Eur. Transp. Res. Rev. (2015) 7: 18
DOI 10.1007/s12544-015-0167-3

ORIGINAL PAPER

A Data Envelopment Analysis approach for accessibility measures: Simulating operational enhancement scenarios for railway across Europe

F. Rotoli[1] **· E. Navajas Cawood**[1] **· P. Christidis**[1]

Received: 2 September 2014 / Accepted: 17 April 2015 / Published online: 10 May 2015

Abstract
Introduction As well known, infrastructure endowment influences competitiveness of a region since the characteristics of a transport system in terms of capacity, connectivity, speeds, etc. determine the advantages/disadvantages of an area compared to other locations. This article attempts to investigate the potential impacts on rail accessibility across Europe when different possible operational enhancement scenarios are simulated.
Methods The simulations are carried out by means of a combination of the TRANSTOOLS rail network and Traffic Analyst, the post-processing analyses are implemented in Matlab and the results for each zone (at NUTS3 level) are reported both in tabular form and in easy-to-read ArcGIS maps. Several accessibility measures are evaluated including two Data Envelopment Analysis (DEA) approaches aiming to construct a composite index for embracing all the complementary information provided by 'partial' accessibility sub-indicators; to better evaluate and understand the results either sensitivity and robustness analyses are performed for both the aggregate indicators.
Results The outcomes provide insight into where major benefits in terms of accessibility can be expected; in particular the current infrastructure endowment already benefits many regions but improvements in speed could still increase significantly rail accessibility across Europe (mainly outside the core area as in Poland, Bulgaria, Romania, Slovakia, etc.). Furthermore both the proposed global indexes, although associating different 'endogenous' weights to the various sub-indicators, appear to be worthy and robust against uniform random noise.
Conclusions Ultimately the results provide information useful for the prioritization of investment needs; moreover even if the interpretation of the partial accessibility indicators is clear and useful for policy-makers, the evaluation of a composite measure could allow planners not only to compare or fully rank the level of accessibility for different regions but even to control for eventual confusing and/or incomplete results that may appear when using only a partial approach.

Keywords Accessibility · European railway system · Transport simulation and policy

1 Introduction

As stated in the White Paper on Transport [1]: "Infrastructure shapes mobility. No major change in transport will be possible without the support of an adequate network and more intelligence in using it. Overall, transport infrastructure investments have a positive impact on economic growth, create wealth and jobs, and enhance trade, geographical accessibility and the mobility of people. It has to be planned in a way that maximizes positive impact on economic growth and minimizes negative impact on the environment". In this context, ascertained the strong relationship between accessibility and socio-economic regional development, in the last decades the

This article is part of the Topical Collection on Accessibility and Policy Making

✉ F. Rotoli
Francesco.Rotoli@ec.europa.eu

E. Navajas Cawood
elena.navajas-cawood@ec.europa.eu

P. Christidis
Panayotis.Christidis@ec.europa.eu

[1] European Commission, Joint Research Centre – Institute for Prospective Technological Studies (IPTS), Economics of Climate Change, Energy & Transport Unit (ECCET), Edificio EXPO - C/ Inca Garcilaso 3, E-41092 Seville, Spain

topic of impacts on accessibility of transport policies has been widely treated in the scientific literature [2–17] and also in several European projects [18–21]. Beyond general assessment papers [2–7], several authors focused their attention either on particular corridors/areas [8–11] or on a wider scale [12–17].

This article simulates various European-wide scenarios assuming different rail operational enhancements (such as improving speeds) and it attempts to assess qualitatively the potential impacts on railway accessibility across Europe.

Accessibility is a complex concept with various facets; inter alia it could be defined as 'the amount of effort for a person to reach a destination' or 'the number of activities which can be reached from a certain location' [3]. Indicators of accessibility measure the benefits households and firms in an area enjoy from the existence and use of the transport infrastructure relevant for their area [13]. As highlighted by Wegener et al. [13], accessibility could be calculated in function of origin, destination, spatial impedance, type and mode of transport. Our analysis focuses on European regions (origins and destinations) and rail passenger services; the spatial impedance between two regions is assumed equal to the travel time along the minimum path between the zones over the rail network.

The study analyses several accessibility indexes offering complementary information and mainly based on two concepts: travel resistance (cost, time, etc.) and attractiveness of urban agglomerations (depending on variables such as population, employment or gross domestic product). In particular four different indicators (emphasizing different cost or attraction attributes) are explored: index of location, relative efficiency of the network, potential and daily accessibility.

Since the location of each zone could influence the measures of the mentioned sub-indicators (core-periphery patterns), the article also explores two different approaches trying to construct an accessibility composite indicator, thought as a synthetic parameter embracing all the complementary information delivered by the other four *'partial'* accessibility measures. In particular the research examines:

- a Data Envelopment Analysis (DEA) approach as already proposed among others by Martin, Gutiérrez, and Roman [22–25],
- a Benefit of Doubts (BOD) approach widely threated in the scientific literature for creating composite indicators (e.g. [26–37]), but according to the authors' knowledge, never used for synthetic accessibility index.

The Data Envelopment Analysis is a non-parametric method for evaluating the relative efficiency of Decision Making Units; in the last years DEA has been extensively applied in several sectors, and also in transportation [38–44]. The methodology consists in obtaining from the available data an approximation of the "best-practice" frontier by means of the linear programming; efficiency measures are then investigated according to this surface [45–48].

Both the proposed approaches are based on the use of the DEA to create a composite indicator, but the main difference resides on the adopted model: in the first case the sub-indicators are either considered as inputs or outputs according to their characteristics and an output-oriented DEA with variable return of scale (VRS) is adopted to take in account the great heterogeneity among of the various European regions. The BOD approach, instead, as pointed out by Despotis [34], is formally equivalent to the original input oriented and constant-returns-to-scale DEA model presented by Charnes et al. [33], with the sub-indicators representing the different outputs and allocating a single 'dummy input' with value unity to each country.

To test the proposed methodologies and also to better understand the results, sensitivity and robustness analyses are performed on both the aggregate indicators; although associating different weights to the various sub-indicators, both the models appear to be robust against uniform random noise. Furthermore the outcomes of the study show how the current European railway endowment already benefits many regions (mainly in Italy, Spain, Germany, Netherlands, UK, Austria, France, Belgium, etc.) but improvements in speed could still increase the accessibility of various areas.

Regarding the structure of this paper, after this introduction and brief review of the technical literature, the next paragraph describes in detail the proposed methodology and the utilized data while paragraph 3 illustrates the results of the analysis and finally paragraph 4 sets out our conclusions.

2 Data & methodology

This article summarizes the results of the model simulations carried out in order to estimate the potential impacts on rail accessibility of infrastructure enhancements across Europe (see also [17]). The model simulations have been performed by a combination of the TRANS-TOOLS rail network and Traffic Analyst: the TRANS-TOOLS ("Tools for transport forecasting and scenario testing") is a European transport network model that has been developed in collaborative projects funded by the European Commission's Joint Research Centre and DG TREN (for more information see [49]) and it provides a quite detailed European rail network for 2005; its assignment module (Traffic Analyst by Rapidis, see [50]) allows the model to capture changes in route choice as a result of hypothesized changes in speed and frequency.

The data for origins and destinations at NUTS 3 level (the Nomenclature of Units for Territorial Statistics is a geocode standard developed and regulated by the European Union) and for the baseline (2005) have been assumed according to the ETIS Plus figures (downloadable at [51]): ETIS Plus is a FP7 project on data collection for transport at European level aiming at providing a bridge between official statistics and applications within the transport policy theme; in practice it consists of an European Transport Policy Information System, combining data, analytical modelling with maps (GIS), and a single online interface for accessing the data.

The post-processing analyses of the results have been carried out with utilities developed in Matlab, while the outcomes for each zone have been also reported in easy-to-read ArcGIS maps (Fig. 1).

Beyond the baseline (2005), three different scenarios have been implemented by changing speeds as follows:

- *Scenario 200 km/h*: speed increased up to 200 km/h for all links that currently have a speed lower than 200 km/h. For links with current speed higher than 200 km/h (high speed trains), no changes were introduced.
- *Scenario 90 km/h*: speed increased up to 90 km/h for all links that currently have a speed lower than 90 km/h. For links with current speed higher than 90 km/h, no changes were introduced.
- *Scenario 45 km/h*: speed decreased to 45 km/h for all links that currently have a speed higher than 45 km/h. For links with current speed lower than 45 km/h, no changes were introduced.

In practice the article assumes the best and worst (hypothetic) network settings by simulating respectively the Scenario 200 km/h and Scenario 45 km/h; this last one hypothesizes a degradation of the current network (or better an imaginary configuration previous to the baseline), to evaluate the benefits of the current infrastructure endowment compared to this lower bound. Subsequently the analysis estimates the effects of a more feasible and realistic interventions such as increases of speed of some links (with current speed lower than 90 km/h).

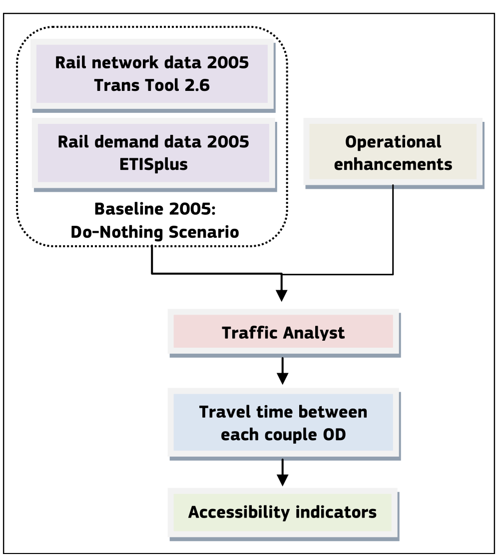

Fig. 1 Flowchart of methodological steps

As also proposed in other studies [18] our analysis considers the centroids of NUTS3 regions as origins and destinations. The all-or-nothing assignment module calculates the minimum paths through the networks, i.e. the path with minimum travel times between the centroids of the regions. According to the new travel times between each couple OD, for each scenario various accessibility indexes have been evaluated.

