Developing a Transport GIS Database for Economic Development Planning

Jiawen Yang

Georgia Institute of Technology, Georgia, Atlanta, USA
jiawenmit@yahoo.com

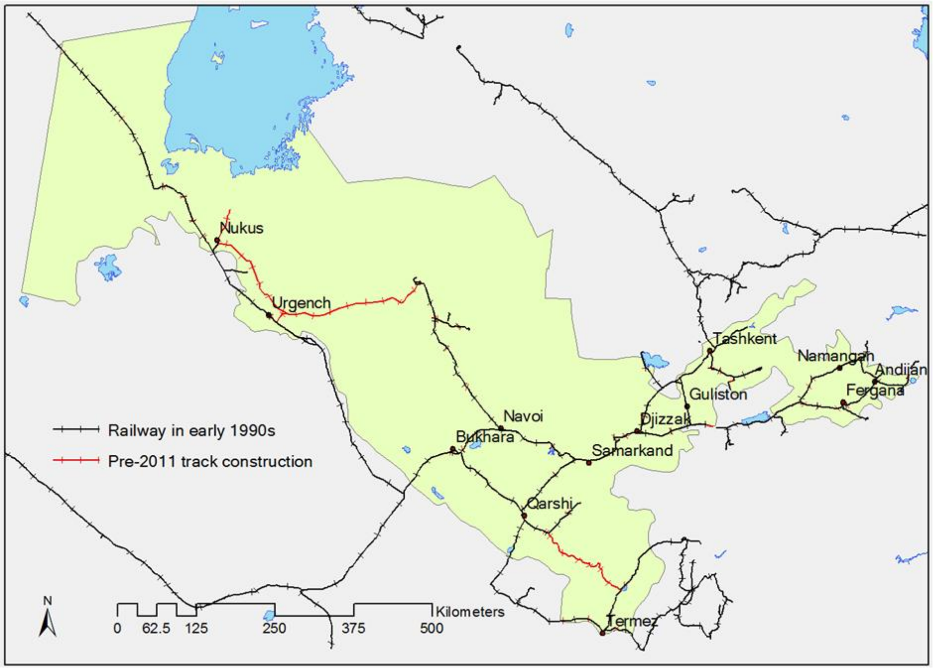
Abstract

Countries in Central Asia face serious challenges in transportation following the collapse of the Soviet Union. In order to facilitate transport infrastructure investment and planning there, we developed a transport database based on geographic information systems (GIS) for two Central Asian countries, Kazakhstan and Uzbekistan. This dataset will be used for assessing the status of current transport network, forecasting transportation demand, evaluating transportation investment plans, and visualizing transport sector development. Significant efforts have been made in order to produce segment-level attribute information for the road network by geocoding a trucking survey. Assessment of the developed dataset indicates sufficient data quality for network analysis and modeling.

1. Introduction

Like many other developing countries, the member states of the former Soviet Union (FSU) are eager for economic development [1, 2]. Unfortunately, the transportation infrastructure has become a bottleneck [3]. For example, in Uzbekistan, central Asia's most populated country, railway was part of the FSU system. It was oriented toward Russia rather than toward domestic connectivity within Uzbekistan. At the time of Uzbekistan's independence, railway connections between Tashkent and other provinces in the Fergana Valley passed through Tajikistan. Connections between Termez, Nukus and Bukhara rely on railway links within Turkmenistan [4]. Kazakhstan, Central Asia's biggest country by land area, still lacks internal connectivity, as its railway network was likewise designed to link various parts of Kazakhstan to Russia, rather than to develop an integrated market for the Kazakh nation [5].

Fig. 1. Railway track in early 1990s and post-independence construction in Uzbekistan



This FSU legacy has roughly dominated railway investment after the independence of these two countries. The map below shows the railway network in early 1990s and the post-independence construction by 2011 in Uzbekistan. The black lines illustrate how the railway in Uzbekistan was interconnected with railway links in neighboring countries in the 1990s. The red lines are railway tracks constructed after independence, which helps to avoid the need to pass through Turkmenistan. Domestic railway connection with Fergana Valley is still missing, partially because of the prohibitive construction cost around the mountainous Kamchik pass.

The Institute of Central Asia Regional Economic Cooperation (CAREC), an institute mainly sponsored by the Asian Development Bank and the World Bank, is promoting regional wide cooperation for economic development [8]. It helps the central Asia countries to strengthen domestic transport connectivity at one hand, to recover trade relationship at the other hand [6, 7, 8, 9 and 10]. Whilst addressing this problem of transportation and trade problems with additional investment projects, the countries are also eager for transport data of improved quality, which could be used to support investment decisions [11]. Transportation infrastructure and services provide mobility and accessibility to individuals and businesses – facilitating their interactions with each other across physical space of various geographical scales. In order to best promote economic development and maximize efficiency, the supply of and the demand for transportation infrastructure and services are expected to match one another not only at the aggregate national or regional level, but also spatially at the sub-national level, or at the corridor level. A mismatch between the demand and supply will reduce the efficacy of the investment and slow down economic development. This spatial perspective is particularly important in the Central Asian region, where the regional economy is still waiting to be better integrated and the growth of trade could be faster if the gaps and the bottlenecks on the transportation network can be identified and problems can be mitigated with fine-tuned investment strategies for transportation infrastructure and services.

