated at all (17.6 km/h and 16.3 km/h). We also estimated values for marked cycling paths (on car streets) to be 19.8 km/h and 18.7 km/h. Estimated speeds in combined pedestrian/cycling streets (18.4 km/h, 16.7 km/h) suggest – from a speed evaluation point of view – that it is relatively more important to separate cyclists from pedestrians than cars. This may have interesting implications for transport infrastructure planning, as cyclists, just like other road user groups, need to minimize travel times.

It seems that our average predicted speeds are in the upper range of values in the literature.[18] In this connection, we have to point out that the concepts "average speed", "cruising speed", "speed on flat roads" and "speed identified by the constant term in regression models" are sometimes confounded in the literature, so clear comparison is not easy. In addition, methodological differences in study designs and cultural differences in the study areas are likely to explain differences in measured or predicted average speed. We regard our model as largely transferable to other cities, but a recalibration of the general speed level might be desirable.[19] It should be noted that cycling culture in Oslo is typical of that of many cities with relatively low cycling shares, in that it is characterized by a relatively high proportion of training oriented cyclists who cycle as a form of exercise (Fyhri et al., 2015), and many workplaces provide lockers and showers that facilitate "high-speed" cycling to work.

On the basis of the speed model and the variation in cycling speed it predicts, we plan to revise and improve the modelling of route and travel mode choices in different Norwegian transport models. We hope that our model contributes to an improved modelling of travel behavior in general and cycling behavior in particular.

7. Conclusion

Our large-scale collection of GPS data on cycling in Oslo shows that cycling speed varies greatly over network links, user segments and type of bicycle (regular bicycle and e-bike). Significant and expected effects were found for most parameters, including type of infrastructure, type of crossing and link gradient. The assumption of constant link cycling speed in the clear majority of transport models is therefore

---

[18] Other studies using GPS measurement have found lower cruising speeds, such as 15.3 km / h (Schleinitz et al., 2017); 14 km / h (Dozza et al. 2016) and 16 km/h (El-Geneidy et al., 2007). In a study from Italy, using cameras to estimate speed average speeds ranged between 14,6 km / h (separate cycle path) and 22 km / h (mixed traffic)(Bernardi & Rupi, 2015).

[19] Technically an implementation to another network/city is done via equation 4. It involves inserting the estimated beta-parameters and calculated scaling factors reported in this paper, and applying it to the city specific network characteristics (the X vector). If not all X-variables are available, there might be the need adjust the constant term given that the normalization level is changed. A calibration to another average speed can directly be done by adjusting the scaling factors.

shown to be restrictive, and an implementation of cycling speed models is expected to increase the precision of transport model forecasts.

References

Allen, D., Rouphail, N., Hummer, J. & Milazzo, J. II. (1998). Operational Analysis of Uninterrupted Bicycle Facilities. Transportation Research Record 1638. 1998: p. 29-36. http://www.enhancements.org/trb%5C%5C1636-005.pdf.

Bernardi, S. & Rupi, F. (2015). An Analysis of Bicycle Travel Speed and Disturbances on Off-street and On-street Facilities. Transportation Research Procedia, 5, 82-94. doi: http://dx.doi.org/10.1016/j.trpro.2015.01.004

Broach, J., Dill, J. & Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. Transportation Research Part A: Policy and Practice, 46(10), 1730–1740. doi: 10.1016/j.tra.2012.07.005

Dill, J., McNeil, N., Broach, J. & Ma, L. (2014). Bicycle boulevards and changes in physical activity and active transportation: Findings from a natural experiment. Preventive Medicine, 69, S74–S78. doi: 10. 1016/j.ypmed.2014.10.006

Ehrgott, M., Wang, J. Y., Raith, A. & Van Houtte, C. (2012). A bi-objective cyclist route choice model. *Transportation research part A: policy and practice*, 46(4), 652-663.

El-Geneidy, A., Krizek, K. J., & Iacono, M. (2007). Predicting bicycle travel speeds along different facilities using GPS data: a proof of concept model Paper presented at the 86 th Annual Meeting of the Transportation Research Board, Washington.

Flügel, S., Hulleberg, N., Fyhri, A., Weber, C. & Ævarsson, G. (2016) Så fort sykler folk i Oslo, Samferdsel. https://samferdsel.toi.no/forskning/sa-fort-sykler-folk-i-oslo-article33490-2205.html (last retrieved 17 April 2017) (in Norwegian).

Flügel, S., Hulleberg, N., Fyhri, A., Weber, C., Ævarsson, G. & Skartland, E.-G. (2017) Fartsmodell for sykkel og elsykkel. TØI report 1557/2017. Oslo: Institute of Transport Economics (in Norwegian).

Fyhri, A., Bjørnskau, T. & Sørensen, M. W. J. (2012) Krig og fred: en spørreundersøkelse om samspill og konflikter mellom bilister og syklister (Vol. 1246/2012). Oslo: Transportøkonomisk institutt.