In the last decades, in fact, accessibility has been measured by several types of indicators often based on two different concepts: travel resistance (cost, time, etc.) and attractiveness of urban agglomerations (depending on variables such as population, employment or gross domestic product). In this analysis we have considered four different accessibility measures offering complementary information and highlighting different cost or attraction attributes.

The location index represents the average travel time between each couple OD weighted on the mass, measured in this analysis by population of the destination regions:

$$Li = \frac{\sum_j t_{ij} * W_j}{\sum_j W_j} \tag{1}$$

where:

- L_i represents the location index of origin i;
- t_{ij} represents the travel time between i and j;
- W_j represents the population of destination j (activities to be reached at j).

Since no distance decay function (and so no discrimination between neighbor or far locations) is considered, the accessibility for each zone depends on the geographical position; remote locations present low accessibility values and even a good transport infrastructure endowment could be not enough to overcome the negative effects of a large geographical distance to the main activity areas [11]. Figure 2 reports the location index for each NUT3 zone in Europe and for each simulated scenario showing clearly the above described core-periphery patterns; as expected the scenarios with maximum speed of 45 km/h or minimum speed of 200 km/h on the whole railway network present respectively the highest and lowest values of the location index.

The network efficiency indicator "represents the distance between the real accessibility against the best accessibility that can be obtained if the zone is connected with all the other regions by the best possible infrastructure" [22], in our case a network with speed on each link of at least 200 km/h.

It offers a measure in terms of the relative ease of access according to the network efficiency; the relative ease of access

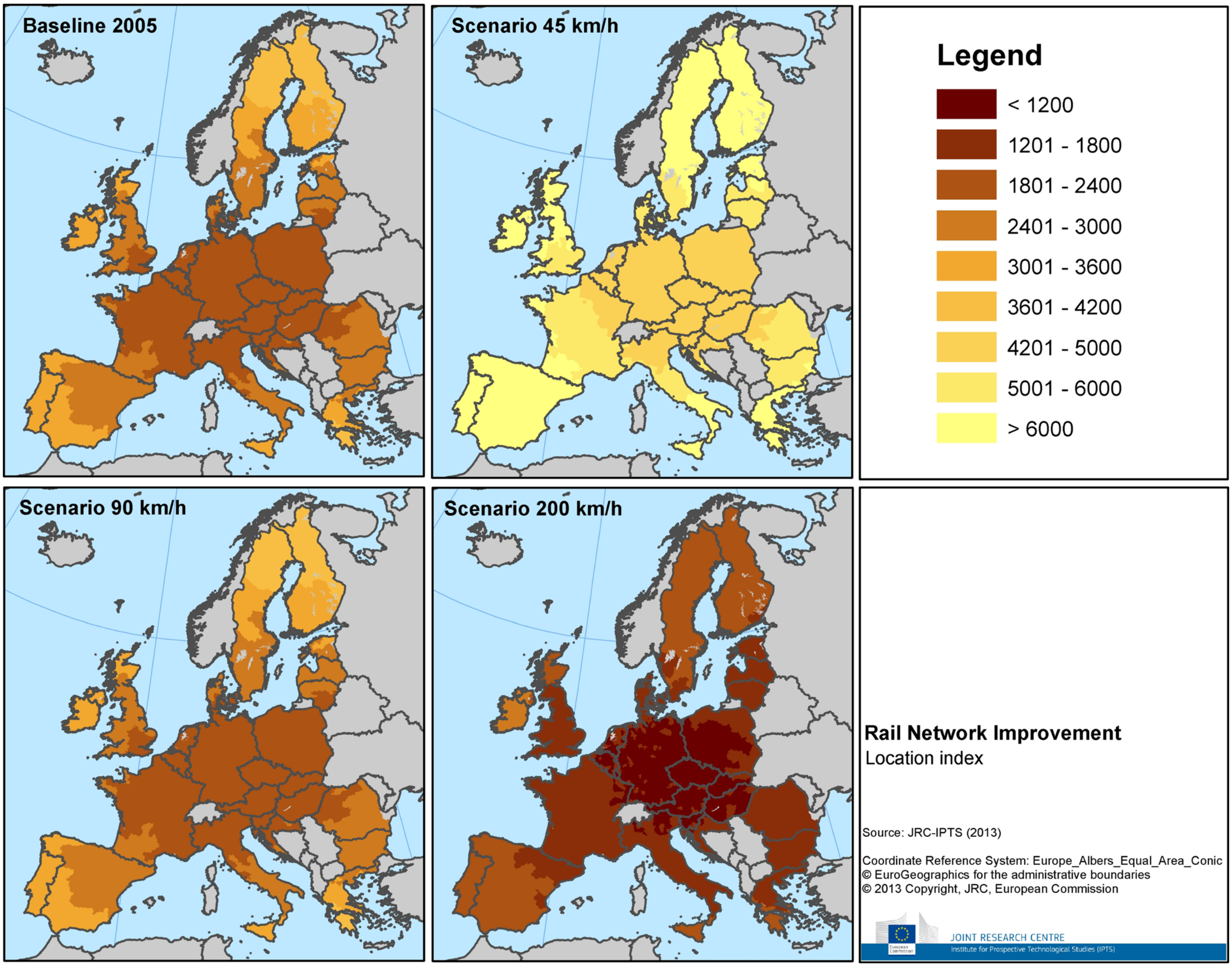

Fig. 2 Location index for each NUTS3 zone and each scenario

is represented by the ratio of the travel time between 2 zones to the ideal travel time between the same zones assuming the best possible infrastructure, i.e. for us scenario 200 km/h):

$$N_i = \frac{\sum_j \frac{t_{ij}}{\underline{t}_{ij}} * W_j}{\sum_j W_j} \tag{2}$$

where:

- N_i represents the network efficiency indicator of origin i;
- t_{ij} represents the travel time between i and j;
- $\underline{t}_{ij}$represents the travel time between i and j assuming the best possible infrastructure (i.e. for us scenario 200 km/h);
- W_j represents the population of destination j.

This indicator provides an idea on how efficient are the connections from a given zone, regardless of its geographic location: it could occur that a region which is peripheral according to the location index is highly accessible in terms of network efficiency [11]. The following figure shows the values of this indicator for each zone and each scenario; of course the scenario with speed of at least 200 km/h on the whole network represents the best possible (ideal) setting ($A=1$ for each region) (Fig. 3).

Finally regarding ***the potential and the daily accessibility***, it is possible to express them as a construct of two functions: the activities function (representing the activities or opportunities to be reached) and the impedance function (representing the effort, time, distance or cost needed to reach them):

$$A_{im} = \sum_j W_j * F(c_{ij}) \tag{3}$$

where:

- A_{im} represents the accessibility of origin i by mode m (i.e. rail in our analysis);
- W_j represents the population of destination j (activities to be reached at j).

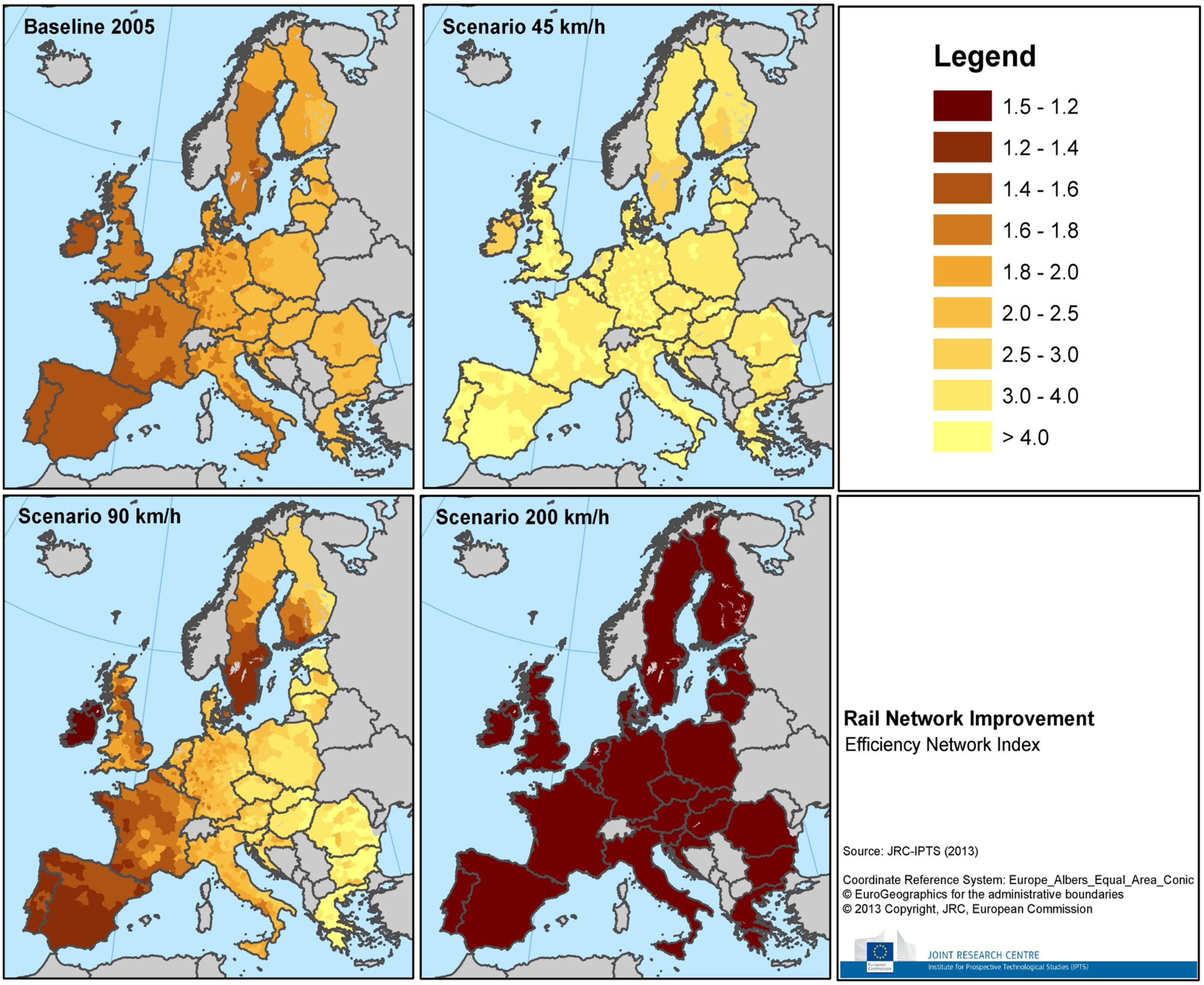

Fig. 3 Efficiency Network Index for each NUTS3 zone and for each scenario

- $F(c_{ij})$ represents the impedance function depending on the generalized cost to reach destination j from origin i;

In practice (3) calculates the total of activities reachable in j weighted by the ease of getting from i to j. As described by the impedance function, the interaction between locations declines with the increasing disutility (distance, time, and/or costs) between them. In general, the perception and valuation of the distance between an origin and a destination differ according to transport modes, purpose of trips, characteristics of the household and of the destination [3]; in the present paper we focused on rail mode and on the population of destination.

