Conference Paper

New Historical Data
The Reconstruction of 1950’s Global Road Network Using American Army Maps

Author(s):
Fuhrer, Raphaël; Axhausen, Kay W.

Publication Date:
2015

Permanent Link:
https://doi.org/10.3929/ethz-b-000097821

Rights / License:
In Copyright - Non-Commercial Use Permitted
New Historical Data: The Reconstruction of 1950’s Global Road Network Using American Army Maps

Raphael Fuhrer (corresponding author)
IVT – Institute for Transport Planning and Systems, ETH Zürich
Stefano-Francini-Platz 5, CH-8093 Zürich, Switzerland
Tel: +41-44-633-67-37 Fax: +41-44-633-10-57; Email: raphael.fuhrer@ivt.baug.ethz.ch

Kay W. Axhausen
IVT – Institute for Transport Planning and Systems, ETH Zürich
Stefano-Francini-Platz 5, CH-8093 Zürich, Switzerland
Tel: +41-44-633-39-43 Fax: +41-44-633-10-57; Email: axhausen@ivt.baug.ethz.ch

Word count: 5066 words text + 4 tables/figures x 250 words = 6066 words

Submission date: 2014-07-31
ABSTRACT
In this paper, a new data set of a historical road network is presented. It represents the situation of 1950 and covers the whole globe. Its generation is documented in detail, its features are reported and, using it as a geographical information system in order to generate historical travel times in the example of Nigeria, its usefulness in (historical) transport studies is demonstrated. The main sources of this data set are maps of the former US Army Map Service and a range of historical statistics. To the knowledge of the authors, such historical transport data has not been available so far, which enables to examine various historical and current transport related research questions using spatial and time-variant data.

Keywords: World – Road network – Nigeria – 1950 – Travel time – Historical transport data
INTRODUCTION AND RELATED WORK

There are a lot of questions in transport research and transport related studies, for example spatial sciences, societal, economic and environmental research areas, or behavioural theory, to which scientists and society in general are urgently looking for an answer. However, it happens that the data available, the model specification, or the sampled results do not allow to answer these questions in the way needed. “If we measured this phenomenon before / if there was an earlier observation of this parameter / if we knew what had happened prior to, etc. – then we would be able to conclude a satisfying answer from our research”, this is the situation researchers and decision makers may be confronted with. By saying so, they imply that historical data and facts are of great value even for very current and future questions and, in the end that one can learn from history.

The question, how was the transport system in about 1950 ?, and, how have the changes since 1950 affected certain developments ?, is posed by many authors and the needed data base – a 1950 global road network – is presented here. Research on historical transport networks has been carried out for some time with different foci. Some researchers compiled numbers and facts on travel in the Middle Ages, the modern times and recent history (for example (1), (2), (3), (4)) in order to reconstruct historical journeys (e.g. (5), (6), (7)). Other researchers use data on historical transport parameters and patterns in order to explain and analyse societal and economic developments, such as slavery (8), trade and growth (for example (9), (10), (11), (12)), commuting (13), regional economic or transport policy (for instance (14), (15, 16), (17), (18)). The direction of the research project, in which the data set presented was generated, follows the rationale of combining historical transport data with other spatial and societal data in order to gain a better understanding of state formation, state reach and the occurrence of societal conflicts. The transport network depict the presence of the state in its different forms (19), as all governmental functionalities such as law enforcement, public services, and so forth require accessibility and thus transport infrastructure. In order to have a quantifiable and measurable proxy for state reach, the global road network of around 1950 was generated in a digital format as a first step. Comparing this data set with earlier or the current situation, models can be developed in order to improve the understanding of state formation, state reach and the occurrence of conflicts.

The paper has the following structure. The second section on its methods details the data generation. It deals with potential sources; explains the programming aspects of the digitalisation, which might be particularly interesting for programmers or geographic information system (GIS) specialists; next it gives an overview on the transport modelling involved. The third section provides information on the historical network’s features, its use and potential applications. The fourth section shows how the data can be used in order to estimate 1950 travel times in Nigeria and gives a brief comparison and interpretation of travel time changes between 1950 and today. The fifth and last section concludes.

METHOD: FROM SCANNED MAPS TO TRAVEL TIMES

As shown in Figure 1, the procedure of converting scanned paper maps into historical travel times between two locations consists of these three main steps: Data source assessment, automated digitalisation (geographical network), and the modelling of transport attributes to the network (logical network).
Data Source Assessment

There are not many sources that provide information on the global mid-twentieth century road network. Most of them served military interests. In order to select the best map provider, a set of assessment criteria were defined (for a description of all sources and the assessment as well as legal issues see (20)):

1. Maps provide global data on classified road infrastructure.
2. Maps come in a globally uniform cartographic style.
3. Maps do not exhibit spatially correlated accuracy.
4. Maps are available to the public.
5. Maps are available at no or low costs.