The aim of our work is to develop such a transport-centered database based on geographic information systems (GIS) for two Central Asian countries, Kazakhstan and Uzbekistan. This GIS database will be part of the foundational dataset for assessing the status of current transportation network: forecasting transportation demand; evaluating transportation investment plans; and visualizing a transportation-sector road map.

2. Database development

Our work started in February 2011 and ended in December 2012. After intensive work on data development, the database now includes the following layers: a. Administrative boundary, including country boundaries and province boundaries;

b. Components of the transportation systems, including airports, railroad tracks, railway stations, roads, pipelines, lakes and rivers;

c. Major production and attraction points, including cities, mine locations, oil and gas locations, ore locations and border crossings;

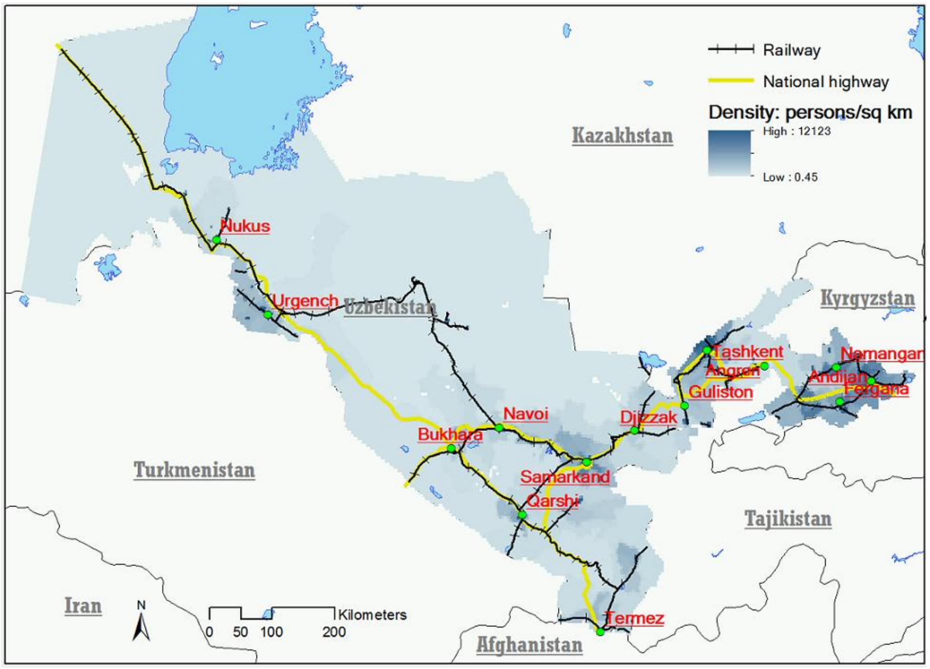


Fig. 2. Example map of Uzbekistan based on developed dataset

d. Various attribute information, including population counts at the province level and in major cities, passenger and freight volumes and service characteristics at major airports, driving speed for each infrastructure segment and major corridors, freight volumes and service characteristics at border crossings.

The map (Figure 2) given as an example was created for Uzbekistan based upon attributes established in our database.

By the end of the project period, 18 GIS files have been added into the database, including 12 data layers extracted from existing datasets of various sources, and 6 data layers created from scratch. The two major data development steps are explained below.

