

2011

# System for Transit Performance Analysis Using the National Transit Database

Albert Gan

*Department of Civil and Environmental Engineering, Florida International University, [gana@fiu.edu](mailto:gana@fiu.edu)*

Feng Gui

*Department of Civil and Environmental Engineering, Florida International University*

Li Tang

*Department of Civil and Environmental Engineering, Florida International University, [ltang001@fiu.edu](mailto:ltang001@fiu.edu)*

Follow this and additional works at: [http://digitalcommons.fiu.edu/cee\\_fac](http://digitalcommons.fiu.edu/cee_fac)



Part of the [Engineering Commons](#)

---

## Recommended Citation

Gan, Albert, et al. 2011. System for Transit Performance Analysis Using the National Transit Database. *Journal of Public Transportation*, 14 (3): 87-107. DOI: <http://dx.doi.org/10.5038/2375-0901.14.3.5> Available at: <http://scholarcommons.usf.edu/jpt/vol14/iss3/5>

This work is brought to you for free and open access by the College of Engineering and Computing at FIU Digital Commons. It has been accepted for inclusion in Department of Civil and Environmental Engineering Faculty Publications by an authorized administrator of FIU Digital Commons. For more information, please contact [dcc@fiu.edu](mailto:dcc@fiu.edu).

# System for Transit Performance Analysis Using the National Transit Database

Albert Gan, Feng Gui, and Li Tang  
Florida International University

## Abstract

*The National Transit Database (NTD) includes comprehensive data on transit organization characteristics, vehicle fleet characteristics, revenues and subsidies, operating and maintenance costs, vehicle fleet reliability and inventory, services consumed and supplied, and safety and security. Some of these data have been used extensively to derive values for transit performance measures and have become the sole source of standardized and comprehensive data for use by all constituencies of the U.S. transit industry. An important application of NTD data has been in trend analysis, which requires multiple years of data. However, accessing NTD data, especially for multiple years, has not been an easy process. One reason is because the data were collected and distributed annually in separate files. This paper introduces a web-based system that integrates over 20 years of NTD data and provides user-friendly tools designed to facilitate the access and analysis of transit performance data.*

## Introduction

Performance analysis can help transit agencies to more objectively evaluate the performance of their systems, thus allowing them to better identify and prioritize problem areas for management actions (Gan et al. 2004). Whether a system is to be analyzed by looking at its performance trends over the years or by comparing

its performances with those of other peer systems, a pre-requisite to these analyses is that the required data be available. In the United States, the National Transit Database (NTD) is increasingly being used for this performance analysis. Collected and distributed by the Federal Transit Administration (FTA), NTD has become the sole source of standardized and comprehensive data for use by all constituencies of the U.S. transit industry (Lyons and Fleischman 1991). NTD includes data on transit organization characteristics, vehicle fleet characteristics, revenues and subsidies, operating and maintenance costs, vehicle fleet reliability and inventory, services consumed and supplied, and safety and security. These data not only provide direct information on the transit systems, but can also be used to derive values for many useful performance measures such as farebox recovery, operating expense per passenger trip served, average speed, etc.

While the majority of the NTD data collected are made available by FTA to the public, access to these data was not easy. This can be attributed to four limitations:

1. NTD data are collected and distributed annually in separate files. To perform a trend analysis for a specific performance measure, one must learn about the file structures, which can vary from one year to another, identify the correct variables from the long list of NTD variables, and manually extract the corresponding data values from the specific files for each of the selected transit systems. For a 10-year trend analysis, for example, this process must be repeated 9 times. The process not only is time-consuming, but also is prone to human errors.
2. NTD data are reported and provided in the most original form while many performance measures must be calculated from the origin data. For example, average speed can be estimated by dividing the total actual revenue miles by the total actual revenue hours. The calculation will become more complicated and time-consuming with when a performance measure involves more than two variables, especially when these variables come from different NTD forms and thus, different files. In addition, the burden is also on the users to be knowledgeable about the correct formula to use and the correct variables to select from the files (Gan et al. 2002).
3. Many NTD variables are reported for individual transit modes and service types (i.e., directly operated or purchased transportation), making it difficult to compare transit performance for the entire agency or state. For example, an analyst may want to know the performance of a transit system without

differentiating between transit vehicles operated directly by the agency itself or by a subcontractor. Similarly, a state official may be interested in statewide transit performance in addition to those of individual transit agencies. Aggregating data is a very time-consuming process, and the user is burdened with using the correct methods of aggregation and selecting the correct variables to use.