Fyhri, A., Heinen, E., Fearnley, N. & Sundfør, H. B. (2015). A push to cycling – Exploring perceived barriers to cycle use and willingness to pay for ebikes with a survey and an intervention study. Transportation, submitted.

Fyhri A., Sundfør, H. B. & Weber, C. (2016) Effekt av tilskuddsordning for elsykkel i Oslo på sykkelbruk, transportmiddelfordeling og CO2 utslipp. TØI rapport 1498/2016 (in Norwegian)

Madslien, A., Rekdal, J. & Larsen, O. I. (2005). Utvikling av den regionale modeller for persontransport i Norge. Oslo: Institute of Transport Economics (in Norwegian).

Menghini, G., Carrasco, N., Schussler, N. & Axhausen, K. W. (2010). Route choice of cyclists in Zurich. Transportation Research Part A: Policy and Practice, 44(9), 754–765. doi: 10.1016/j.tra.2010.07.008

Parkin, J. & Rotheram, J. (2010). Design speeds and acceleration characteristics of bicycle traffic for use in planning, design and appraisal. Transport Policy, 17(5), 335-341. doi: http://dx.doi.org/10.1016/j.tranpol.2010.03.001

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y. & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. Journal of Applied Psychology, 88(5), 879.

Schleinitz, K., Petzoldt, T., Franke-Bartholdt, L., Krems, J. & Gehlert, T. (2017). The German Naturalistic Cycling Study – Comparing cycling speed of riders of different e-bikes and conventional bicycles. Safety Science. doi: http://dx.doi.org/10.1016/j.ssci.2015.07.027Dozza et al. (2016).

Sener, I. N., Eluru, N., & Bhat, C. R. (2009). An analysis of bicycle route choice preferences in Texas, US. Transportation, 36(5), 511-539.

URBANET (2013) Reisevaner i Oslo og Akershus. Analyser av Ruters markedsinformasjonssystem (MIS) PROSAM report nr. 202 (http://prosam.org/rapporter/pdf/202prosamrapport202bearbeidingavmis.pdf) (last retrieved 17 April 2017).

Winters, M. & Teschke, K. (2010). Route preferences among adults in the near market for bicycling: Findings of the cycling in cities study. American Journal of Health Promotion, 25(1), 40–47. doi: 10. 4278/ajhp.081006-QUAN-236.

Ziemke, D., Metzler, S. & Nagel, K. (2017) Modeling bicycle traffic in an agent-based transport simulation. 6th International Workshop on Agent-based Mobility, Traffic and Transportation Models, Methodologies and Applications, ABMTRANS 2017.

Attachment

Table A1 descriptive statistics and a-priori expectation of related coefficients.

Explanatory variable		bicycle			e-bike			a-priori expect sign of coefficient
		Mean	Min	Max	Mean	Min	Max	
male	dummy	.6672	0.00	1.00	.5362	0.00	1.00	+
work-related trips	dummy	.1849	0.00	1.00	.1808	0.00	1.00	+
gradient below -9%	dummy	.0094	0.00	1.00	.0102	0.00	1.00	+
gradient between -9% and -7%	dummy	.0135	0.00	1.00	.0178	0.00	1.00	+
gradient between -7% and -6%	dummy	.0180	0.00	1.00	.0184	0.00	1.00	+
gradient between -6% and -5%	dummy	.0265	0.00	1.00	.0292	0.00	1.00	+
gradient between -5% and -4%	dummy	.0453	0.00	1.00	.0462	0.00	1.00	+
gradient between -4% and -3%	dummy	.0527	0.00	1.00	.0459	0.00	1.00	+
gradient between -3% and -2%	dummy	.0726	0.00	1.00	.0753	0.00	1.00	+
gradient between -2% and -1%	dummy	.1029	0.00	1.00	.0957	0.00	1.00	+
gradient between -1% and 0%	dummy	.1398	0.00	1.00	.1444	0.00	1.00	+
gradient between 0% and 1%	dummy	.1505	0.00	1.00	.1477	0.00	1.00	Norm.
gradient between 1% and 2%	dummy	.1056	0.00	1.00	.0985	0.00	1.00	-
gradient between 2% and 3%	dummy	.0733	0.00	1.00	.0728	0.00	1.00	-
gradient between 3% and 4%	dummy	.0591	0.00	1.00	.0507	0.00	1.00	-
gradient between 4% and 5%	dummy	.0493	0.00	1.00	.0525	0.00	1.00	-
gradient between 5% and 6%	dummy	.0300	0.00	1.00	.0361	0.00	1.00	-
gradient between 6% and 7%	dummy	.0249	0.00	1.00	.0280	0.00	1.00	-
gradient between 7% and 9%	dummy	.0170	0.00	1.00	.0201	0.00	1.00	-
gradient above 9%	dummy	.0095	0.00	1.00	.0107	0.00	1.00	-
average gradient of all inbound links	continuous	-.0010	-.18	.20	-.0006	-.18	.20	-
horizontal curvature	continuous	.0287	0.00	1.50	.0287	0.00	1.50	-
cycling lane	dummy	.1314	0.00	1.00	.1213	0.00	1.00	+
walk/cycle path	dummy	.2300	0.00	1.00	.2546	0.00	1.00	+
cycle path	dummy	.0446	0.00	1.00	.0613	0.00	1.00	+
remaining type of roads	dummy	0.594	0.00	1.00	0.563	0.00	1.00	Norm.
T-crossing at start; link < 30 m	dummy	.0522	0.00	1.00	.0476	0.00	1.00	-
T-crossing at end; link < 30 m	dummy	.0440	0.00	1.00	.0403	0.00	1.00	-
X-crossing at start; link < 30 m	dummy	.0053	0.00	1.00	.0045	0.00	1.00	-
X-crossing at end; link < 30 m	dummy	.0047	0.00	1.00	.0054	0.00	1.00	-
T-crossing at start; link ϵ [30m,100m)	dummy	.2496	0.00	1.00	.2596	0.00	1.00	-
T-crossing at end; link ϵ [30m,100m)	dummy	.2475	0.00	1.00	.2575	0.00	1.00	-
X-crossing at start; link ϵ [30m,100m)	dummy	.0697	0.00	1.00	.0736	0.00	1.00	-
X-crossing at end; link ϵ [30m,100m)	dummy	.0723	0.00	1.00	.0751	0.00	1.00	-
T-crossing at start; link ≥ 100 m	dummy	.0669	0.00	1.00	.0639	0.00	1.00	-