Several forms of distance decay function have been already used and described in past accessibility studies; this analysis considers two different shapes depending on travel time (Fig. 4):

- a negative exponential function to represent the ***potential accessibility***:

$$F\left(t_{ij}\right) = e^{-\beta t}; \tag{4}$$

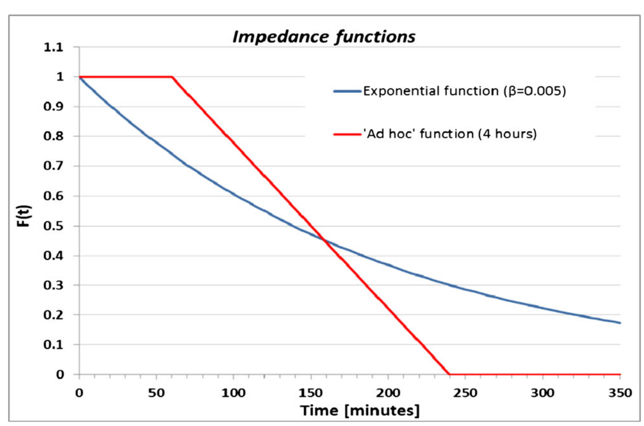

Fig. 4 Considered impedance functions

as proposed also in [18] in our analysis "the parameter β has been set to 0.005. That means that assuming a travel time between two regions of 0 min (which does not occur in reality), the population of the destination region would be included with its full value in the potential accessibility of the origin region, while for a travel time of little more than 2 h the weight is 0.5, and for a travel time of little more than 5 h the weight goes down to 0.2 only" (see Fig. 4).

- an *'ad hoc'* impedance function dropping linearly from 1 to zero with travel times between 1 and 4 h (see Fig. 4) to represent the ***daily accessibility***; this indicator calculates the amount of population or economic activities that can be reached from each zone within a limited period of travel time (in our analysis 4 h), so that it is possible to go and return within the same day and carry out some activities at the destination. The proposed decay function assigns weight 1 to destinations with travel time less than or equal to 1 h and linearly decreasing weights to farther locations with travel time within 4 h. Since the calculations of accessibility have been implemented in Matlab with a post-processing application, it has not been difficult to reproduce the proposed shape.

Of course different shapes of the impedance function could represent diverse aspects (and could provide different indications) of accessibility; the described exponential decay function (potential accessibility) allows to consider the population (activities) of all the reachable zones even if with a diverse weight depending on travel time, while the proposed 'ad hoc' decay function associates the accessibility measures only to short trips with travel time within 4 h (allowing 'daily' commuting) giving a different and more specific indication. The following figures report the potential and daily accessibility values for each NUT3 zone and each considered scenario.

To allow a first analysis of the results, Table 1 reports the percentage of variation (at country level) of all the considered accessibility indicators comparing each alternative scenario and the baseline 2005 (do-nothing configuration); it presents also the average speed on the network (for each country) weighted on the length of the links.

As evident, all the proposed scenarios show a positive impact on the *'partial'* accessibility measures for each country but the differences among of the indicators vary according to the specific area; also the baseline scenario (2005) presents a positive variation in accessibility compared to the hypothetic degraded configuration (maximum speed of 45 km/h on the whole network).

Analyzing the results of the previous table and the Figs. 2, 3, 5 and 6 it is not difficult to recognize the different approach of each indicator, such as for example the core-periphery and border patterns of the location index and of the potential accessibility.

As pointed out in [24], in fact, "the interpretation of the partial accessibility indicators is clear and useful for policy-makers, but it does not provide a synthetic and global measure that allows planners to compare or fully rank the level of accessibility for different regions or cities within Europe. Besides this, it is necessary to control for some contradictory results that may confuse planners who only use a partial approach."

Table 2 reports the Pearson and the Spearman correlation coefficients among of the considered accessibility indexes and for each scenario to assess how well their relationship can be described using respectively a linear or a monotonic function; results seem to justify the assumption that the different *'partial'* indicators can be considered complementary more than substitutive. As expected the potential and the daily accessibility indexes present the strongest correlation; we have retained both the indicators in our successive analyses since, how already noticed, they allow us to measure accessibility from different perspectives (overall or *'within-4-h/daily'* activities reachable from each zone).

To better analyze the accessibility impacts of the simulated operational enhancements and trying to summarize the complementary information provided by the four described indicators, the authors have attempted to explore the construction of a composite indicator (CI) by means of two different approaches:

- Data Envelopment Analysis (DEA) [22–25];
- Benefit of Doubts (BOD) [26–36].

The first methodology is based mainly on the estimation of a DEA-accessibility index as already proposed by Martin, Gutiérrez, Roman and also Reggiani [22–25] while the second method recalls the wide scientific literature for creating composite indicators in analysis at large scale [26–36].

A frequent issue concerning the development of a global index is related to the weights to assign to the partial indicators; in this context, the main appeal of DEA-based composite indicators (CIs) is that they "look for endogenous weights, which maximize the overall score for each decision-making unit given a set of other observations" [31].

Indeed the Data Envelopment Analysis is a non-parametric methodology for evaluating the relative efficiency of Decision Making Units; practically the proposed DEA approach suggests solving the following multiple objective problem of accessibility for both the scenarios with and without interventions:

$$\begin{aligned}&\min_j(LocationAccessibilityIndex) = \min_j(L_j)\\&\min_j(EfficiencyNetworkIndex) = \min_j(N_j)\\&\max_j(PotentialAccessibilityIndex) = \max_j(PA_j)\\&\max_j(PotentialDailyAccessibilityIndex) = \max_j(PDA_j)\end{aligned} \tag{5}$$

Table 1 Percentage of variation in accessibility (alternative scenarios vs baseline) and average speed on the network

Country	Baseline 2005 vs scenario 45 km/h				Scenario 90 km/h VS baseline 2005				Scenario 200 km/h vs Baseline 2005				Average speed on the network weighted on the length of links			
	LA	NA	PA	PDA	LA	NA	PA	PDA	LA	NA	PA	PDA	Scen 45	Scen 2005	Scen 90	Scen 200
Austria	49 %	50 %	353 %	143 %	11 %	10 %	27 %	28 %	50 %	48 %	390 %	584 %	44.6	86.0	94.4	200.0
Belgium	52 %	55 %	434 %	302 %	10 %	8 %	7 %	19 %	47 %	43 %	113 %	439 %	44.8	92.1	101.1	201.1
Bulgaria	46 %	44 %	84 %	28 %	13 %	14 %	53 %	28 %	52 %	53 %	506 %	291 %	44.9	67.3	90.1	200.0
Czech Republic	47 %	48 %	196 %	13 %	13 %	12 %	42 %	29 %	51 %	51 %	449 %	543 %	45.0	70.8	91.2	200.0
Germany	50 %	53 %	414 %	305 %	11 %	9 %	14 %	29 %	49 %	46 %	193 %	425 %	44.7	94.0	105.6	200.2
Denmark	50 %	52 %	236 %	50 %	12 %	11 %	22 %	8 %	47 %	45 %	285 %	103 %	44.6	83.0	98.7	200.0
Estonia	44 %	43 %	41 %	9 %	16 %	16 %	39 %	12 %	54 %	55 %	631 %	56 %	43.8	62.0	90.0	200.0
Spain	60 %	62 %	433 %	81 %	7 %	5 %	1 %	0 %	39 %	34 %	65 %	45 %	45.0	147.1	147.9	201.9
Finland	47 %	46 %	74 %	4 %	19 %	17 %	7 %	1 %	48 %	49 %	228 %	20 %	45.0	92.0	96.9	200.0
France	54 %	58 %	740 %	189 %	8 %	6 %	2 %	1 %	44 %	39 %	94 %	102 %	44.4	140.9	142.2	204.9
Greece	46 %	45 %	96 %	48 %	13 %	13 %	32 %	18 %	54 %	54 %	418 %	206 %	45.0	77.8	94.8	200.0
Croatia	47 %	47 %	198 %	22 %	12 %	12 %	47 %	18 %	50 %	50 %	656 %	258 %	44.7	83.3	98.7	200.0
Hungary	47 %	47 %	130 %	62 %	12 %	13 %	48 %	34 %	51 %	52 %	535 %	484 %	44.6	67.4	90.2	200.0
Ireland	50 %	50 %	69 %	25 %	21 %	22 %	17 %	2 %	38 %	36 %	73 %	123 %	45.0	80.3	93.8	200.0
Italy	52 %	54 %	280 %	166 %	9 %	8 %	9 %	15 %	48 %	45 %	176 %	249 %	44.9	101.6	107.7	200.1
Lithuania	45 %	44 %	81 %	9 %	16 %	16 %	61 %	17 %	53 %	53 %	674 %	129 %	44.4	77.4	90.7	200.0
Luxembourg	51 %	55 %	662 %	177 %	10 %	8 %	8 %	17 %	48 %	45 %	151 %	798 %	45.0	78.2	91.4	200.0
Latvia	45 %	44 %	76 %	0 %	15 %	15 %	26 %	1 %	53 %	53 %	489 %	54 %	45.0	80.1	97.1	200.0
Netherlands	51 %	53 %	289 %	202 %	10 %	8 %	8 %	16 %	48 %	45 %	159 %	395 %	44.8	87.3	94.1	200.0
Poland	47 %	46 %	124 %	19 %	13 %	14 %	49 %	26 %	51 %	52 %	454 %	322 %	44.7	69.4	92.3	200.4
Portugal	61 %	63 %	417 %	226 %	6 %	4 %	0 %	0 %	38 %	34 %	56 %	57 %	45.0	138.3	139.1	200.0
Romania	46 %	44 %	99 %	15 %	12 %	13 %	33 %	8 %	52 %	53 %	425 %	132 %	45.0	75.2	90.9	200.0
Sweden	50 %	51 %	133 %	0 %	14 %	13 %	16 %	0 %	41 %	40 %	174 %	4 %	43.8	91.6	104.4	200.0
Slovenia	48 %	49 %	256 %	111 %	13 %	12 %	48 %	93 %	51 %	51 %	548 %	707 %	44.0	68.7	90.7	200.0
Slovak Republic	47 %	47 %	134 %	38 %	13 %	13 %	50 %	23 %	52 %	52 %	518 %	365 %	44.5	68.7	90.4	200.0
United Kingdom	55 %	58 %	267 %	229 %	8 %	7 %	7 %	9 %	44 %	41 %	125 %	246 %	44.7	96.6	103.5	200.0