4. There was no system that could provide the tools necessary to quickly select the input data and to perform analysis using these data.

This paper describes a web-based system designed to overcome these limitations. It allows the user to focus on the analysis rather than the data preparation and data access. The system is known as the Integrated National Transit Database Analysis System (INTDAS) and is a component program of the Florida Transit Information System (FTIS). The specific development efforts of the INTDAS system involve (1) integrating the NTD data from multiple years to form a single database that allows single access to multiple-year data, (2) developing a set of major performance measures from NTD, (3) aggregating agency-wide and statewide NTD performance data, and (4) developing a user-friendly interface and analysis tools for easy data access and analysis. The remainder of this paper elaborates each of these efforts.

## **Data Integration**

As listed in Table 1, the latest NTD data are reported on a set of 21 standard forms covering 7 different reporting areas, referred to as modules in NTD. Among these forms, Forms B-10, B-20, B-30, F-10, F-20, and F-40 are reported for the entire transit agencies, while the other forms are reported for individual transit modes (motor-bus, heavy rail, etc.) and/or service types. A service type can be either DO (directly operated), if the service is operated in-house by the transit agency itself, or PT (purchased transportation), if the service is subcontracted out to a service provider(s). When a service is subcontracted out, the NTD data may be reported either by the transit agency or by the subcontractor(s).

**Table 1. NTD Reporting Forms**

Module Name	Form Name	Form Title Description
Basic Information	B-10	Transit Agency Identification
	B-20	Transit Agency Contacts
	B-30	Contractual Relationship
Financial	F-10	Sources of Funds—Funds Expended and Funds Earned
	F-20	Uses of Capital
	F-30	Operating Expenses
	F-40	Operating Expenses Summary
Asset	A-10	Stations and Maintenance Facilities
	A-20	Transit Way Mileage
	A-30	Revenue Vehicle Inventory
Service	S-10	Transit Agency Service
	S-20	Fixed Guideway Segments
Resource	R-10	Employees
	R-20	Maintenance Performance
	R-30	Energy Consumption
Federal Funding Allocation Statistics	FFA-10	Federal Funding Allocation Statistics
Safety and Security	SS-10	Safety and Security Setup
	SS-20	Ridership Activity
	SS-30	Safety Configuration
	SS-40	Major Incident Reporting
		Non-Major Incident Reporting

From the perspective of database structure, NTD variables generally can be divided into two categories. The first involves variables that are used to identify the transit agencies and the transit systems (or modes) that they operate. These are herein referred to as the system variables. They include a unique four-digit ID that is used to identify a transit agency, a mode code, and a service type. The second category of variables consists of over 1,000 data attributes that are used to describe the various characteristics associated with each transit agency. They include such variables as service area population, unlinked passenger trips, operating expenses, etc.

Because NTD data are stored and distributed annually on different data files, the first step in the database development process was to combine these data files from different years for each form. Because multiple years of NTD data were inte-

grated into a single data table, a fourth system variable must be added to every data record to identify the corresponding NTD data year. Combining data from different years was a major undertaking mainly because the NTD data have continued to evolve over the years, resulting in inconsistent data variables and file structures from year to year. These data must be reconciled before they could be combined.