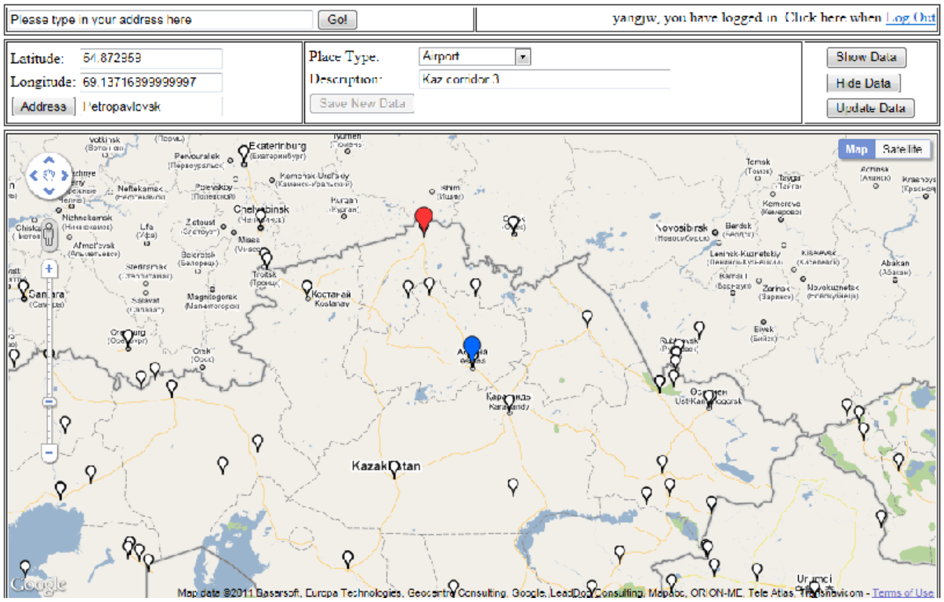
First, we created baseline data files by extracting features from existing datasets of several organizations, including the Environment System Resource Institute (ESRI) and the United States Geological Survey (USGS). Those GIS file typically cover the whole globe. We deleted features that are distant from these two target countries. Most datasets were developed in the late 1990s and they need significant update in order to reflect the changing situation. However, these data provide a starting point for information visualization, updating and integration. Most of these datasets come with metadata, whose accuracy has been examined carefully whenever applicable. Geographic data have been projected into the same coordinate systems and integrated into a single geodatabase. Given the publication dates of the above datasets, some attribute data have been updated to reflect the most recent situation. Given population growth's particular relevance to transportation demand, we have collected provincial-level population data for Kazakhstan and Uzbekistan from Wikipedia's website. The 2008 population counts were downloaded from http://en.wikipedia.org/wiki/Kazakhstan and http://en.wikipedia.org/wiki/Uzbekistan.

Unlike administrative and geographic terrain, transportation and trade relevant information evolves much faster than the baseline datasets. We created six additional transport-centered GIS data layers as they were not readily available but are essential to transportation analysis. Those files include the following:

a. A GIS file of railway stations in Kazakhstan and Uzbekistan is created based upon the station list from wikipedia. (http://en.wikipedia.org/wiki/Railway_stations_in_Kazakhstan; http://en.wikipedia.org/wiki/Transport_in_Uzbekistan#Railways). Each station is identified with a point in the GIS file.

b. A separate GIS file of border crossings has been created for Kazakhstan and Uzbekistan. Major border crossings have been identified based on the intersection of the corridors and county boundaries. The list of the board crossings comes from the consultant report for the CAREC 2008 Transport Sector Strategy Study (page 28) [6].

c. A separate GIS file showing six major CAREC corridors is coded into GIS. These corridors are identified by CAREC in 2009 and documented in relative detail in the CAREC 2008 Transport Sector Strategy Study.



Please type in your address here

yangjw, you have logged in. Click here when Log Out

Latitude: 54.872959

Longitude: 69.13716399999997

Address: Petropavlovsk

Place Type: Airport

Description: Kaz corridor 3

Fig. 3. Interface of the web application for data development

NB: This web application is now hosted in a Georgia Tech website and password protected

d. Separate GIS files showing the location of ports, airports, and logistic centers on the corridors or in these two countries are also created. In order to accurately locate those places on maps, we have created a web application based on Google map API, which is linked to a MySQL database. We searched for the location first and then we inserted the location coordinates together with the place name and other attributes into our database. The work log would be automatically updated as we would continue to insert new data points into the database. The dataset was originally stored in the

MySQL database and then converted into an ArcGIS point shapefile. A screenshot of the web application is in Figure 3.

On our field trips to Kazakhstan and Uzbekistan, we brought hard copies of maps showing the position of these transport-relevant locations. We updated those maps after returning from our trip with feedback from local professionals and international organizations.

