Improved Modal Split Indicators for transport statistics based on an effective distance matrix

Authors: Ekkehard Petri*, Boryana Milusheva, Gunter Schäfer – DG Eurostat/ European Commission

The European Commission White Paper on transport sets the target of a 60% reduction of greenhouse gas emissions for the transport sector by 2050. One of the proposed measures is a shift to more environment-friendly transport modes in particular for road freight journeys over 300km. Eurostat will produce the Modal Split Indicators (MSI) to measure the achievement of this goal. To calculate MSI a coherent approach for all transport modes (road, rail, inland waterways, air, maritime) is needed: tonne-kilometres as unit measure, broken down by distance class (above/ below 300km) per country. For road freight transport we propose a road distance matrix between NUTS level 3 regions based on a detailed road network using a fastest-path algorithm. We use transport data reported by the countries at region-to-region level, and compute tonne-kilometres performed on the countries' territories. This will enable Eurostat to calculate MSI per country without creating additional reporting obligations. A key success factor is to agree on the methodology with Member States. The method was piloted for road freight transport, and its principles will now be extended to the other transport modes.

1. Introduction

In 2011 the European Commission published a White Paper „Roadmap to a single European transport area – towards a competitive and resource efficient transport system" [1]. It sets 10 goals for a competitive and resource efficient transport system: benchmarks for achieving the 60% Green House Gas emission reduction target."

Shifting the transport of goods and passengers from more polluting to less polluting and more energy efficient transport modes is one possibility to achieve this target. In concrete terms the White Paper proposes to shift 30% by 2030 and 50% by 2050 of road freight travelling over distances longer than 300 km to rail or waterborne transport.

To monitor the achievement of this modal shift target, comparable data for all freight transported modes are required. The development of MSI initially focuses on inland transport modes and requires building time series for road, rail, and inland waterways. The current

---

*contact Ekkehard.Petri@ec.europa.eu

[1] COM(2011) 144 final "Roadmap to Single European Transport Area – Towards a competitive and resource efficient transport system.

system of relevant legal acts consists of one legal act per transport mode (road [2], rail [3], inland waters [4]). They require Member States (MS) to report their Transport Performance (TP) (tonnes transported multiplied by distance in kilometres = tonne kilometres or TKM) per mode.

2. Current situation and problem statement

The reporting on road freight transport (RFT) differs significantly from the other inland transport modes. For rail and inland waterways freight transport MS report the TP taking place on their territory (territoriality principle). For RFT the reporting is done according to the registration of the haulier. Countries design their surveys to cover journeys performed by their national vehicles and the TP is attributed to the country of registration. Hence, even cross border TP is fully attributed to the country of origin of the haulier. For MSI calculations Eurostat had to develop a coherent approach for measuring TKM across transport modes at MS and EU level, and apply the territoriality principle also to RFT (see [5] for additional information). This requires that for cross border RFT the territorial TP has to be computed, and journeys be broken down into national segments.

One way of modelling national contributions to cross-border TP is the use of a distance matrix, calculated along a road network. MS are confronted with a similar problem when they want to measure the inland TP of foreign hauliers on their territory, and follow similar approaches [6]. As distances are one factor in TP their computation has to be as accurate as possible. At the same time the territorialised TP should be coherent with data reported by MS. The so-called D3-tables [5] based on MS reporting contain the tonnes forwarded between NUTS level 3 (NUTS3) regions and the total TKM. However, it turned out that multiplying the D3 tonnes with the length of the national segments of cross border journeys as calculated by Eurostat yields TKM figures that are different from TKM figures in the D3-tables To solve this issue, Eurostat, in agreement with MS decided to adopt a different approach and calculate

---

[2] Regulation (EU) No 70/2012 European Parliament and of the Council of 18 January 2012 on statistical returns in respect of the carriage of goods by road (recast).

[3] Regulation (EC) No 91/2003 of the European Parliament and of the Council of 16 December 2002 on rail transport statistics.

[4] Regulation (EC) No 1365/2006 European Parliament and of the Council of 6 September 2006 on statistics of goods transport by inland waterways and repealing Council Directive 80/1119/EEC.