T-crossing at end; link ≥ 100 m	dummy	.0623	0.00	1.00	.0581	0.00	1.00	-
X-crossing at start; link ≥ 100 m	dummy	.2822	0.00	1.00	.2723	0.00	1.00	-
X-crossing at end; link ≥ 100 m	dummy	.2745	0.00	1.00	.2715	0.00	1.00	-
Main cycling paths	dummy	.1331	0.00	1.00	.1579	0.00	1.00	+
city centre road, car speed limit ≤ 30 km/h	dummy	.1821	0.00	1.00	.1760	0.00	1.00	-
city centre road, car speed limit > 30 km/h	dummy	.2694	0.00	1.00	.2812	0.00	1.00	-
non city centre road, car speed limit ≤ 30 km/h	dummy	.3233	0.00	1.00	.3548	0.00	1.00	-
non city centre road, car speed limit > 30 km/h	dummy	.2252	0.00	1.00	.1880	0.00	1.00	+

Table A2 Note about technical coding of explanatory variables.

Explanatory variable	Technical coding
male	reported from the survey
work-related trips	identified by the app (optional changed by respondent)
gradient below -9%	Based on z-coordinates at start- and endnode
gradient between -9% and -7%	
gradient between -7% and -6%	
gradient between -6% and -5%	
gradient between -5% and -4%	
gradient between -4% and -3%	
gradient between -3% and -2%	
gradient between -2% and -1%	
gradient between -1% and 0%	
gradient between 0% and 1%	
gradient between 1% and 2%	
gradient between 2% and 3%	
gradient between 3% and 4%	
gradient between 4% and 5%	
gradient between 5% and 6%	
gradient between 6% and 7%	
gradient between 7% and 9%	
gradient above 9%	
average gradient of all inbound links	the mean of the gradient for all inbound links[20]
horizontal curvature	The ratio between shortest path and shape length minus value 1 (such that a straight link gets value 0)
cycling lane	Combining the definition of lanes and road type in the network
walk/cycle path	
cycle path	
remaining type of roads	
T-crossing at start; link < 30 m	Combining number of inbound links and length of link. The number of inbound links gives type of crossing (1 = None, 2 = T, 3 or more = X). Length is a network attribute.
T-crossing at end; link < 30 m	
X-crossing at start; link < 30 m	
X-crossing at end; link < 30 m	
T-crossing at start; link ∈ [30m,100m)	

---

[20] Inbound links are found by comparing start- and endnodes.

T-crossing at end; link ϵ [30m,100m)	
X-crossing at start; link ϵ [30m,100m)	
X-crossing at end; link ϵ [30m,100m)	
T-crossing at start; link ≥ 100 m	
T-crossing at end; link ≥ 100 m	
X-crossing at start; link ≥ 100 m	
X-crossing at end; link ≥ 100 m	
Main cycling paths	3 roads identified as main cycle roads. Coded manually by road number and geographic position.
city centre road, car speed limit ≤ 30 km/h	Combining speed limit and definition of city centre area. The centre zone is a concept variable from SSB (Statistics Norway). To be defined as a centre zone it must consist of several different functions and industries (e.g. it must have public administration or health services). We define a city centre road as a road located in the city centre area. The speed limit is defined in the network.
city centre road, car speed limit > 30 km/h	
non city centre road, car speed limit ≤ 30 km/h	
non city centre road, car speed limit > 30 km/h	