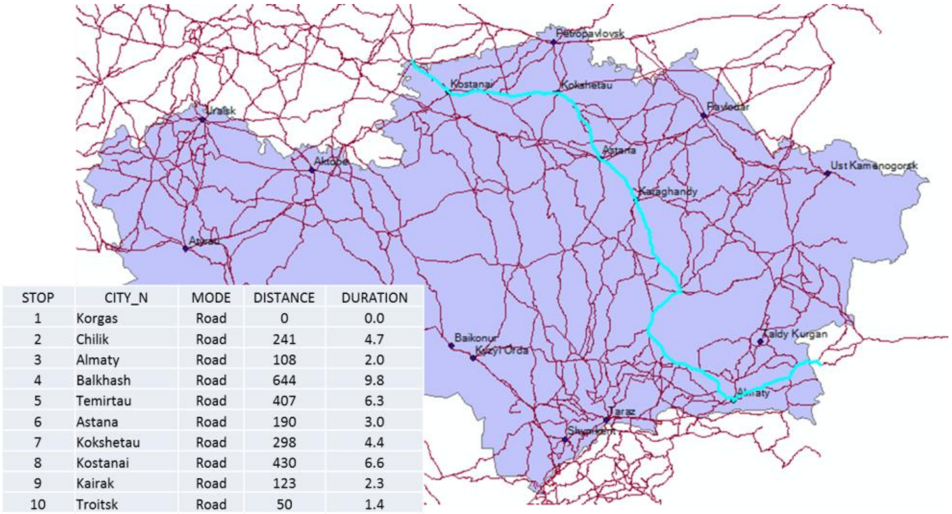
3. Geocoding attribute data

Governments in Kazakhstan and Uzbekistan did not have GIS data. The maps available in these two countries were typically created in graphic processing software such as Photoshop or CorelDRAW. We have used the meetings during our field trip to fill the gaps in attribute data. We submitted a data request in written format, in both English and Russian and received no response. Quite a percentage of information on transport infrastructure and status is presented haphazardly in government documents, rather than in a systematic database. We also asked for government documents from both countries, particularly relevant investment plans. As a result, we received a limited amount of documents from relevant government units. Most of these documents are in Russian; key segments of those documents were identified and translated into English.

Additional efforts were carried out to produce service characteristics for major roads at detailed geographic levels. This was necessary for transportation network analysis and investment impact assessment. Data at this level was not available. Fortunately, the Asian Development Bank’s program of Corridor Performance Measurement and Monitoring (CPMM) has conducted a survey for trucking activities throughout the Central Asia region. Participating truck drivers were asked to fill in detailed stop by stop information for each selected trucking trip, including start-point, endpoint and each intermediate stop. Relevant time stamp information is also reported.

The CPMM data possesses a very good geographic coverage. Asian Development Bank has identified six major corridors for this region based on traffic volume, connection to external markets, potential economic growth and other criteria. CAREC has been collecting survey from freight carriers and forwarders since 2009 and a group of aggregated performance indicators has been published, either for the whole region or by corridor. Published performance measures (http://cfcfa.net/cpmm/data/) indicate significant performance variability among different corridors, suggesting a great potential to improve transportation performance through investment in infrastructure modernization, and improvement of transportation and trade processes. Upon our request, the CPMM group emailed us the survey data collected in 2010. It contains over 2000 freight trips and over 35,000 data records of stop-by-stop information on cost, time, and distance. Link- and node-level service characteristics of the transport network can be extracted from this survey data once it is geocoded into our database.

The example below illustrates how CPMM data records a typical trucking trip and how our team geocode the CPMM data into the GIS database and how we estimate driving speed for relevant road segments. The example freight trip starts from Korgas at the China border and ends at Troisk at the Russian border. It passes intermediate steps including Almaty and Kostanai. The survey includes distance (in km) and travel duration (in minutes) between adjacent stops, based on which travel speed can be calculated.



STOP	CITY_N	MODE	DISTANCE	DURATION
1	Korgas	Road	0	0.0
2	Chilik	Road	241	4.7
3	Almaty	Road	108	2.0
4	Balkhash	Road	644	9.8
5	Temirtau	Road	407	6.3
6	Astana	Road	190	3.0
7	Kokshetau	Road	298	4.4
8	Kostanai	Road	430	6.6
9	Kairak	Road	123	2.3
10	Troitsk	Road	50	1.4

Fig. 4. Geocoding an example CPMM trip

We first geocode every stop recorded in the CPMM survey. Over 600 stops have been added into our database. We then geocode all trip segments for each trip one by one. For the first trip segment, which runs from Korgas to Chilik, we assume it follows the shortest distance path between these two places. The travel speed is then assigned to relevant road links on this path. Adding all segment paths together, we have the path for the whole trip, which is highlighted in the Figure 4.

A single road link could be used in the geocoded paths for multiple instances because multiple trips traverse that road link. This situation enables us to estimate the average speed, average shipment time, as well as the variability of speed and shipment time. The geocoding efforts results in over 2000 road links with CPMM service characteristics. The map below shows the average speed on different road segments within Kazakhstan, with green color showing relative high driving speed and red for relatively low speed.