Criterion 2 follows from the fact that data on transport infrastructure will be extracted automatically using an image processing procedure. Ensuring comparability for spatial statistics dictates the third criterion. If the five criteria stated are to be satisfied, US Army Map Service (AMS) maps are the best choice. To be precise, it is series 1301 “Maps of the World 1:1,000,000”. Their biggest advantages include almost uniform mapping style amongst different sheets and editions, and clear documentation of the age and reliability of the mapped information (1).

Next, all available paper maps must be scanned. The maps used are provided by the ETH Library (21) and a set of already scanned maps by The University of Texas at Austin (22). All relevant meta data of each map are stored in one data base. Transport related meta data includes the road types mapped, the surface of the road types (hard, loose, gravel, metalled), dependence on

FIGURE 1 Schematic diagram of the three main steps including subordinating steps.
weather, and the year of the survey for the map. Maps are included if the main part of the network
represented was surveyed around 1950. However they are still included, if some map features or
some parts of the map are within the interval 1930 to 1970. This relatively generous bandwidth is
necessary to avoid too many gaps in the global coverage due to excluded maps. The other part of
meta data relates to geographic information, such as projection and standard parallels.

Subsequently, all scans are cropped to the map properly, this means deleting legend, title,
etc. spaces, and projected in World Geodetic System 1984 Mercator (WGS84). Last, national
borders are painted over with another colour, since they would be misinterpreted as roads by the
automated digitalisation algorithm, which is presented below.

Automated Digitalisation Algorithm
The algorithm is written in the language and programming environment R (23). The algorithm
needs a map prepared as described in the section above and two training sets – one containing only
road pixels and one containing only background pixels. The processing of one standard AMS map
on a server with 12 cores at 2.66 GHz each and 192 GB RAM takes between 10 and 120 minutes.
The result is a link-node description of the network stored in ESRI compatible shape file
representing the network. Primary roads are distinguished from other road types. The following
sub-sections give an overview on the single steps in the algorithm. See Figure 2 below for the
interim results of each step [numbers in brackets]. Its details are provided in (24).

Extraction of road pixels[1]
The main element is a supervised learning model, which evaluates each pixel on its colour
(red-green-blue shares) and, accordingly, categorises the pixel into foreground (road pixel) or
background pixel (non-road pixel). The model is based on a support vector machine (SVM) that
models the decision rule. This SVM is trained on the two training sets and optimised for predicted
accuracy.

Cleaning I: Removing Circles [2]
All circles, indicating road numbers on the map, need to be removed to ensure smooth and correct
post-processing. By using the Circular Hough Transform of the image, every pixel is evaluated on
its probability of serving as the centre of a potential circle. The surface of identified circles is filled
with road pixels and will be reduced to the original road line in the next step.

Thinning of the Pixel Image (Skeletisation) [3]
The Zhang-Suen thinning algorithm (25) is applied next. In every iteration, the outermost layer of
each connected shape is deleted until there are no surplus pixels. Thus, pixel sequences of the size
of exactly one pixel are the result, resembling a skeleton. In this way, the input for the vectorisaton
in the next step is clearly defined.

Vectorisation [4]
This process converts the raster image into vector format. The shapes of the line geometry are
generated by a naïve tracing algorithm. It starts at one pixel in a pixel sequence and connects the
pixels into a line until it has reached both ends of the whole pixel sequence.

Cleaning II: Standardise and smooth lines
This step improves the quality of the line shape geometries. Very short line segments (~points) are
deleted. Line geometries are broken at all intersections of line geometries. There, nodes are
defined and consecutive line segments between two nodes are connected into one line segment. This improves line segmentation and link-node sequence. Duplicate line segments are removed. Smoothing eliminates raster-like pattern of vectorised line geometries.

**Snap Line Segments Together [5]**
This step fills network gaps that have arisen out of errors in previous steps. Different line segments are snapped together into a link-node network. Potential nodes are identified and evaluated on their spatial position relative to other nodes. Nodes (intersections) are produced if the spatial position and the role in the network of the potential node improves network connectivity and other network parameter above a certain threshold. This ensures that only gaps due to processing errors rather than real gaps are reconnected.

**Classification Road Segments [6]**
Two road types, main and secondary road, are assigned to each segment based on network characteristics and a probability model. This hidden Markov model (26) follows all roads and decides when to switch from type one to type two. In the next section, the modelling of transport on this network is explained.