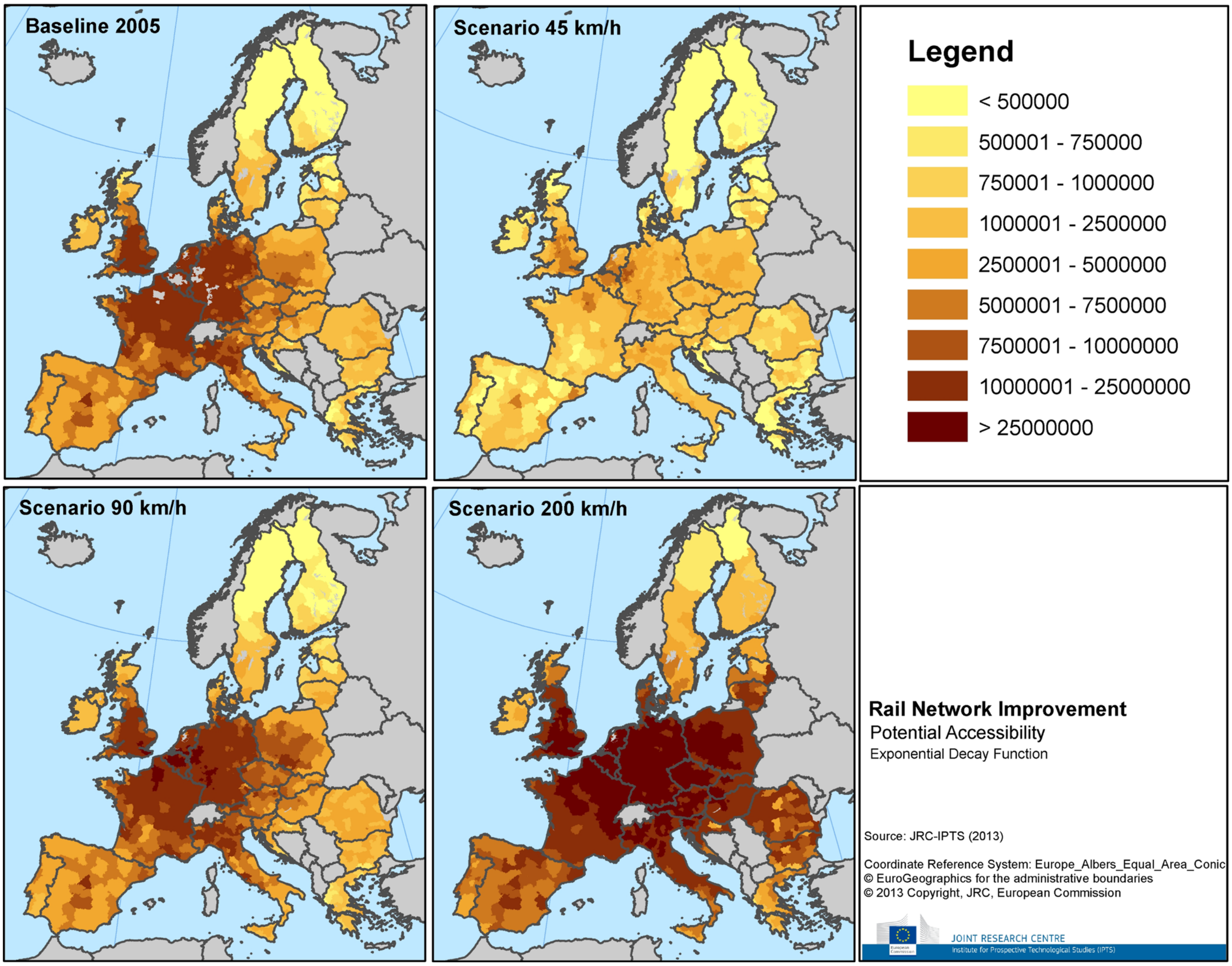

Fig. 5 Potential accessibility for each NUT3 zone and for each scenario

where j represents the generic European NUT3 zone and L, N, PA and PDA represent respectively the location index, the network efficiency indicator, the potential and the daily accessibility.

In other words this "synthetic DEA-based indicator is calculated as the inverse of the maximum proportional accessibility outputs that can be obtained for the indicated accessibility inputs" [10].

The model determines for each scenario the most efficient zones (from an accessibility perspective) to individuate the frontier of the envelopment surface; the regions not lying on the frontier are inefficient and the measurement of the grade of inefficiency is represented by their distance from this 'best-practice' frontier.

When the data cannot be easily interpreted as inputs or outputs, a general rule suggests to consider the variables for which lower levels are better as inputs (in our case Location index and Network efficiency), and to treat as outputs those variables for which higher amounts are better (potential and daily accessibility in our analysis) [10].

We have assumed variable returns to scale (VRS) because of the great heterogeneity among of the various EU regions and an output orientation.

The Benefit of Doubts approach, instead, is rooted in the copious literature concerning CIs. Considering *m* sub-indexes and *n* regions the idea is to merge the sub-indicators' values per region into a single number, e.g. their weighted average. In absence of reliable information about the weights, the proposed approach endogenously determines the weights maximizing the CI value for each region. In practice it comes to solving the following linear programming problem for each zone j:

$$\begin{aligned} &CI_j = \sum\nolimits_i y_{ij} * w_{ij} \\ &\sum\nolimits_i y_{ij} * w_{ij} \le 1 \, (bounding\ constraint) \\ &w_{ij} \ge 0 \ (non{-}negativity\ constraint) \end{aligned} \tag{6}$$

where:

- CI_j represents the value of the composite indicator for region j;

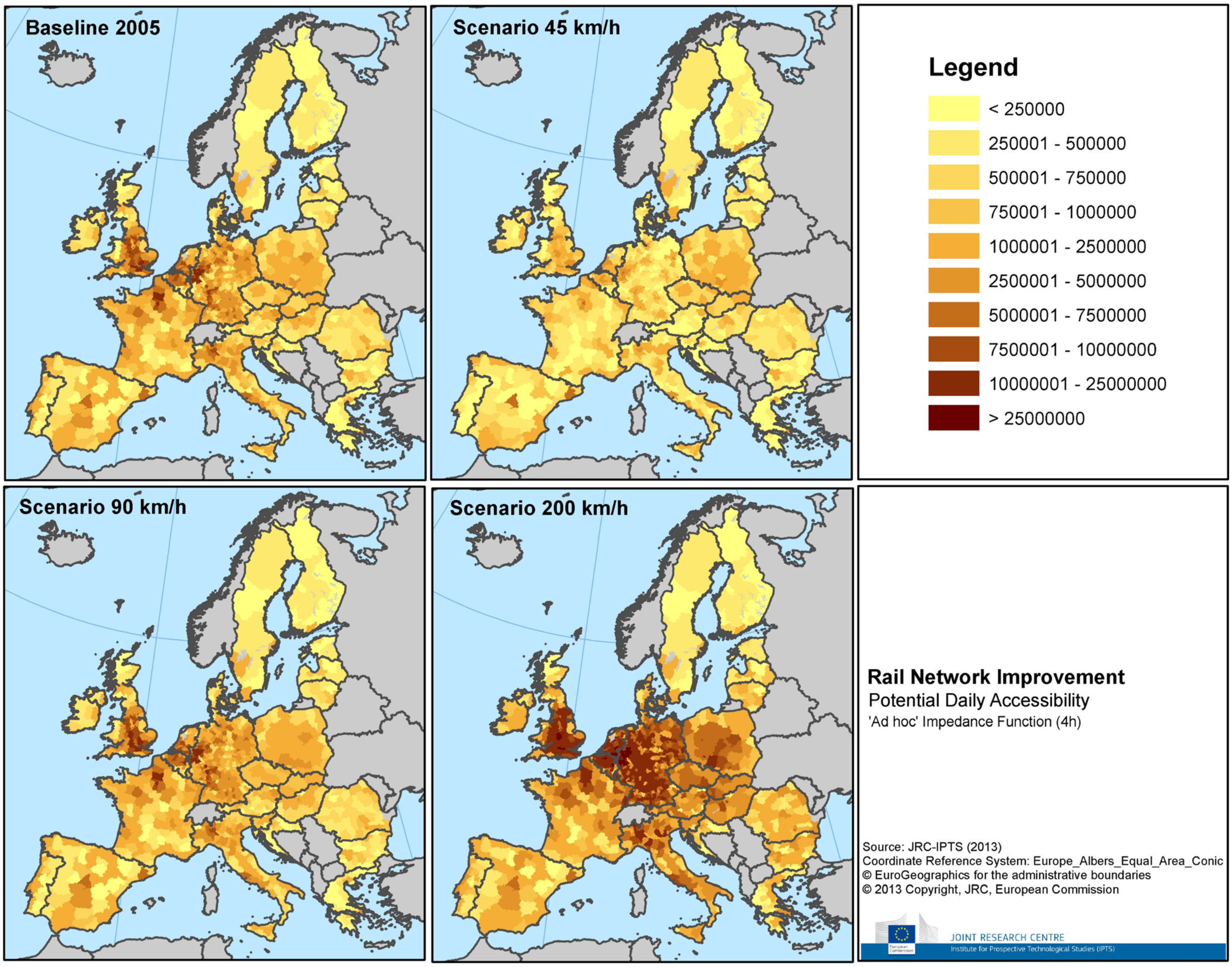

Fig. 6 Potential daily accessibility for each NUTS3 zone and for each scenario

- y_{ij} represents the original value of sub-indicator i for region j;
- w_{ij} represents the weight of sub-indicator i for region j;

The benefit-of-the-doubt interpretation of the methodology resides in the fact that the method chooses those weights maximizing the resulting indicator: the highest relative weights are accorded to those sub-indicators for which the zone j achieves the best performance (in relative terms) when compared to the other regions in the sample.