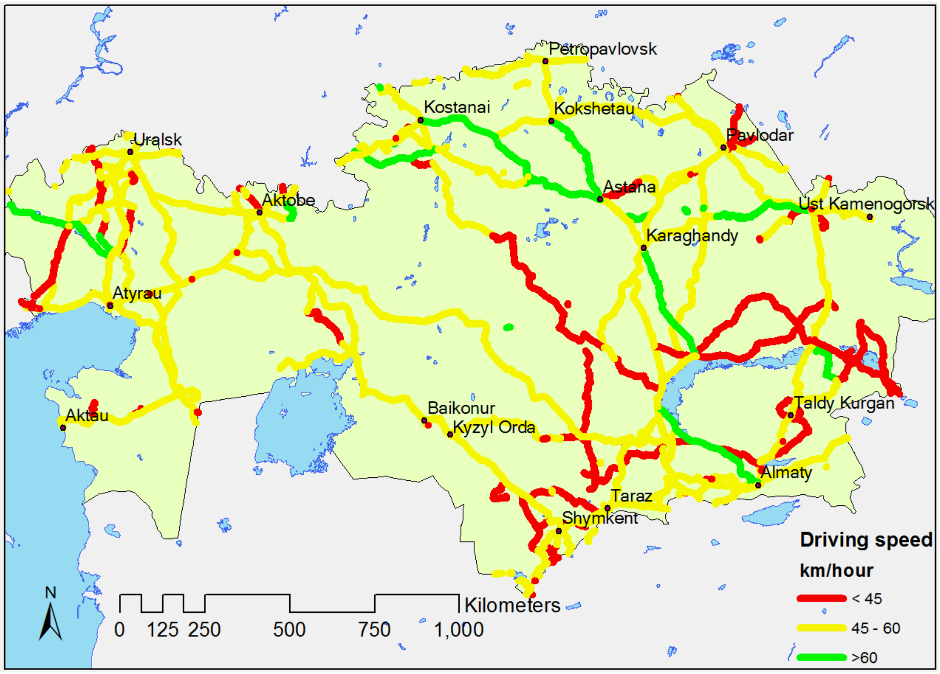


Fig. 5. Estimated driving speed based on CPMM survey

Development of investment plans for the road and railway sectors can greatly benefit from spatially detailed information on infrastructure condition and usage. Corridors with heavy traffic but poor infrastructure are priority investment targets. The Asian Highways data from the United Nations provide detailed information at the road segment level, which helps to identify investment opportunities. Following a similar approach, we have geocoded this dataset. The traffic volume map below shows average annual daily traffic counts on different segments of Asian Highways, highlighting the heavily traveled corridor from Tashkent to Fergana Valley and from Tashkent to Bukhara. It also shows highway quality, which depicts the distribution of four-lane highway roads on the Asian Highway. Thin segments between bold segments indicate infrastructure bottlenecks, particularly when these links also have high traffic volumes. Examples include roads from Tashkent to Fergana Valley and from Bukhara to Qarshi.

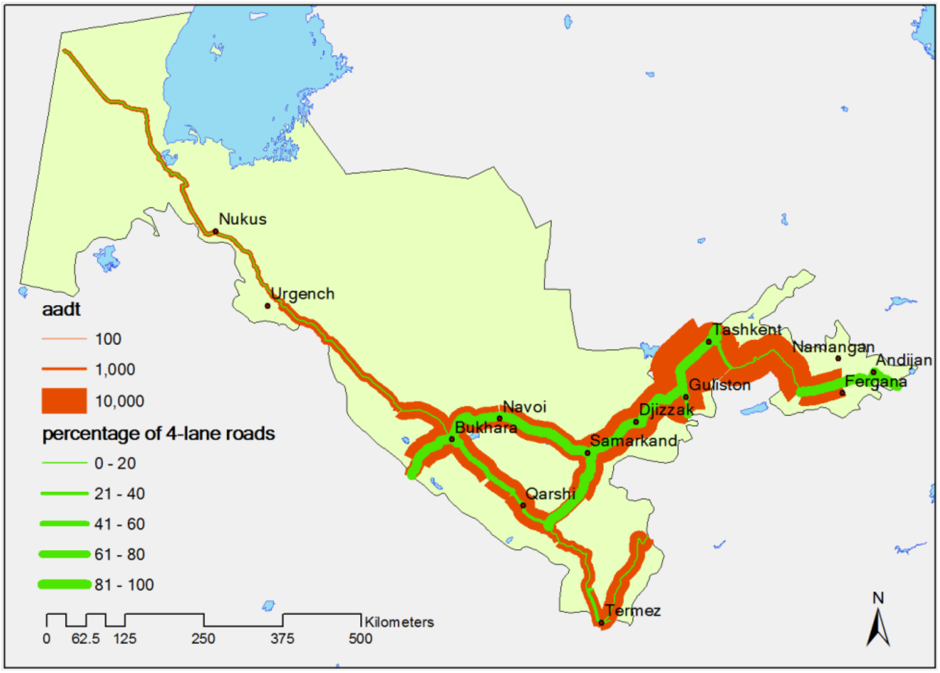


Fig. 6. Asian highway traffic volumes and four-lane highways 4. Assessing data quality

4. Assessing Data Quality

Spatial data accuracy is a typical issue when integrating data from multiple sources for visualization and analysis purposes. Different layers which come from different sources may not line up well with each other. Our GIS database includes GIS layers from multiple sources, as indicated in prior sections. We have also generated our own data layers. Once we overlay these data for visualization purpose, we add satellite images. An initial test in Google Earth shows that data accuracy is good enough for visualization purpose. For example, in Figure 7, we overlay data from three data sources: satellite image from Google Earth, airport locations from the self-developed datasets, and the provincial boundaries and corridors extracted from the baseline dataset.