Changes to the NTD data over the years have included NTD reporting forms that were added, dropped, combined, or restructured. In general, data were reconciled to match the latest version of the NTD forms. The reconciliation process included rearranging variables from different years, combining variables from multiple forms, moving variable from one form to another, standardizing inconsistent mode codes used in various years, standardizing measurement units, etc. Once the data from different years were reconciled and integrated, they were imported into a SQL Server database. The current database includes NTD data from 1984 through 2009 for most of the NTD forms and represents the most comprehensive database available for NTD data.

## **Performance Measure Development**

After the NTD data were integrated, performance measures could be developed from these data. Below is a list of 20 commonly-used performance measures that are included in INTDAS (CUTR 2000):

1. **Average Age of Fleet.** This is a service quality measure based on the age of the vehicle fleet. It is derived by first multiplying the total number of active vehicles of each fleet of the same mode code and service type with their years in service (i.e., NTD reporting year subtracted by the year of manufacture), then summing up the totals from all fleets. The final number is then obtained by dividing this sum by the number of total active vehicles in all fleets. All of the variables involved in the calculations come from Form A-30.
2. **Average Fare.** This is an indicator of the average level of fare charged to transit riders and is calculated as the passenger fare revenues from Form F-10 divided by the total number of unlinked passenger trips from Form S-10.
3. **Average Speed.** This is the average speed of vehicles in revenue service operation (i.e., not including travel to and from the garage or any other deadhead) and is calculated by dividing the total actual vehicle (for non-rail modes) or train (for rail modes) revenue miles by the total actual vehicle/train revenue hours. Both of the variables come from Form S-10.

4. Average Headway (in minutes). This is an important measure of service frequency. It is computed by first dividing the total directional route mileage from Form S-10 by the system's calculated average speed, as defined above, to obtain an estimate of the number of hours it takes to traverse the entire system's total route miles. This time (in hours) is then divided by the system's average weekday total vehicles from Form S-10 to determine the amount of time in hours it takes for a vehicle to complete its portion of the total route miles, one time. The resulting time is then multiplied by 60 for conversion from hours to minutes.
5. Average Trip Length. This is the average trip length for all passenger boardings and is calculated as the total passenger miles divided by the total unlinked passenger trips. Both variables come from Form S-10.
6. Farebox Recovery. This is an indicator of the share of the total operating costs that is covered by the passenger fares. It is calculated by dividing the passenger fare revenues from Form F-10 by the total modal (operating) expenses from Form F-30. The resulting number is multiplied by 100 to express the share as a percent.
7. Operating Expense per Peak Vehicle. This is a measure of the resources required per vehicle to have one vehicle in operation for a year. It is calculated as the total modal expenses from Form F-30 divided by the number of vehicles operated in maximum service (i.e., peak vehicle) from Form S-10.
8. Operating Expense per Passenger Trip. This is a key indicator of the cost efficiency of transporting riders and is calculated as the total modal expenses from Form F-30 divided by the total unlinked passenger trips from Form S-10.
9. Operating Expense per Revenue Hour. This is a measure of the cost efficiency with which service is delivered. It is calculated as the total modal expenses from Form F-30 divided by the total actual vehicle or train revenue hours from Form S-10.
10. Operating Revenue per Operating Expense. This is a measure of how much the total operating expenses are covered by the total operating revenues, which include both directly generated fare and non-fare revenues. It is calculated as the ratio of the total operating revenue (including passenger fares, special transit fares, school bus service revenues, freight tariffs, charter service revenues, auxiliary transportation revenues, subsidy from other

sectors of operations, and non transportation revenues) reported in Form F-10 and the total modal expenses from Form F-30.