[5] http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Freight_transport_statistics_-_modal_split.

[6] Karner, T. Weninger, B. & Scharl, S. (2014), Estimation of the Inland Transport Performance from the Consolidated European Road Freight Transport data, Austrian Journal of Statistics, Vol. 43, 49--62.

the total distance between NUTS3 region pairs, and the length of each national segment. In a second step the shares of each national segment in the total length of the journey are computed, and then applied to the TKM data in the D3-tables, thus preserving reported TKM totals.

The ILSE matrix (Index of Locations for Statistics in Europe) forming the basis for this territorialisation method has been in use since 2008. For each pair of NUTS3 regions itineraries along a European road network and the national shares have been pre-calculated. The road network (see Figure 1, left) which formed the basis for the route calculation was fairly generalised and lacked further attributes such as speed information or road classes. Due to missing road class attributes, distance calculations had to be based on a shortest-path algorithm.

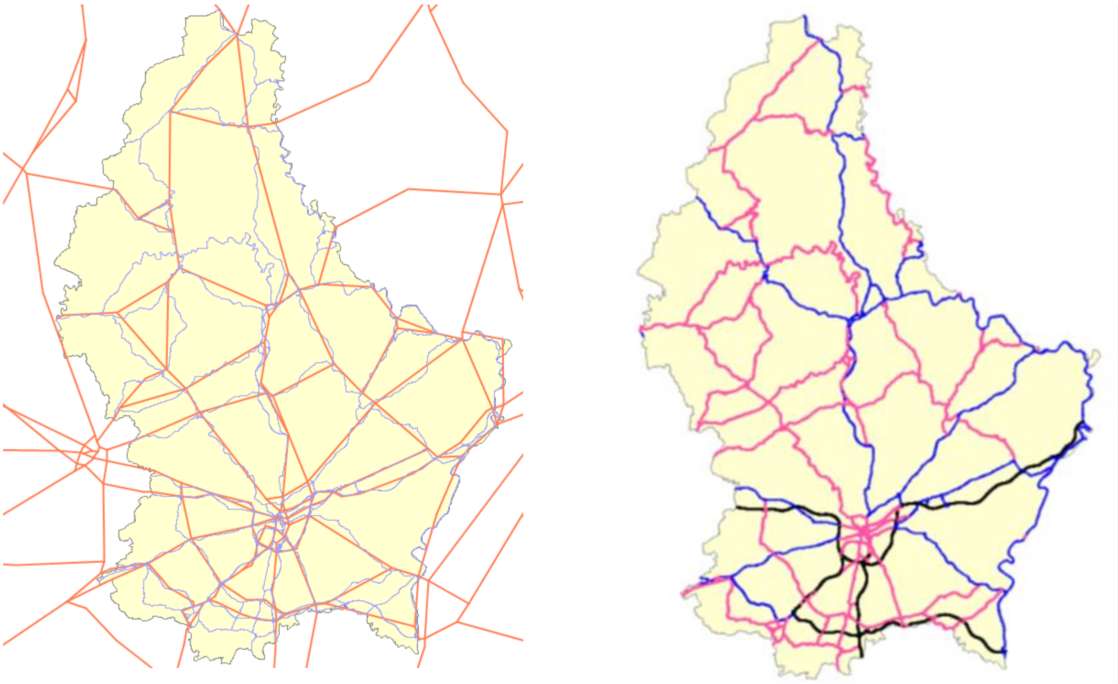

Figure 1: Old ILSE road network GISCO-IRPUD 1997 ©EuroGeographics (left) and New ILSE road network TomTom© MultiNet 2011, extract for Luxembourg.

An internal study conducted by Eurostat revealed that the computed distances in the ILSE matrix were significantly shorter than proposed itineraries by commercial internet route finders (Google©, ViaMichelin©, see Table 1). These differences can be explained by the fact that the ILSE road network was last updated in 1997, and used strongly generalised road lines compared to the detailed networks in modern route finders (see Figure 1, right). Also they normally use a fastest-path routing algorithm. Eurostat therefore decided to improve the quality of the ILSE distance matrix by updating the road network and the route calculation algorithm.