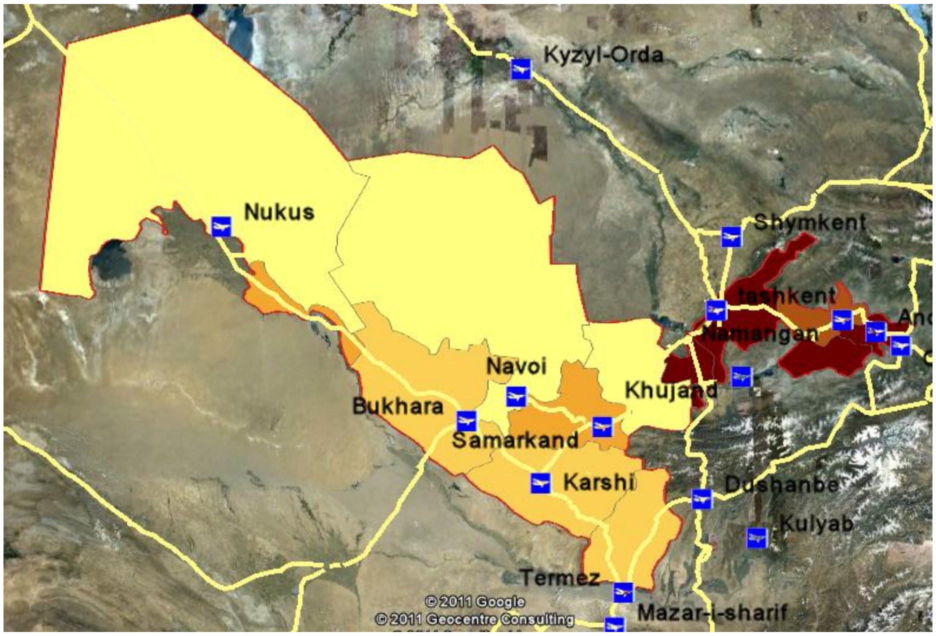


Fig. 7. Example of population density map using GIS data from multiple sources

For forecasting purposes, the GIS database should enable network-based forecasting. A preliminary network model has already been developed to test the connectivity of the baseline road file. Shortest path analysis has been conducted and a matrix of shipment distances between major origins and major destinations has been created. Figure 8 shows the desire lines produced out of this network analysis. The attribute table of this desire line layer contains road-based distances from every Uzbekistan airport to every other airport in this region. Preliminary testing confirms that the data is topologically sufficient for network-based modeling.

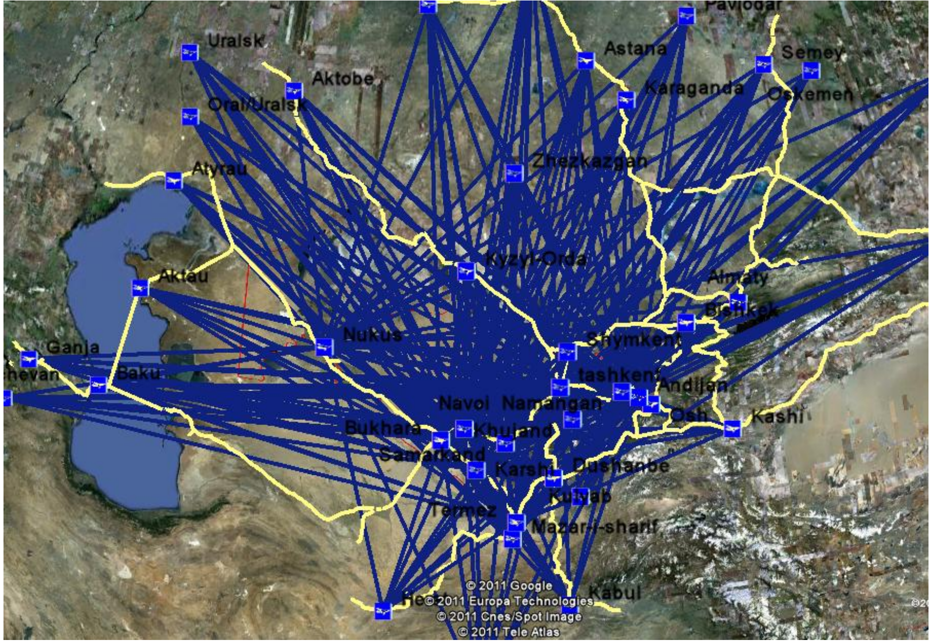


Fig. 8. Desire lines based on network analysis

Network model based on a completely updated dataset can enable us to develop and evaluate transportation investment strategies. Investment in both physical and non-physical projects can be evaluated at varied geographic levels.