This model is equivalent to the original input oriented, constant-returns-to-scale DEA model presented by Charnes et al. [33] (for more details see [29]), with the sub-indicators representing the different outputs and a single 'dummy input' with value unity to each country.

Although, as pointed out in [30], a BOD-based indicator meets the property of units invariance, which makes the normalization stage redundant, the selected four sub-indicators have not the same direction, meaning that for some of them (location index and network efficiency) higher values represent worse performances. To overcome this issue the "min-max" normalization method has been applied to the results of the *partial* indexes according to the original direction of the variables:

$$I_{ij} = \frac{y_{ij} - \min_j(y_i)}{\max_j\left(y_{ij}\right) - \min_j(y_i)} \tag{7}$$

$$I_{ij} = \frac{\max_j(y_i) - y_{ij}}{\max_j\left(y_{ij}\right) - \min_j(y_i)} \tag{8}$$

where:

- i and j represent respectively the sub- indicator and the region
- I_{ij} represents the normalized value of sub-indicator i in region j;
- y_{ij}represents the original value of sub-indicator i in region j;

Table 2 Pearson correlation and Spearman rank correlation (* for values with significance levels below 0.05)

		Scenario 45				Baseline 2005				Scenario 90				Scenario 200			
		LA	NA	PA	PDA	LA	NA	PA	PDA	LA	NA	PA	PDA	LA	NA	PA	PDA
Pearson	LA	1.0000				1.0000				1.0000				1.0000			
	NA	0.0180	1.0000			−0.0232	1.0000			−0.2456*	1.0000			n/a	–		
	PA	−0.3761*	0.2732*	1.0000		−0.6422*	−0.3144*	1.0000		−0.6521*	−0.1209*	1.0000		−0.7664*	n/a	–	
	PDA	−0.1300*	0.2375*	0.8152*	1.0000	−0.3872*	−0.2372*	0.8407*	1.0000	−0.3853*	−0.1130*	0.8401*	1.0000	−0.4661*	n/a	0.8472*	–
Spearman	LA	1.0000				1.0000				1.0000				1.0000			
	NA	0.0784*	1.0000			−0.1027*	1.0000			−0.2230*	1.0000			n/a	–		
	PA	−0.5011*	0.2527*	1.0000		−0.7782*	−0.3418*	1.0000		−0.7537*	−0.2206*	1.0000		−0.8193*	n/a	–	
	PDA	−0.1986*	0.3199*	0.7540*	1.0000	−0.5974*	−0.2708*	0.8395*	1.0000	−0.5973*	−0.1520*	0.8547*	1.0000	−0.6600*	n/a	0.9126*	–

To better compare and also to test the two proposed approaches, sensitivity and robustness analyses are performed on both the aggregate indicators for the baseline 2005 (do-nothing scenario). We have carried out the Data Envelopment Analysis several times eliminating the sub-indicators one by one, to evaluate the weight, the importance associated to each dimension in both the procedures and also to estimate the changes in scores.

Tables 3, 4, 5 and 6 report the Spearman correlation coefficients across the sub-indicators (eliminating one of them per time) and across the scenarios respectively applying the DEA or the BOD approach. In particular:

- *Scenario 45* (or *45*), *Baseline 2005* (or *2005*), *Scenario 90* (or *90*) and *Scenario 200* (or *200*) indicate the simulated scenarios;
- *-LA* indicates the DEA or BOD approach carried out neglecting the location index;
- *-NA* indicates the DEA or BOD approach carried out neglecting the network efficiency;
- *-PA* indicates the DEA or BOD approach carried out neglecting the potential accessibility;
- *-PDA* indicates the DEA or BOD approach carried out neglecting the potential daily accessibility;
- *All* indicates the DEA or BOD approach carried out considering all the sub-indicators

The tables point out the different set of “endogenous” weights (and the consequent relevance) associated to the sub-indicators applying the two methods and also how they vary across the scenarios; while for the DEA the most relevant indicator comes to be the potential accessibility, for the BOD procedure the major changes in results are obtained eliminating the location index. This is due to the different approach of the methods; as noticed by Martin and Reggiani [24], the DEA results “can be interpreted as the distance that separates every region from the most accessible location that can be found taking into account all the observations”. In other words this “indicator measures how accessible a region is with respect to all the other regions included in the sample”. The BOD approach, instead, maximizes the CI value (see formula (6)) for each region and so it takes in account also the heterogeneity of the sub-indicators among of the regions: the highest relative weights are accorded to those dimensions for which the region j achieves the best performance (in relative terms) when compared to the other regions in the sample; in other words “each region has its own weights which are optimal and guarantee the best possible position for the associated region among all other regions in the sample. With any other weighting scheme, the relative position of that region would be worse.”[35].

Table 3 Spearman rank correlation across indicators applying the DEA approach

Scenario 45	45	- LA	- NA	- PA	- PDA
- LA	0.977	-			
- NA	0.993	0.974	-		
- PA	0.826	0.808	0.805	-	
- PDA	0.996	0.972	0.994	0.802	-
Baseline 2005	2005	- LA	- NA	- PA	- PDA
- LA	1.000	-			
- NA	1.000	1.000	-		
- PA	0.847	0.847	0.847	-	
- PDA	0.999	0.999	0.999	0.840	-
Scenario 90	2005	- LA	- NA	- PA	- PDA
- LA	1.000	-			
- NA	0.973	0.973	-		
- PA	0.856	0.856	0.819	-	
- PDA	0.999	0.998	0.974	0.849	-
Scenario 200	200	- LA	- NA	- PA	- PDA
- LA	0.998	-			
- NA	1.000	0.998	-		
- PA	0.998	0.996	0.998	-	
- PDA	0.925	0.922	0.925	0.918	-

The main disadvantage of this last method is that without additional setting constraints countries performing very well only in one indicator can be considered successful. This is why the scientific literature has focused on different ways to take in account experts' (subjective) opinions or to set additional relative weight constrictions (e.g. pie share constraints). Anyway we have explored the original BOD approach and so no additional constraints are added except those already described in (6). The described effect is quite evident in Tables 5 and 6 where the Scenario 200 is less relevant than the others since the network efficiency index presents unity value for each region (reference "best-possible" infrastructure).

Finally a robustness analysis has been performed on both the synthetic indicators adding a uniform casual noise to the normalized value of each sub-indicator, according to the formula:

Table 4 Spearman rank correlation across scenarios applying the DEA approach

ALL	2005	45	90	200
45	0.872	-		
90	0.971	0.872	-	
200	0.909	0.860	0.893	-
-LA	2005	45	90	200
45	0.881	-		
90	0.971	0.882	-	
200	0.916	0.832	0.900	-
-NA	2005	45	90	200
45	0.890	-		
90	0.992	0.900	-	
200	0.909	0.878	0.942	-
-PA	2005	45	90	200
45	0.828	-		
90	0.969	0.808	-	
200	0.742	0.576	0.728	-
-PDA	2005	45	90	200
45	0.882	-		
90	0.970	0.876	-	
200	0.871	0.921	0.865	-

Table 5 Spearman rank correlation across indicators applying the BoD approach

Scenario 45	45	- LA	- NA	- PA	- PDA
- LA	0.481	-			
- NA	0.956	0.379	-		
- PA	0.994	0.444	0.951	-	
- PDA	0.993	0.463	0.957	0.989	-
Baseline 2005	2005	- LA	- NA	- PA	- PDA
- LA	0.558	-			
- NA	0.766	0.101	-		
- PA	1.000	0.558	0.766	-	
- PDA	0.997	0.546	0.769	0.997	-
Scenario 90	2005	- LA	- NA	- PA	- PDA
- LA	0.474	-			
- NA	0.906	0.224	-		
- PA	1.000	0.474	0.907	-	
- PDA	0.998	0.466	0.907	0.998	-
Scenario 200	200	- LA	- NA	- PA	- PDA
- LA	n/a				
- NA	n/a	n/a			
- PA	n/a	n/a	n/a		
- PDA	n/a	n/a	n/a	n/a	-

$$I_{ij}^{k} = I_{ij} + Unif(-\alpha, \alpha)*\underline{I_i} \tag{9}$$

where

- I_{ij}^{k} represents the normalized value of sub-indicator i in region j and for the casual extraction k;
- I_{ij} represents the normalized value of sub-indicator i in region j;
- $Unif(-\alpha,\alpha)$ represents a uniform casual distribution with values between $-\alpha$ and α. The value of α has been set to 0.05 (5 %) in our analysis.
- $\underline{I_i}$ represents the average of the normalized values of sub-indicator i across the regions;

Table 6 Spearman rank correlation across scenarios applying the BoD approach

ALL	2005	45	90	200
45	0.570	-		
90	0.848	0.845	-	
200	n/a	n/a	n/a	-
-LA	2005	45	90	200
45	0.322	-		
90	0.935	0.520	-	
200	n/a	n/a	n/a	-
-NA	2005	45	90	200
45	0.934	-		
90	0.988	0.964	-	
200	0.937	0.972	0.970	-
-PA	2005	45	90	200
45	0.553	-		
90	0.848	0.832	-	
200	n/a	n/a	n/a	-
-PDA	2005	45	90	200
45	0.569	-		
90	0.852	0.835	-	
200	n/a	n/a	n/a	-

Table 7 Mean and standard deviation of the variation coefficient

Variation coefficient	DEA approach	BOD approach
Mean	0.015	0.017
Standard deviation	0.046	0.070

In practice we have performed a Monte Carlo simulation by extracting 100 random values (within the range defined) of noise for each sub-indicator and assessing the related variation in outcomes. To evaluate the dispersion of the results, for each region we have calculated the variation coefficient, defined as the ratio of the standard deviation to the mean:

$$C_{Vj} = \frac{\sigma_j}{\mu_j} \tag{10}$$

where

- μ_i is the mean of the 100 simulated scores (in terms of ranking) of the composite indicator for the region j;
- σ_j represents the standard deviation of the 100 simulated scores (in terms of ranking) of the composite indicator for the region j;

Table 7 reports the mean and the standard deviation across all regions of the variation coefficient C_{Vj}. Both the approaches seem to be robust against the assumed uniform casual noise: by hypothesizing a value of α equal to 5 % the results show a variation coefficient of less than 2 % in both the cases.