The main challenge was to select a European road network that would lead to a significant increase in the accuracy of distance calculations, and is of comparable quality in all EU MS, thus avoiding any national bias. Implementing a fastest-path routing requires additional road attributes, such as speed information and road classes.

For measuring modal shift over time, time series are essential, and hence the new distance matrix ideally should not off-set the time series existing since 2008. An assessment of the fulfilment of this condition was part of this study.

3. Data and methodology

We use the TomTom© MultiNet product as road network. We select NET2CLASS 0, 1, and 2 due to the interregional nature of transport [7]. We use pgRouting [8] in a PostgreSQL/PostGIS database to perform the routing calculation. The fastest-path is calculated based on a time attribute in the TomTom© data. The speed model distinguishes three speed levels: NET2CLASS level 0 = 130km/h, level 1 = 70km/h, level 2 = 40km/h) (see Figure 1, right). The lowest cost is calculated using the dijkstra_sp_delta method.

Each NUTS3 region is represented by its most populated settlement, as derived from EuroGeographics© EuroRegionalMap. If the settlement point is not located close to a road network entry point, the distance to the nearest road entry points is calculated as-the-crow-flies and added to the itinerary. We calculate the total distance between each pair of NUTS3 regions (422 000 pairs) and then intersect the path with country boundaries (EuroBoundaryMap ©EuroGeographics). This yields the national segment of each path. For quality benchmarking we draw a sample of journeys and compare them with routes from Google© Maps and ViaMichelin©.

We compute the territorialised national TP by splitting up the reported total TKM between NUTS3 pairs in proportion to the national share in the total distance. The TKM data are taken from the Eurostat D3 tables.

4. Results of the matrix update

Table 1 gives an overview of distances of a random selection of journeys between NUTS3 regions from the old ILSE matrix, the new ILSE matrix and two internet route finders. On average the difference between the new ILSE matrix and the average of the two route finders

[7] TomTom (2011), MultiNet user manual.

[8] http://pgrouting.org/

is +3%. This is more than the differences between the two route finders, which use different road networks and potentially routing algorithms, and therefore might be considered systematic and significant. It might be explained by limitations of our choice of cost factors for the routing. Where our distances are shorter this might be explained by our method of connecting the origin or destination point to the road network.

Table 1: Comparison of old ILSE and new ILSE distances with internet route finders.

From	From City	To	To City	Old WebILSE (shortest)	New WebILSE (fastest)	Google (EUROSTAT)	Michelin (EUROSTAT)	AverageCommercial	New WebILSE vs commercials
PT183	EVORA MONTE	DE211	NEUBURG A.D.D.	2124	2502	2356	2416	2386	5%
DEB38	SPEYER	RO414	SLATINA	1489	1763	1626	1622	1624	9%
ITC31	SAN REMO	SK023	NITRA	1056	1323	1274	1252	1263	5%
DEA1C	VELBERT	DE225	FREYUNG	622	690	693	689	691	0%
BE331	HUY-WANZE	CH057	RORSCHACH	524	612	741	743	742	-18%
PT16B	CALDAS DA RAINHA	ITG14	LICATA	2680	3117	3422	3402	3412	-9%
RO213	IASI	RO311	PITESTI	401	507	455	451	453	12%
DEB3E	GERMERSHEIM	BG323	RUSE	1645	2019	1788	1972	1880	7%
NL332	S GRAVENHAGE	DE731	KASSEL	412	444	429	429	429	4%
RO316	PLOIESTI	DEE08	ZEITZ	1412	1765	1647	1564	1605,5	10%
ITD51/ITH51	PIACENZA	DE426/DE408	RATHENOW	989	1158	1164	1146	1155	0%
RO224	GALATI	DE125	HEIDELBERG	1724	2197	2032	2026	2029	8%
UKN04	LONDONDERRY	UKK23	TAUNTON	579	872	866	806	836	4%
FR514	LE MANS	BE353	PHILIPPEVILLE	462	504	476	494	485	4%
AT130	WIEN	NL226	NIJMEGEN	949	1087	1077	1048	1062,5	2%
DE255	SCHWABACH	BG342	SLIVEN	1601	1845	1764	1777	1770,5	4%
DE21G	WALDKRAIBURG	HU313	SALGOTARIJAN	592	747	700	726	713	5%
DE219	EICHSTAETT	DEB34	LUDWIGSHAFEN A.R.	238	303	303	259	281	8%
LV006	RIGA	DE12B	MUEHLACKER	1481	1907	1860	1734	1797	6%
TRA23	IGDIR	DEE0D	STENDAL	2706	3979	3939	3830	3884,5	2%
								average	3%