At the route level, we can examine how technical and regulatory innovations reduce transport cost (whether it is distance, time, or monetary) and increase service reliability. At the corridor level, we can analyze how multi-modal strategies can lead to a better integration of the system, stimulate transport demand and change mode split, which should also have environment and energy implications. At the network level, we can evaluate how transportation investment and geographically differentiated population/employment growth can provide better access to economic opportunities for the typical citizen.

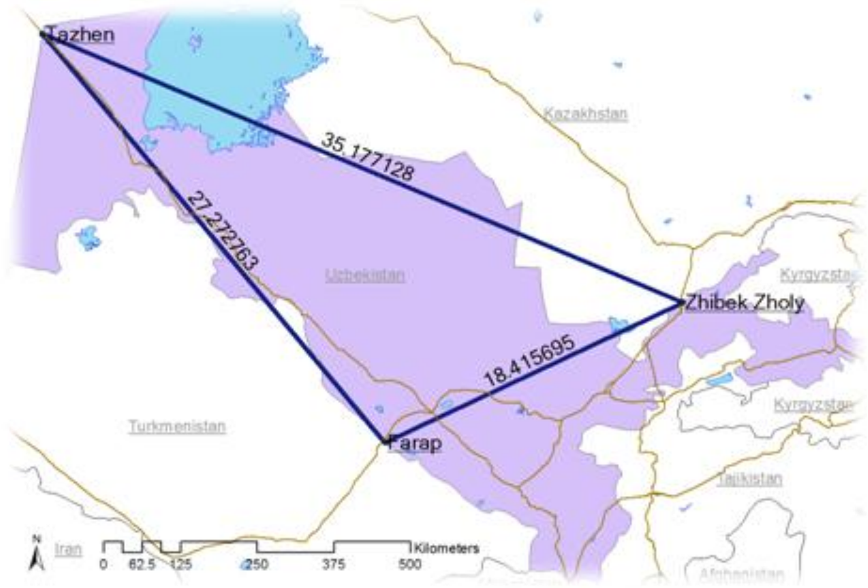


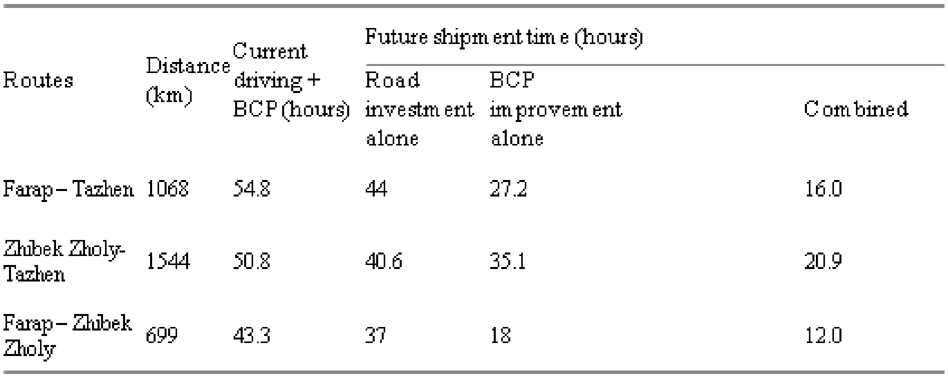
Fig. 9. Driving hours on CAREC corridors after BCP improvement

The example (Figure 9) illustrates how we use the geodatabase and network model to assess the impacts of investment in Uzbekistan's road corridor and border crossings. Given Uzbekistan's central position in Central Asia, transport investment should have large impacts on regional transportation. Uzbekistan cannot build a modern economy without an efficient regional transportation system that facilitates trade among neighboring countries. The quality of regional transportation service, in terms of delay and cost, however, has become less reliable as the result of custom and border control at the country boundaries and pavement quality of road infrastructure. A continuous 500-km trip along a CAREC corridor is traveled at an average speed of 35.6 kilometers per hour and costs an estimated 984 USD (12, 13). Border-crossing points (BCPs) are major bottlenecks in the region and need a clearance time on average of 9.0 hours at a cost of about 159 USD. Waiting time alone at Beyneu, Irkeshtan (both PRC and KGZ), Torugart, and Khorgos is longer than 10 hours. Removing or reducing this border relevant cost can provide immediate and significant gain in transportation and trade.

In order to determine how Uzbekistan's investment on road infrastructure and BCP improvement can benefit regional transit, we pick up three important border crossing points in neighboring countries: 1) Farap in Turkmenistan on CAREC corridor 3; 2) Tazhen is in Western Kazakhstan on CAREC corridor 2; and 3) Zhibek Zholy in Southern Kazakhstan on CAREC corridor 3.