**FIGURE 2 Interim results of the main steps in the automated digitalisation algorithm.**

**The Modelling of Transport Related Attributes**
The network’s characteristics do not only include the spatial position of roads but also their technical details such as pavement, weatherability, capacity, etc. The overall objective of this step is to achieve a reasonable estimate of the network’s properties in order to associate speeds on the links, which are historically appropriate. Knowing these and the length of the roads, it will be possible to define historical travel times for journeys in 1950.

Speed on roads is influenced on the one hand by the infrastructure and on the other hand by the
vehicle. The maximum speed is the minimum of these two speed values. If we assume an average European car around 1950, the conclusion from various sources (e.g. see (27), (28), (3), (29)) can be drawn that, in general, it was the infrastructure rather than the vehicle that defined the maximum speed. However, there are two exceptions: First, some steep roads where cars could have driven faster if they would have had more horsepower; second, completely developed motorways. In these two rather rare cases, the maximum speed was determined by the maximum vehicle speed, which was around 120km/h for an average European passenger car ((1, 29)). Therefore, in 1950 it is the infrastructure that limits the speed of a road link. Using different information, the logical network model converts the input information – whether a road segment in a specific region is of type “main road” or of type “all other roads” – into a maximum speed on this road segment. First, there is a logical connection between the mapped and the real-world road type established. Second, this information is connected to further parameters, e.g. surface quality. Third, speeds are attached to the first and second item in order to obtain a reasonable maximum speed per road link. These parameters are shown in Table 1.

### TABLE 1 Link attributes: Resolution, values, and sources

<table>
<thead>
<tr>
<th>Information</th>
<th>1. Mapped road type</th>
<th>2. Corresponding real road type</th>
<th>3. Surface types</th>
<th>7. Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per unit</td>
<td>Per link</td>
<td>Per mapped type</td>
<td>Per mapped/real type</td>
<td>Per surface type</td>
</tr>
<tr>
<td>Cases</td>
<td>• Main road</td>
<td>Different terms, e.g.</td>
<td>• Hard</td>
<td>• Paved: 80km/h</td>
</tr>
<tr>
<td></td>
<td>• Secondary road</td>
<td>• Motorway</td>
<td>• Loose</td>
<td>• Gravel: 64 km/h</td>
</tr>
<tr>
<td></td>
<td>(all other roads and trails)</td>
<td>• Track, path</td>
<td>• Gravel</td>
<td>• Earth: 56 km/h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Trail</td>
<td>• Metalled</td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>AMS map meta data</td>
<td>AMS map meta data</td>
<td>AMS map meta data</td>
<td>World Bank (30)</td>
</tr>
<tr>
<td>Per unit</td>
<td>Per real road type</td>
<td>Per country</td>
<td>Per mapped type</td>
<td></td>
</tr>
<tr>
<td>Cases</td>
<td>• Paved</td>
<td>Percentage of paved road of the total length of all roads</td>
<td>• All weather</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Gravel/crushed stone</td>
<td></td>
<td>• Dry only</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Earth graded</td>
<td></td>
<td>• Unclear</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Unimproved</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>IRF (31), network data</td>
<td>IRF (32)</td>
<td>AMS map meta data</td>
<td></td>
</tr>
</tbody>
</table>

AMS: Army Map Service; IRF: International Road Federation; 1 km is 0.621 mi

After compiling the different items presented in Table 1, a specific maximum speed is attached to every link in the whole global network based on the link’s known attributes derived from the AMS.
map meta data and additional statistics. These items interact with each other. For example, a road link of type “main road” may be a paved motorway lane in one location, but can be a track with hard surface in another map sheet. The speeds in Table 1 are global maximum values and are adjusted downwards considering the information in Table 1 on local conditions.

DATA SET AND POTENTIAL APPLICATIONS

Format
The final data set is stored as a shape-file (data.shp) and can be imported with most GIS software. This data set includes the line geometries of roads around the year 1950 and their transport related attributes concerning quality and speed (some of TABLE 1 and others, for example from which map they are scanned). This GIS uses the reference and projection standard World Geodetic System 1984 Mercator (WGS84).

Coverage and Completeness
AMS maps were produced globally except for areas over 85° North and 60° South respectively. However, at the time of preparation of this paper, there are some gaps, especially in Northern Europe, Russia and South Africa. Some sheets are editions surveyed considerably after 1950 in Australia and South America. This situation is improved continuously by researching further map sources. The question whether the historical maps include the complete road network or whether links are missing cannot be answered in retrospect. However, it seems very reasonable to assume that the AMS staff carefully compiled and mapped the geographic features, since they themselves relayed on these maps.