11. **Passenger Trips Per Service Area Capita.** This is a measure of the extent to which the public utilizes transit in a given service area and is calculated by dividing the unlinked passenger trips from Form S-10 with the total population living in the transit service area from Form B-10. The total service area population is usually taken as the sum of all population living within 0.75 mile of transit routes. However, because the number is reported for all modes operated by an agency, agencies that operate a demand response mode in a county may report the entire county population, making the measure difficult to compare across different modes and/or agencies. Accordingly, this measure should only be applied within an agency for comparisons across different years, i.e., trend analysis.
12. **Passenger Trips per Revenue Hour.** This is a key indicator of service effectiveness that is influenced by the levels of demand and the supply of service provided and is calculated by dividing the total unlinked passenger trips from Form S-10 with the total actual vehicle or train revenue hours, also from Form S-10.
13. **Revenue Miles per Vehicle Mile.** This is a measure that reflects how much of the total vehicle operation is in passenger service. It is calculated as the total actual revenue in vehicle/train miles divided by the total actual vehicle or train miles. All variables come from Form S-10.
14. **Revenue Miles between Failures.** This is an indicator of the average frequency of delays due to a problem with the equipment and is calculated as the total actual vehicle or train revenue miles from Form S-10 divided by the total revenue vehicle system failures from Form R-20.
15. **Revenue Miles per (Total) Vehicle.** This is measure of the level of vehicle utilization in terms of total revenue miles driven. It is calculated as the total actual vehicle or train revenue miles divided by the total vehicles available for maximum service. All variables come from Form S-10.
16. **Revenue House per (Total) Vehicle.** This is a measure of the level of vehicle utilization in terms of total revenue hours driven. It is calculated as the total actual vehicle or train revenue hours divided by the total vehicles available for maximum service. All variables come from Form S-10.



17. Revenue Hours per Employee. This is a measure of the overall labor productivity and is calculated as the ratio of the total actual vehicle/train revenue hours of service from Form S-10 and the total full-time employees from Form R-10.
18. Spare Ratio. This is an indicator of the level of spare vehicles available for service. It is calculated as the difference between vehicles operated in maximum service and vehicles available for maximum service, divided by vehicles operated in maximum service. Both variables come from Form S-10.
19. Vehicle Miles per Gallon or Kilowatt-Hour. This is an efficiency measure of energy utilization and is calculated as the total actual vehicle or train miles from Form S-10 divided by the total gallons or kilowatt-hour consumed from Form R-30.
20. Weekday Span of Service. This measure reports the number of hours that transit service is provided on a representative weekday. It is determined by calculating the number of hours between the average weekday time service begins and time service ends reported on Form S-10.

The values for these performance measures are pre-calculated and stored in the system, and can be quickly retrieved by the users for performance analysis.

## **Aggregated Performance Measures**

Performance analysis does not necessarily have to be performed at the transit system level. Measures also can be compared at the state, regional, national, or any other geographic level of interest. For example, a state Department of Transportation (DOT) may be interested in comparing its statewide performance with the other peer states. For such comparisons, data collected at the transit system level must be aggregated to the specific level of interest. The system currently provides the aggregated performance measures for: (1) both service types combined, (2) all modes combined, (3) all rail modes combined, (4) all non-rail modes combined, (5) all fixed-route modes combined, (6) all transit systems within a state combined (i.e., statewide), and (6) all combinations of the above aggregates.

For illustration purposes, assume that a transit agency operates two modes, A and B. If a performance measure involves two variables, V1 and V2, and the formula for the performance measure calculation is  $V1/V2$ , the aggregated value for both modes will be calculated as  $(V1A + V1B) / (V2A + V2B)$ . It is obvious that one major issue with such aggregation has to do with the treatment for missing and incorrect

values. A missing or incorrect value for one of the transit systems involved in the aggregation will have a direct impact on the final aggregated value.

In the case of a missing value, no aggregated values will be calculated when there is a missing value. This rule is applied on the basis that it is better to provide no data than to provide inaccurate or wrong data. A missing value simply could be because the number was not reported by the agency, or the agency is not required to report for a specific mode or service type. For example, for Revenue Hours per Employee, the measure is not calculated for the combined service types (i.e., DO + PT) because reporting of employee information is not required for purchased transportation.