The differences between the old and the new ILSE data are more significant. As expected all new distances are systematically longer. The average increase is 24%. As explained above we assume that this increase is mainly due to the more detailed road network geometry. The contribution of the modified routing algorithm to these differences cannot be evaluated, as we could not replicate the calculations on the new network using the old routing software RouteFinder. The software was not able to process the large amount of data of the TomTom© network.

5. Discussion

The comparison between the old and the new ILSE matrix with commercial route finder distances indicates an overall improvement of the accuracy of the distances in the new ILSE matrix compared to the old one. This should lead to a more accurate computation of territorialised TP. As explained above, increased distances will not result in increased total TP used for MSI. As for TP per MS, due to our approach, the main controlling factor on relative changes of territorialised TP between the old and the new ILSE matrix are shifts in national shares. Hence, provided that the absolute increase is fairly proportionally distributed across all countries the application of the new ILSE matrix should not shift the national territorialised TP between countries and the time series of MSI would be preserved.

We therefore investigated potential changes in the shares of national segments in total distances between the old and the new ILSE matrix. We found that not only the absolute distances have changed but also the shares of the national segments. This is due to the fact that for many NUTS3 pairs, the routing software based on the upgraded road network and the fastest path algorithm calculated a different itinerary, often with different border crossings points.

In general terms, modifications may be attributed to a different itinerary with comparable distances or to a different itinerary with increased or in rare cases decreased length. The following examples may illustrate our findings.

Example 1: Comparable distance but different territorial shares

A journey from the Spanish region ES212 (Guipuzacoa, in the Spanish Basque country) to the French region FR81 (Pyrénées-Orientales) can take a route north or south of the Pyrenees. In the old ILSE matrix the distance was 494km, 83% of which were attributed to Spain and 17% to France. This is the southern route. The new ILSE matrix proposes the northern route with a distance of 564km. This represents a moderate increase of 14% which is even below the average increase of 24%. However, now only 4% of the journey is attributed to Spain and 96% to France.

Example 2: Increased distance

The old ILSE matrix indicated for a journey from ITC4A (Cremona in the North of Italy) to PT118 (Alto Tras-os-Montes in the North of Portugal) a distance of 944km, with 15% in IT, 83% in ES and 2% in PT, avoiding France. Hence, a ferry between IT and ES was taken. In the new ILSE matrix the distance is 1884km with 18% in IT, 50% in FR, 31% in ES and 2% in PT. This increase and redistribution can be explained by a purely land based itinerary without ferry use.

Example 3: Ferry bias

Another systematic effect of the application of the new ILSE matrix is the inclusion of the length of the ferry segments. For example, the new matrix increases the length of the journey from AT341 (Bludenz) to FR831 (Corse-du-Sud) from 485km to 828km, including a 190km ferry trip from Italy to Corsica. The old ILSE matrix excluded ferry legs and only included road segments. This inclusion of distances covered by ferries cannot be avoided due to the choice of the fastest-path routing algorithm, which has to take into account all segments of a journey. As data reported by MS only include actual road transport and thus do not include TKM covered by ferries these 190km cannot be attributed to any of the countries, and thus reduce in our example their TP by 1% to 8%. Hence in certain cases the new ILSE matrix may even reduce the accuracy of the territorialisation for certain origin-destination pairs.