3 Results

As reported below the outcomes of this study show how the current scenario already benefits many regions but improvements in speed could still increase significantly rail accessibility across Europe. For example the above reported Table 1 indicates that the current infrastructure endowment (baseline

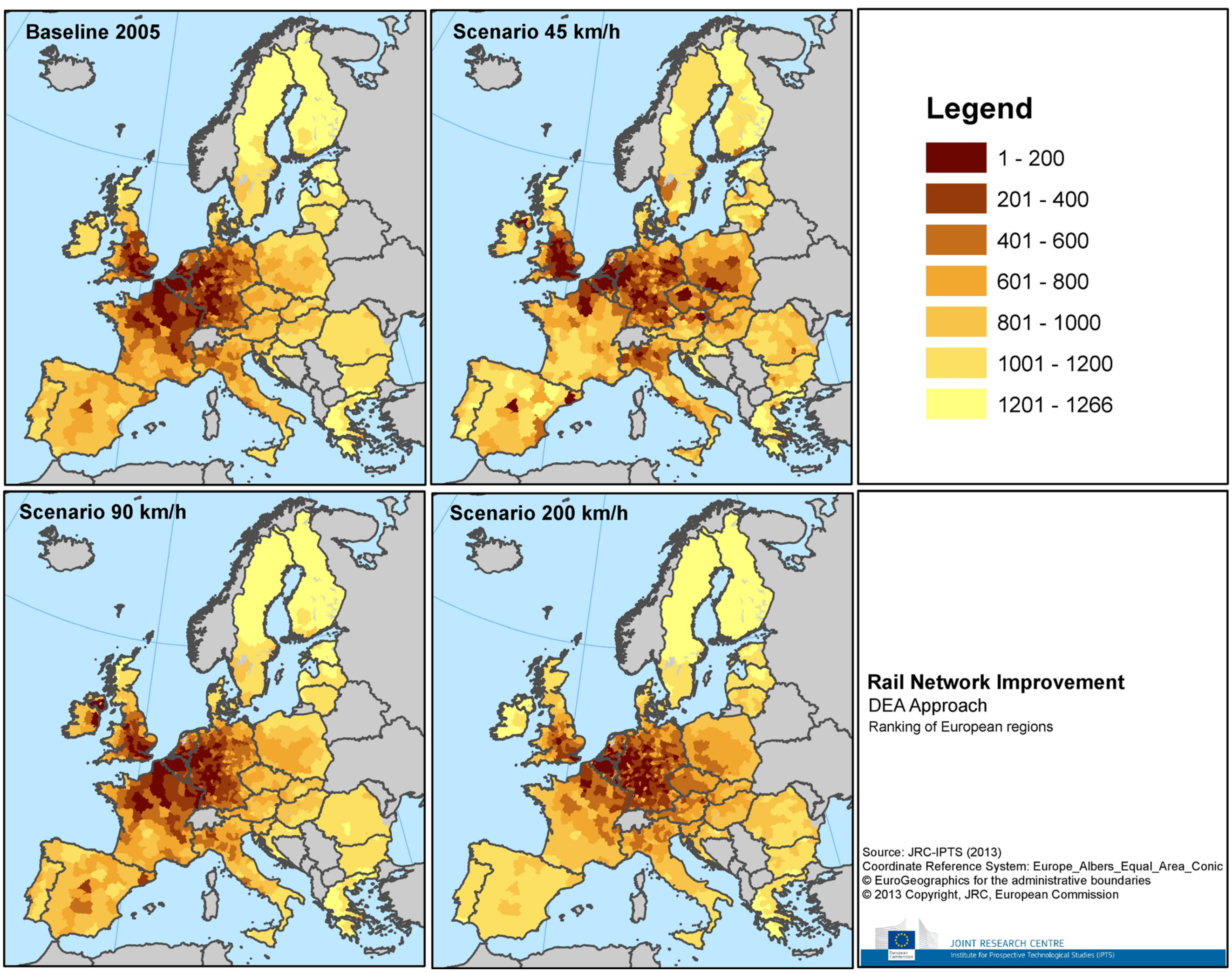

Fig. 7 DEA-based ranking of accessibility across Europe and for each scenario

2005 versus scenario 45) benefits in terms of accessibility mainly the countries in the European core (such as Austria, Germany, Belgium, Luxembourg, Netherlands, Italy, United Kingdom) and even the Iberian Peninsula and France.

The same outcomes are also evident in Figs. 7, 8, 9 and 10, reporting the rankings of accessibility (and the variation in it) across Europe and for all the scenarios. At this stage is worthy to notice that, as highlighted in the scientific literature, both the proposed composite indicators do not provide scores, but ranks; in other words, either DEA or BOD provides values but they are directly incomparable, and this explains why the results are reported in relative rankings. In particular Figs. 9 and 10 show the variation in ranking of accessibility for each scenario versus the baseline 2005; a negative value means that a region is better ranked when moving from the baseline towards the assumed scenario.

Table 1 and the next figures show also how slight improvements in speed (at least 90 km/h for all the links) could benefit mainly peripheral and border regions (in Bulgaria, Romania, Poland, Greece, Hungary, Slovakia, etc.) and how eastern regions would gain accessibility advantages by ideally improving the speed on the current railway network at least at 200 km/h; obviously, although technically not feasible (or hardly achievable), this hypothesis tries only to represent an extreme *'optimum'* situation, an upper limit for the rail network (as well as the scenario 45 km/h represents a lower bound) with which comparing more plausible and less ambitious interventions.

Moreover, analyzing the DEA and BOD maps reported below, it seems quite reasonable that they embrace all the different information provided by the *'partial'* accessibility measures, even if with a different set of endogenous weights, as already noticed in the previous paragraph.

Looking for example at the maps for the baseline (2005) in Figs. 7 and 8, it is not difficult to recognize the outcomes already observed in Tables 3, 4, 5 and 6: while the DEA index is mainly influenced by the potential accessibility, the BOD ranking takes more in account the location index and the