The current driving speed on CAREC corridors are calculated based on the geocoded CPMM data. The updated wait-duration, in hours, at BCPs for relevant border crossings also comes from the same CPMM survey. Our network analysis calculates shipping hours, which includes driving hours on relevant CAREC corridors, BCP hours for the three BCPs mentioned above, as well as BCP hours for the Uzbek control points. A shipment traversing Uzbekistan from Farap to Tazhen will need 54.8 hours in order to cover the 1,068 km distance, as shown in Table 1.

Table 1. Proposed investments which would significantly decrease shipment hours on



Routes	Distance (km)	Current driving + BCP (hours)	Future shipment time (hours)		
			Road investment alone	BCP improvement alone	Combined
Farap – Tazhen	1068	54.8	44	27.2	16.0
Zhibek Zholy-Tazhen	1544	50.8	40.6	35.1	20.9
Farap – Zhibek Zholy	699	43.3	37	18	12.0

Note: Farap is in Turkmenistan on CAREC corridor 3; Tazhen is in Western Kazakhstan on CAREC corridor 2; and Zhibek Zholy is in Southern Kazakhstan on CAREC corridor 3. All are BCPs.

Assuming that the construction of Uzbekistan's national highway increases travel speed to 100 km per hour, but the proposed investment in BCP improvement does not happen, the same trip will be reduced from 54.8 hours to 44 hours. Alternatively, assuming the proposed BCP improvement reduces BCP time to 2 hours at each BCP point for each relevant country, but the proposed investment in road infrastructure does not happen, the same trip can be covered in 27 hours. If both road investment and BCP investment happens, the 54.8 hour trip shall be reduced to 16 hours.

5. Summary

With limited GIS data availability and data sharing possibilities in Kazakhstan and Uzbekistan, the biggest challenge and most time consuming task has been to identify the sources of data and to match the obtainable data with the infrastructure planning needs. Despite the poor data availability, we have successfully developed a useful GIS database for transport mapping and network modeling. One significant potential use for our dataset and network model is for transport improvement assessment. To this end, our GIS database is capable of including physical and non-physical projects that are either in the implementation process or of being considered for implementation by the country or an international organization such as the Asian Development Bank. Within the framework of this GIS-based network model, we are able to evaluate the benefits of a single project, or the consummate impacts of multiple projects. Once we have all relevant transport improvements coded in our database, we can even estimate the benefit of the entire investment plan for a complete corridor and even for the nation as a whole.

Important for any far-reaching study is the need for better data. Although data are readily available for aggregate transport demands at a national or even subnational level, absent are more general historical data on modal corridor (i.e. point-to-point) flows. Estimating models and generating forecasts at the corridor level would sharpen aggregate demand estimates and facilitate more detailed investment strategies.

References

Decree of the President of the Republic of Uzbekistan No PP-1446 dated 21 December 2010, “On Acceleration of development of infrastructure, transport and Communication construction in 2011-2015”.

Decree of the President of the Republic of Kazakhstan, March 19, 2010, Government program on the accelerate industrial-innovative development of the Republic of Kazakhstan, 2010-2014, Astana, 2010

Asian Development Bank, 2003, Central Asia: Reassessment of the Regional Transport Sector Strategy, by Ian Jenkins and Paul Pezant

Transport Sector Road Map for Uzbekistan, 2012, by Patrick McCarthy and Jiawen Yang

Transport Sector Road Map for Kazakhstan, 2012, by Patrick McCarthy and Jiawen Yang

Asian Development Bank, 2008, STRATEGY 2020: The Long-Term Strategic Framework of the Asian Development Bank 2008–2020, Mandaluyong City, Phil.: Asian Development Bank.

Asian Development Bank, 2011, CAREC 2020: A Strategic Framework for the Central Asia Regional Economic Cooperation Program for 2011-2020

Asian Development Bank, 2005, CENTRAL ASIA REGIONAL ECONOMIC COOPERATION: Harmonization and Simplification of Transport Agreements, Cross Border Documents and Transport Regulations Prepared by Mr. Michael Sims, consultant report for ADB East and Central Asia Department Transport and Communications Division.

Asian Development Bank, 2009, CAREC Institute Foundation Study: Transport and Trade Facilitation

Asian Development Bank, 2009, CAREC Transport and Trade Facilitation: Partnership for Prosperity, Printed in the Philippines.

Central Asia Regional Economic Cooperation, 2011, CAREC Corridor Performance, Measurement and Monitoring: Annual Report (January 2010 to December 2010)

Central Asia Regional Economic Cooperation, 2011, CAREC Corridor Performance, Measurement and Monitoring: 2011 Q1 report

Central Asia Regional Economic Cooperation, 2011, CAREC Corridor Performance, Measurement and Monitoring: 2011 Q2 report