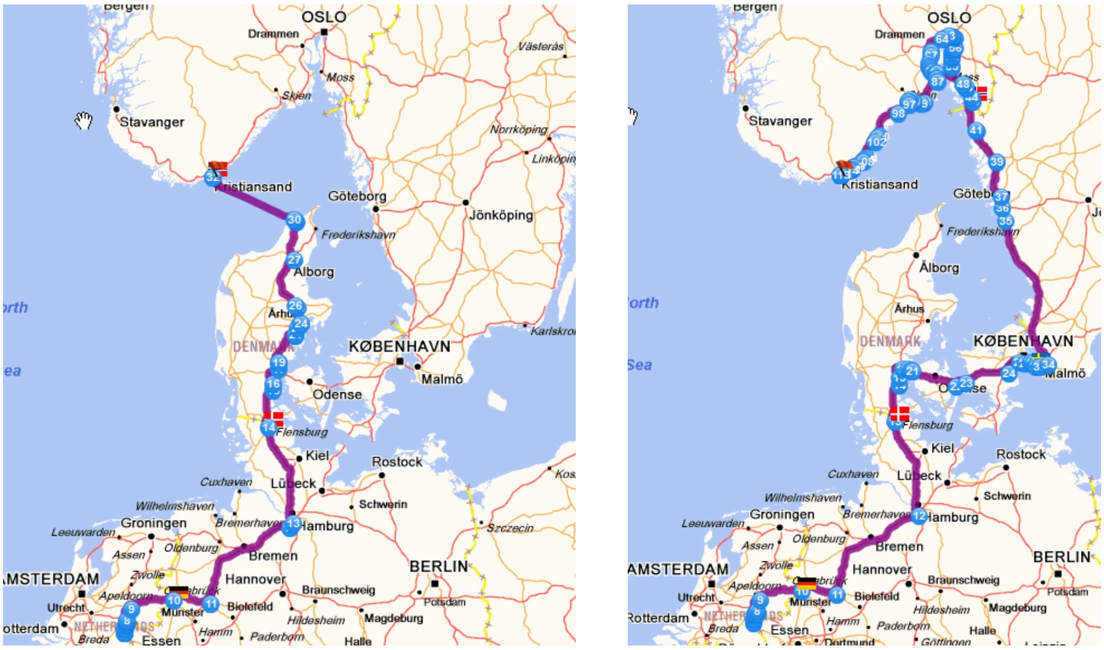

Figure 2: Journey between NL and NO with (left) and without (right) ferry usage.

As already explained above, the new ILSE matrix tends to avoid ferries and favour land based itineraries. This might reflect real patterns like in example 2, but also might be an artefact, e.g. between NL and NO where highly performing ferry connections can cover part of the journey (see Figure 2 with screenshots from ViaMichelin©). The route on the left is recommended by ViaMichelin© and uses a ferry, while forcing the route onto land yields the distance which is contained in the new ILSE matrix. This avoidance of ferries is probably due to the choice of speed parameters in our cost model.

6. Consequences of the application of the new ILSE matrix on MSI

One of the principles of quality in statistics is the comparability of data over a certain period of time [9]. Breaks in time series should be avoided and changes in the data between years should be due to real changes in TP. The new ILSE matrix was first tested on 2011 data. We noted differences in TP that could not be entirely explained by real changes in the trend and assumed that the observed changes can be partially attributed to the new distance matrix and to the shifts between national segments resulting from its application. We therefore considered that the new matrix would lead to a break in the time series that was not acceptable and decided to recalculate all national TP data using the new matrix for the currently disseminated time period (from 2008 onwards). Table 2 shows the relative differences in TP between the old and new ILSE matrix ranging from -28% (CZ 2008, 2010) to +28% (BG, 2008). These substantial changes show that the effort for recalculation of previous data was justified and necessary to rebuild consistent time series.

Table 2: Comparison of TP based on the old and the new ILSE matrix, between 2008 and 2010.