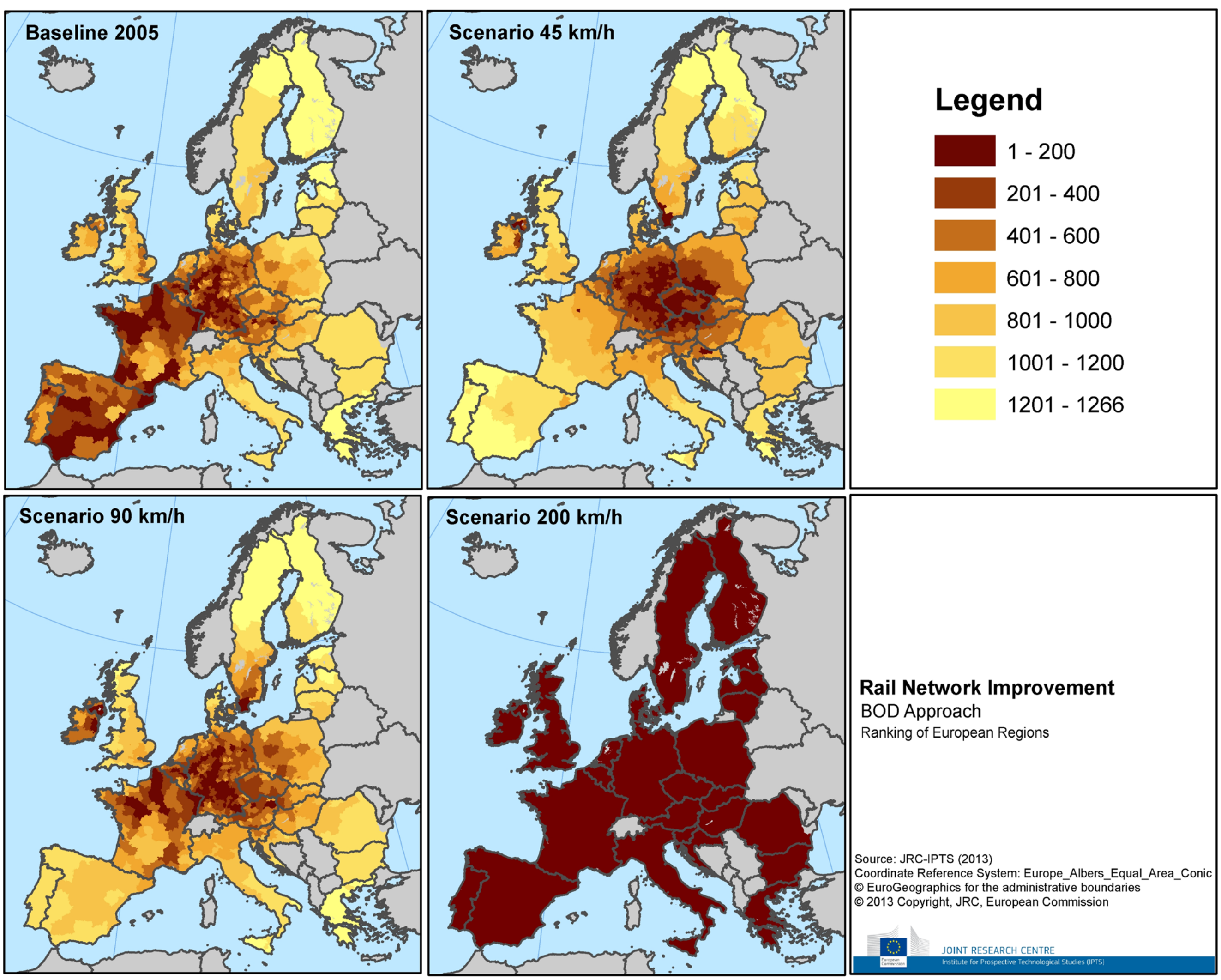

Fig. 8 BOD-based ranking of accessibility across Europe and for each scenario

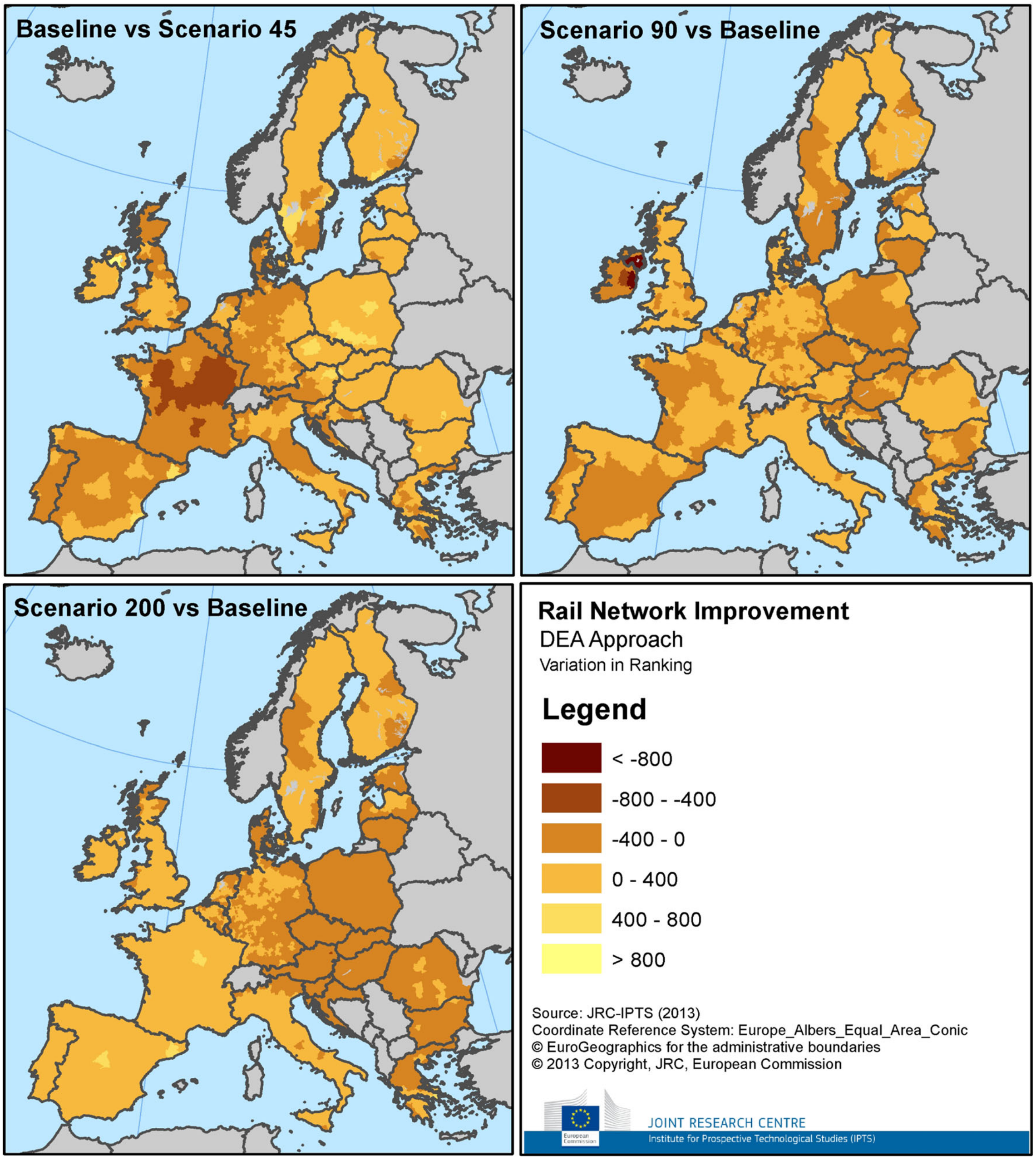

Fig. 9 DEA-based variation in ranking of accessibility for each scenario vs the baseline 2005

network efficiency. This difference affects also Figs. 9 and 10: the BOD-based variation in ranking between the baseline and the degraded scenario 45 for the Iberian Peninsula, for example, is bigger than the corresponding variation applying the DEA approach.

Anyway in the authors' point of view both the methods are worthy and as already mentioned above they only present a different endogenous weighting scheme: while in the DEA approach the regions with highest values are considered the most efficient ones in terms of accessibility since they present the shortest distances from the "*best-practice*" frontier, the BOD attempts to be sensible to national policy priorities, meaning that the set of optimal weights for each region guarantees its best position against all other zones in the sample, however the index is calculated so that regions with specializations in a particular component are not penalized for this.

4 Conclusions

As already highlighted above, this article has tried to explore the impacts of improvements of the European railway infrastructures in order to evaluate how these could potentially

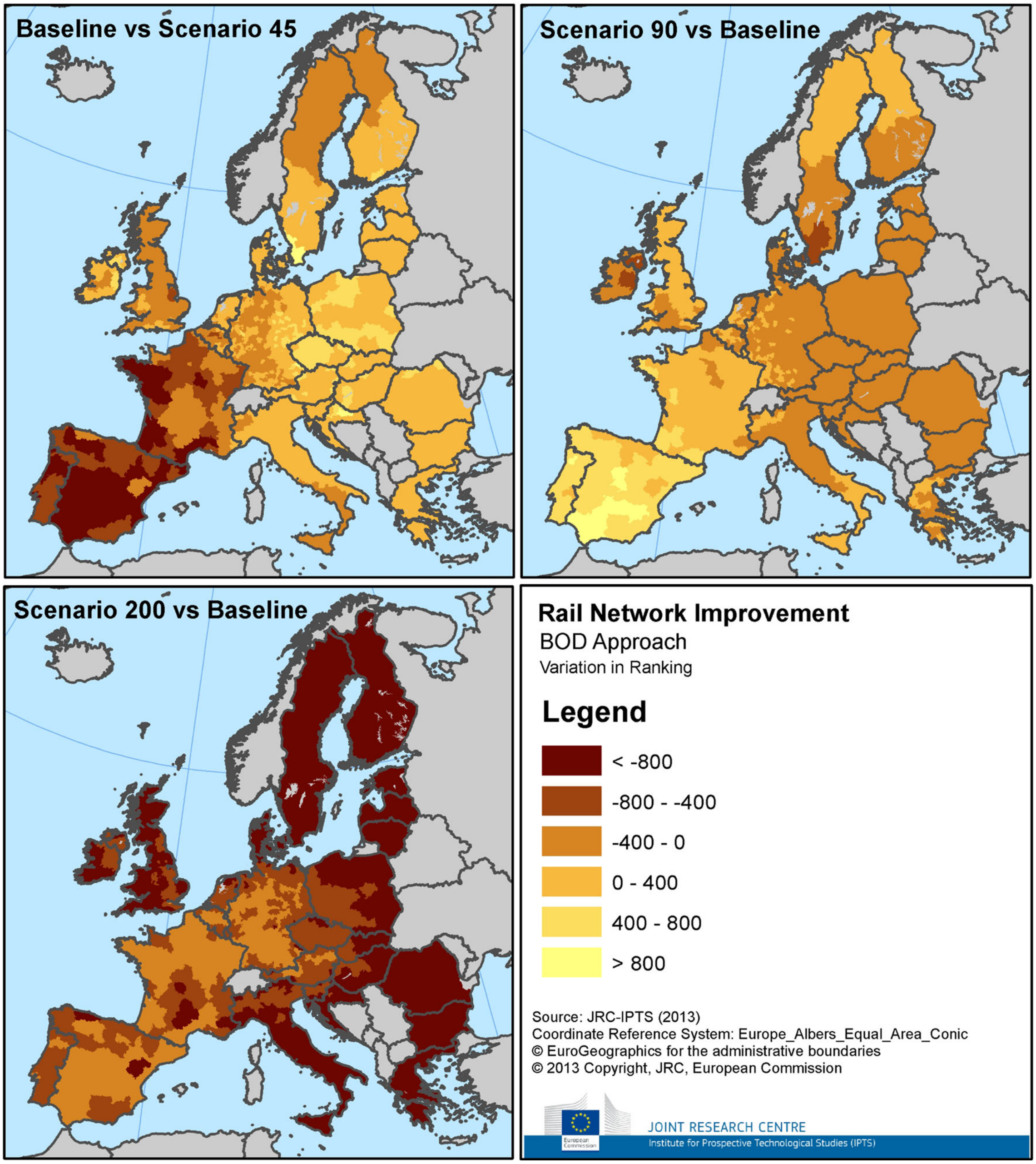

Fig. 10 BOD-based variation in ranking of accessibility for each scenario vs the baseline 2005

increase accessibility for passengers across European regions; it summarizes the results of the model simulations carried out with a combination of the TRANSTOOLS rail network and its assignment module (Traffic Analyst).

Beyond the baseline (year 2005), three different scenarios have been tested: two scenarios simulating increases of speed at least up to 90 and 200 km/h on the whole rail network and the last one assuming a decrease down to 45 km/h.

Furthermore the study has considered four accessibility indicators providing different and complementary information: location index, relative efficiency of the network, potential and daily accessibility. Since the results suggest that the location of each zone (core-periphery patterns) could influence the measures of the various accessibility indicators, the article has also evaluated two different approaches (DEA and BOD) to create a composite accessibility index embracing all the complementary information delivered by the other four *'partial'* accessibility measures. In fact, as already pointed out among others by Martin and Reggiani [24], despite the clear and useful results provided by the sub-indicators, a composite measure could allow policy makers to control for eventual confusing and/or incomplete scenarios which may appear when using only a partial approach. Sensitivity and robustness analyses have been carried out to test the proposed methodologies and to better analyze the results.