Accuracy and Granularity
Accuracy is indicated on a sheet-by-sheet basis as “good”, “fair” and “poor” by AMS. Most parts of available maps are labelled as “good”. Doing overlays of randomly chosen digitised networks with their original scanned maps, a high precision of the result is observed. Locations where the mapped road line is frequently interrupted by other cartographic features or locations where a lot of different road lines coincide lead to minor deviations. If the historical road network is compared with current aerial photos at roads that are present in both data sets, the difference of the overlapping road geometry is small. In general, differences for subordinate roads are larger than for main roads.

Given the observed local accuracy and the comprehensive representation of the road, track, path, trail, etc. network at a scale of 1:1 000 000, this data is suitable for projects on an international, national, regional and in some cases even on a municipal level.

Potential Applications
A digital and global network data set representing the mid twentieth century road based transport infrastructure is interesting in transport related as well as other studies. As mentioned in the introduction, the data will be used in the context of the analysis of inner state conflicts, conflicts due to natural resources, and state formation. The road network and the associated travel times are a measure of the degree of state presence or in other words state reach (19).

The data set presented here can be used in order to expand existing transport models back in time. It allows to compute historical travel times and, adding structural data (population, jobs, etc.), to generate historical accessibility values. Trends in road-based transport supply can be
visualised and quantified. Obtained results can be compared to infrastructure expenditures and political goals regarding transport supply and development by international and national policies. Adding further spatially explicit data, many other studies are possible.

**TRAVEL TIME ANALYSIS 1950 AND TODAY**

An illustrative research question is how travel times between municipalities in Nigeria have changed from 1950 to today. Nigeria lies in West Africa, was part of the British Empire until 1960 when it became a federal republic and is currently the most populous African country. It consists today of 775 local government areas (33). A transport model with two scenarios – 1950 and 2010 – is constructed, which consists of the following input data:

- Shape file of all current 775 municipalities for both scenarios (33)
- Our generated network 1950 with two road categories
- A current network by Open Street Map (OSM) (34)

The model is generated in the software VISUM (35). The municipalities are used as the zones and they are connected to the network by connectors. These connect the centroid of every zone to the nearest node in the network. In the comparison between the AMS and OSM networks, the respective maximum speeds are compared; the actual driven speed might be lower. In the current network, speed limits for every link are included in the OSM data file and define maximum speeds (see Figure 3). For the historic network, the method described above was employed. Maximum speeds are downgraded to 64 km/h for main roads and 31 km/h for all other roads. This is due to a low pavement share of 17% on main roads, most have loose surface (extracted from (32) and AMS map meta data). The secondary road category consists of approximately 20% of tracks with loose surface and maximum speed 56 km/h and 80% of trails with only a third of the track’s speed (extracted from (31) and AMS map meta data). The speed of trails is derived from the speed value of tracks. A trail is regarded as the disadvantageous possible track specification on which a car is still able to drive. Accordingly, all corresponding speed reduction amounts (surface, weatherability, etc.) are subtracted from the track’s value (30). The connectors represent a footpath to the next passable infrastructure and have a speed of 4 km/h. The current OSM network provides a detailed network of motorways, primary, and secondary roads, but is lacking reliable information on subordinate links (tracks, trails and so forth). Thus, an overlay of the two data sets is done to identify roads that have been improved and in order to update the OSM data file. This means, the OSM file is enriched with some of the information of trails from the 1950 AMS network. For every pair of municipality the shortest (in time units) path is found and travel time matrices are computed.
1950: Total travel time per municipality and complete road network

2010: Percentage of change per municipality in total travel time and primary road network

FIGURE 3 Comparison of the travel times and networks 1950 and 2010. Total travel time is the travel time from a municipality to all other municipalities. 1 km is 0.621 mi.
The results in Figure 3 show that in 1950 the network of improved roads had limited coverage. Several municipalities were not reached by this type of road. The structure of this higher-level network shows two parts, a denser southern part and a looser northern part. This coincides with the two British protectorates: Southern and Northern Nigeria. There are several gaps in the higher-level network, which shows the construction in progress, whereas the subordinate network is dense and complete, especially in central-northern areas. The travel time values are absolute values per municipality. It is the sum of all travel times from one municipality to all other municipalities. One would expect that more central areas have shorter travel times. This is partly true, since most of the municipalities with short travel time are in a central location, but there is also a cluster in the south-eastern area. This area seceded in the late 1960s, shortly after oil extraction started, from Nigeria and declared itself as the Republic of Biafra, but it was reintegrated into Nigeria in 1970 (36). The spatial correlation between the main road network and areas with low total travel time is not obvious. This effect is only observed in areas where, in addition to the main network, the subordinate network is dense. The variation between neighbouring municipality could be due to local variance in network density or gaps in the network (e.g. rivers or canyons), which can have very local effects.