Territory where TKM are performed	2008				2009				2010			
	OLD	NEW	new-old abs	new-old %	OLD	NEW	new-old abs	new-old %	OLD	NEW	new-old abs	new-old %
AT	27 262	26 006	-1256	-4.6	23 023	21 891	-1131	-4.9	24 476	23 582	-894	-3.7
BE	28 861	27 903	-957	-3.3	26 508	25 759	-749	-2.8	26 673	25 973	-700	-2.6
BG	2 042	2 626	584	28.6	2 006	2 432	426	21.2	2 453	2 787	334	13.6
CH	12 307	12 395	88	0.7	11 030	11 090	61	0.5	12 155	11 983	-172	-1.4
CY	:	:			:	:			:	:		
CZ	23 329	16 760	-6570	-28.2	21 239	15 621	-5618	-26.5	24 137	17 432	-6706	-27.8
DE	147 550	151 155	3606	2.4	130 625	134 941	4315	3.3	143 763	148 597	4834	3.4
DK	7 257	7 820	563	7.8	6 247	6 609	362	5.8	5 703	6 282	579	10.2
EE	954	937	-16	-1.7	756	744	-12	-1.6	778	770	-8	-1.0
EL	2 097	1 882	-215	-10.3	1 873	1 675	-198	-10.6	2 424	2 158	-266	-11.0
ES	44 002	40 901	-3101	-7.0	40 295	37 497	-2797	-6.9	40 256	37 478	-2779	-6.9
FI	1 454	1 065	-389	-26.8	1 257	936	-321	-25.5	1 707	1 233	-474	-27.8
FR	109 086	113 921	4836	4.4	94 975	99 346	4370	4.6	99 691	103 878	4186	4.2
HR	4 880	3 716	-1164	-23.8	3 920	3 231	-689	-17.6	3 972	3 137	-834	-21.0
HU	14 747	14 201	-546	-3.7	10 675	10 196	-479	-4.5	10 460	9 984	-476	-4.6
IE	1 551	1 415	-136	-8.7	1 351	1 241	-109	-8.1	1 432	1 239	-193	-13.5
IT	38 244	34 298	-3945	-10.3	33 175	29 266	-3909	-11.8	36 062	32 497	-3566	-9.9
LI	41	4	-37	-90.4	45	3	-42	-93.0	48	3	-45	-92.8
LT	3 028	2 886	-142	-4.7	2 466	2 355	-111	-4.5	3 055	2 748	-307	-10.0
LU	1 947	1 620	-326	-16.8	1 667	1 423	-244	-14.7	1 767	1 520	-248	-14.0
LV	1 565	1 341	-224	-14.3	1 210	1 060	-149	-12.3	1 431	1 153	-279	-19.5
NL	22 896	18 885	-4011	-17.5	20 753	17 197	-3556	-17.1	21 294	17 689	-3605	-16.9
NO	2 978	2 727	-252	-8.4	2 824	2 454	-370	-13.1	2 959	2 645	-314	-10.6
PL	37 169	29 643	-7525	-20.2	33 382	28 079	-5303	-15.9	41 789	34 539	-7250	-17.3
PT	6 329	6 556	227	3.6	6 194	6 442	248	4.0	5 749	6 015	266	4.6
RO	8 517	7 717	-800	-9.4	5 060	4 141	-920	-18.2	4 484	3 523	-962	-21.4
SE	10 095	10 011	-84	-0.8	7 955	7 951	-4	-0.1	9 143	8 649	-494	-5.4
SI	4 897	5 564	667	13.6	4 456	4 823	367	8.2	4 545	4 906	361	7.9
SK	6 834	6 294	-539	-7.9	6 275	6 071	-204	-3.3	6 587	6 498	-89	-1.4
UK	17 169	14 481	-2689	-15.7	14 066	11 896	-2170	-15.4	13 836	12 082	-1755	-12.7

[9] Eurostat (2011), European Statistics Code of Practice.

7. Conclusions

The presented results show that spatial data and spatial analysis can inform the production of statistics. MSI produced by Eurostat are now based on a transparent methodology that is comparable for all MS. This should result in an increased traceability and plausibility of MSI. Similar distances matrices will now be constructed or updated for the other transport modes. However, the effect on time series also shows that using spatial modelling might introduce a significant methodological change that requires a careful impact assessment. If data are to be used for monitoring trends the expected improvement in data quality has to be balanced against the risk of time series breaks. The use of new or more accurate spatial data should be synchronised with other methodological changes. The metadata need to carefully document the quality of geographical data and the spatial aspects of data modelling.