The outcomes provide insight into where major benefits in terms of accessibility can be expected; this information, in turn, could also be useful for the prioritization of investment needs. In particular the results of the study show how the current European railway infrastructure already benefits many regions (mainly in Italy, Spain, Germany, Netherlands, UK, Austria, France, Belgium, etc.) but improvements in speed could still increase the accessibility of various areas (mainly outside the core) of Europe (as in Poland, Bulgaria, Romania, Slovakia, etc.).

Moreover the proposed composite indicators appear worthy and robust against casual noise. Their biggest advantage is represented by their different weighting schemes: the weights of the partial indicators do not need to be fixed 'a priori' but they are endogenously derived by the methodology.

Disclaimer The views expressed are purely those of the authors and may not in any circumstances be regarded as stating an official position of the European Commission.

References

1. EUROPEAN COMMISSION (2011) WHITE PAPER. Roadmap to a single European transport area –towards a competitive and resource efficient transport system
2. Schürmann C, Spiekermann K, Wegener M (1997) Accessibility indicators. Berichte aus dem Institut für Raumplanung 39. Dortmund: IRPUD
3. Geurs KT, Ritsema van Eck JR (2001) Accessibility measures: review and applications. RIVM report 408505 006. National Institute of Public Health and the Environment, Bilthoven
4. Geurs KTV, Wee B (2004) Accessibility evaluation of land-use and transport strategies: review and research directions. J Transp Geogr 12:127–140
5. Gutiérrez J, Monzón A, Piñero JM (1998) Accessibility, network efficiency, and transport infrastructure planning. Environ Plan A 30:1337–1350
6. Morris JM, Dumble PL, Wigan MR (1978) Accessibility indicators for transport planning. Transp Res 13A:91–109
7. Van Wee B, Geurs K (2011) Discussing equity and social exclusion in accessibility evaluations. Eur J Transp Infrastruct Res 11(4):350–367
8. Gutiérrez J (2001) Location, economic potential and daily accessibility: an analysis of the accessibility impact of the high-speed line Madrid-Barcelona-French border. J Transp Geogr 9:229–242
9. Martin JC, Gutiérrez J, Roman C (2006) Evaluating accessibility gains produced by new high-speed train services. In: Jourquin B, Rietveld P, Westin K (ed) Towards better performing transport networks, part III. Routledge Studies in Business Organizations and Networks
10. Salze P, Banos A, Oppert J-M, Charreire H, Casey R, Simon C, Chaix B, Badariotti D, Weber C (2011) Estimating spatial accessibility to facilities on the regional scale: an extended commuting-based interaction potential model. Int J Health Geogr
11. López E, Gutiérrez J, Gómez G (2008) Measuring regional cohesion effects of large-scale transport infrastructure investments: an accessibility approach. Eur Plan Stud 16(2):277–301
12. Spiekermann K, Neubauer J (2002) European accessibility and peripherality: concepts, models and indicators. Nordregio working paper. Nordregio, Stockholm, p 9
13. Wegener M, Eskelinnen H, Fürst F, Schürmann C, Spiekermann K (2002) Criteria for the Spatial Differentiation of the EU Territory: Geographical Position. Forschungen 102.2, Bonn: Bundesamt für Bauwesen und Raumordnung.
14. Linneker B (1997) Transport infrastructure and regional economic development in Europe: a review of theoretical and methodological approaches, TRP 133. Department of Town and Regional Planning, Sheffield
15. Vickerman RW (1995) The regional impacts of Trans-European networks. Ann Reg Sci 29
16. Martin JC, Gutiérrez J, Roman C (2007) Accessibility impacts of the Trans-European railway network. In: Van Geenhuizen M, Reggiani A, Rietveld P (ed) Policy analysis of transport networks. Part II. Ashgate Publishing Company
17. Rotoli F, Christidis P, Vannacci L, Lopez-Ruiz HG, Navajas CE, Ibáñez RN (2014) Potential impacts on accessibility and consumer surplus of improvements of the European railway system. Transp Res Procedia 3:319–328
18. ESPON (2007) Update of selected potential accessibility indicators. Final Report. Spiekermann & Wegener Urban and Regional Research (S&W), RRG Spatial Planning and Geoinformation
19. EUROPEAN COMMISSION (2013) TIPTAP ESPON project: territorial impact package for transport and agricultural policies. http://www.espon.eu/main/Menu_Projects/Menu_AppliedResearch/tiptap.html
20. NECTAR-Network on European Communications and Transport Activity Research. http://www.nectar-eu.org/Clusters/Cluster6/Cluster6.htm
21. EUROPEAN COMMISSION (2010) DG regional policy, fifth report on economic, social and territorial cohesion. http://ec.europa.eu/regional_policy/information/reports/index_en.cfm
22. Martin JC, Gutiérrez J, Roman C (2003) An accessibility DEA index to measure the impact of new infrastructure investment. The case of the HST corridor Madrid-Barcelona-French Border. ALdE (Asociación Libre de Economia), VI Encuentro de Economia Applicada,, Granada, 6 Junio 2003
23. Martin JC, Gutiérrez J, Roman C (2004) Data Envelopment Analysis (DEA) index to measure the accessibility impacts of new infrastructure investments: the case of the high- speed train corridor Madrid-Barcelona-French border. Reg Stud 38(6):697–712
24. Martín JC, Reggiani A (2007) Recent methodological developments to measure spatial interaction: synthetic accessibility indices applied to high-speed train investments. Transp Rev 27(5):551–571
25. Hesse C, Evangelinos C, Bohne S (2013) Accessibility measures and flight schedules: an application to the European air transport, European Transport, Issue 55, Paper n° 6
26. OECD-JRC (2008) Handbook on constructing composite indicators: methodology and user guide. Organisation for Economic Co-operation and Development, Paris
27. Nardo M, Saisana M, Saltelli A, Tarantola S (2005) Tools for composite indicators building. Joint Research Centre, Ispra, **EUR 21682**

28. Cherchye L, Moesen W, Rogge N, Van Puyenbroeck T (2008) Benefit of the doubt composite indicators. Workpackage 5. Deliverable 5.3. KEI-project supported by European Commission by funding from the Sixth Framework Programme for Research
29. Cherchye L, Moesen W, Rogge N, Van Puyenbroeck T, Saisana M, Saltelli A, Liska R, Tarantola S (2006) Creating composite indicators with DEA and robustness analysis: the case of the technology achievement index. KEI-project supported by European Commission by funding from the Sixth Framework Programme for Research
30. Cherchye L, Moesen W, Rogge N, Van Puyenbroeck T, Saisana M, Saltelli A, Liska R, Tarantola S (2008) Creating composite indicators with DEA and robustness analysis: the case of the technology achievement index. J Oper Res Soc 59(2)
31. Cooper WW, Seiford LM, Zhu J (2011) Handbook on data envelopment analysis (2nd edition). International Series in Operations Research & Management Science 164
32. Cherchye L, Kuosmanen T (2002) Benchmarking sustainable development: a synthetic meta-index approach. Working paper
33. Charnes A, Cooper W, Rhodes E (1978) Measuring the efficiency of decision-making units. Eur J Oper Res 2:429–444
34. Despotis DK (2005) A reassessment of the human development index via data envelopment analysis. J Oper Res Soc 56:969–980
35. Hudrliková L (2013) Composite indicators as a useful tool for international comparison: the Europe 2020 example. Prague Econ Pap 2013(4):459–473
36. Vidoli F, Mazziotta C (2013) Robust weighted composite indicator by means of frontier methods with an application to European infrastructure endowment. Stat Appl Ital J Appl Stat 23(2)
37. Saisana M, Saltelli A, Tarantola S (2005) Uncertainty and sensitivity analysis techniques as tools for quality assessment of composite indicators. J R Stat Soc 168(Part 2):307–323
38. Lozano S, Gutiérrez E, Salmerón JL (2009). Network DEA models in transportation. Application to airports. German Aviation Research Society Seminar on Airport Benchmarking, Berlin, 20–21 November 2009
39. Schaar D (2010) A method for stakeholder-based comparative benchmarking of airports. Dissertation submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at George Mason University
40. Wang TF, Song DW, Cullinane K (2002) The applicability of data envelopment analysis to efficiency measurement of container ports. IAME Panama 2002 Conference Proceedings
41. Lin LC, Tseng LA (2005) Application of DEA and SFA on the measurement of operating efficiencies for 27 international container ports. Proceed East Asia Soc Transp Stud 5:592–607
42. Yu MM, Lin ETJ (2008) Efficiency and effectiveness in railway performance using a multi-activity network DEA model. OMEGA Int J Manag Sci 36:1005–1017
43. Barnum DT, McNeil S, Hart J (2007) Comparing the efficiency of public transportation subunits using data envelopment analysis. J Public Transp 10(2)
44. Oum TH, Waters WG, Yu C (1999) A survey of productivity and efficiency measurement in rail transport. J Transp Econ Policy 33(Part 1):9–42
45. Coelli TJ, Rao DSP, O'Donnell CJ, Battese GE (2005) An introduction to efficiency and productivity analysis, 2nd edn. Springer, New York
46. Coelli TJ (1996) A guide to DEAP version 2.1: a data envelopment analysis (computer) program. Centre for Efficiency and Productivity Analysis (CEPA). Working Paper N. 8/96
47. Erkoce TE (2012) Estimation methodology of economic efficiency: stochastic frontier analysis vs data envelopment analysis. Int J Acad Res Econ Manag Sci 1(1)
48. Botasso A, Conti M (2011) Quantitative techinques for regulatory benchmarking. Cerre (Centre on regulation in Europe) study
49. EUROPEAN COMMISSION, JRC, IPTS: Trans-tools model. http://energy.jrc.ec.europa.eu/transtools/
50. RAPIDIS, Traffic analyst. http://www.rapidis.com/products/traffic-analyst/
51. EUROPEAN COMMISSION, ETIS plus Project. http://www.etisplus.eu/default.aspx