In the second part of Figure 3, the network drawn represents only the higher-level roads with speed limits greater than 80km/h. As visible, but also checked with spatial analysis, most of former main roads have been developed into current primary roads, some 1950 subordinate roads as well. Completely new roads were built in a few cases, particularly the motorway in the Southeast between Lagos and Ibadan, two cities with several million inhabitants each. The coverage with higher-level roads is much better than it was in the 1950. The structure is denser, especially around urban centres. The network is more complete. Rivers are bridged now. Here, the travel time is given as the change in percent compared to 1950. In the southern half of the country the largest changes occurred in the former “white gaps” in the upper part of Figure 3, so areas relatively badly connected in 1950. This holds true for the area near the coast, just north of the coast arc, and for the one between the 1950 central cluster and the coastal area. The in 1950 relatively badly connected areas in the northwest and east are improved, but to a lower extent than the ones towards the coast. The situation in the whole northern part and in the southeast has improved less. Especially, the northeast, which already was not well connected, has not improved a lot. As mentioned above, the 2010 subordinate road network is updated with the 1950 one. This might cause errors in the travel time calculations. However, as observed in previous studies, the main effect by large is due to the higher-level network ((14), (37)) and this is modelled very reliably. Thus, we can also rely on the spatial distribution of the travel time values. Interesting is the mentioned improvement in travel time in the area just north of the coast arc. In this area adjacent to the Niger delta, a lot of oil fields are located – oil was first discovered there in 1957 (38), which apparently gave incentive for infrastructure improvement. The southern part of Nigeria is generally reported to be economically stronger than the northern part and more dominant in national politics (39). Also, the religious border follows this division, as the south is predominately Christian and the north Muslim. The application of our data set to Nigeria in this initial example, which should be expanded with in-depth analyses, already shows the versatility of this historical transport infrastructure data set.

**CONCLUSION AND OUTLOOK**

The presented data is to the knowledge of the authors the first GIS of a historical global road network. The most crucial difficulties are accessibility and suitability of sources, uniformity in
mapping style and year of surveying, and the very large amount of manual work despite the
development of an automated algorithm. Hence, the data generation is feasible, though costly. In
order to obtain a logical network that includes the basic information of road networks, additional
substantial information and work is required. On the other hand, the advantages of such a data set
are obvious. Time series data model settings become possible, transport models can be expanded
back in time. This kind of data makes changes in transport visible and measurable, and it can help
to improve models in traffic behaviour or transport and land use interaction. At the same time, it is
a valuable basis for further models, for example in spatial econometric models.
First priority is given to filling the existing gaps in the network due to missing map sheets. Any
information about available maps of the AMS series 1301 or 1106 as paper or scanned maps to the
Corresponding author is appreciated. Second, the logical network will be refined by including a
digital elevation model in order to account for the effect of gradients on the maximum speed.
Third, rail and navigable waterways, that may be faster than roads in some cases, are to be
included. Calculating accessibilities (see (40)) is also an important goal, for which the global
population data is currently prepared. In order to expand the model in Western Europe back to
1500, the required data is currently compiled.

ACKNOWLEDGMENT
This research is funded by the ETH Risk Center project ETH48 “The effect of resource
endowments and limited state capacity on political conflict” and has been carried out in
collaboration with Philipp Hunziker from the International Conflict Research group. Daria Wild
and Patrice Frei, our two research assistants, supported us in the data preparation.

REFERENCES

   Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich, 2004.
   Raumwiderstand. In Verkehrsgeschichte = Histoire des Transports (H.-U. Schiedt, L. Tissot,
5. Czysz, W. In der römischen Reisekutsche übers Gebirge. In Über die Alpen - Menschen,
   Wege, Waren (G. Schnekenburger, ed.), Konrad Theiss Verlag / Archeologisches
   the internal migration of Americans, 1850–2000. The journal of American history, Vol. 91,
7. Livet, G. Histoire des routes & des transports en Europe. Presses universitaires, Strasbourg,
   2003.
Columbia University, and Centro Internacional de Agricultura Tropical - CIAT; by NASA Socioeconomic Data and Application Center (SEDAC), Palisades, 2000.