In the case of potentially incorrect data, the detection is more difficult, as it is generally more difficult to look at a value and be able to tell if it is incorrect. However, an incorrect value is likely to be identified more easily after it is used in a formula. For example, an incorrect low value reported for the total actual revenue hours will cause the average speed to fall outside the normal speed range, e.g., a bus system with an average speed of over 80 mph. To further detect incorrect data, a major effort was taken to examine each calculated value by visually identifying sudden changes in the annual data trend. For example, it will be quite unlikely for a system to have its average speed jump by 10 mph from one year to another. Fortunately, INTDAS provides an interactive charting function (described in the next section) that can greatly facilitate this task. All calculated values that were deemed suspicious were subsequently removed from the database.

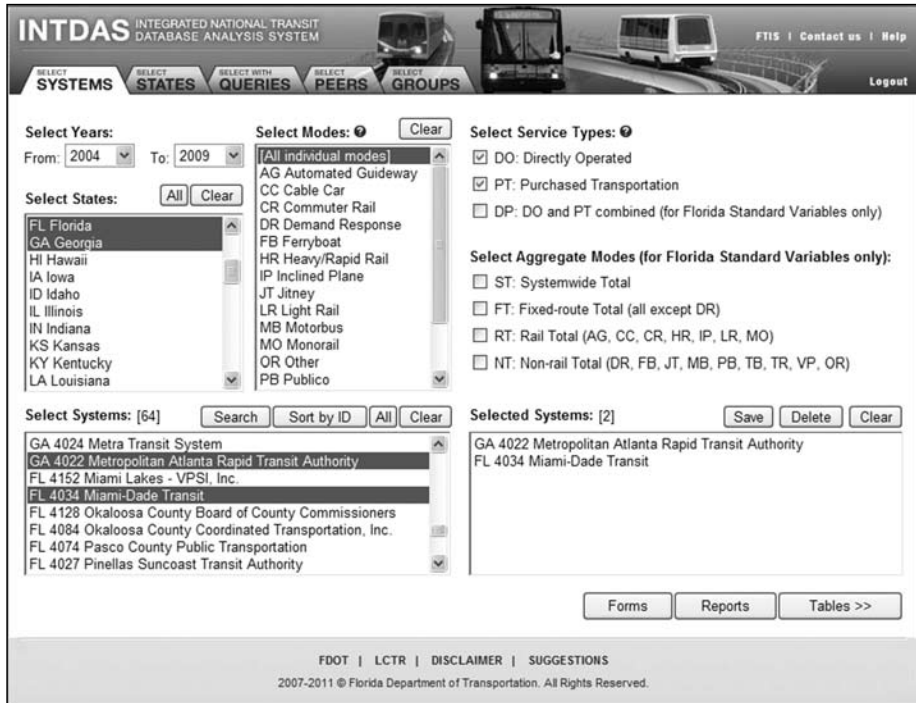
## **User Interface and Functionalities**

This section introduces the functionalities of INTDAS via three case applications. The system is freely accessible at <http://www.ftis.org/intdas.html> upon user registration. The use of the web platform for the system not only facilitates user access, but also allows the developers to update the system quickly and as frequently as needed.

### ***Case Application 1: Quick Information Lookup for Known Agencies***

In this case application, the user would simply like to look up information for two major bus systems, the Miami-Dade Transit (MDT) and Metropolitan Atlanta Rapid Transit Authority (MARTA), located in the Southeast region of the United States. The user has registered to access the system, so she can quickly log on to the system using her assigned password. Upon entering the system, the user is

first shown the main screen in Figure 1. On this screen, INTDAS selects the latest six years of NTD data, by default, for all individual transit modes, and all individual service types. The user can, however, choose to retrieve any years of data between 1984 and 2009 for any specific modes and service types.



**Figure 1. Screen for selecting transit systems**

To select the two transit systems, the user clicks to select Florida and Georgia on the Select States list box to quickly shortlist the number of transit systems to only those in Florida and Georgia. The shortlisted systems are displayed in the Select Systems list box. The user then can identify the desired systems from the shortlist simply by clicking on the agency names. Selected transit systems are listed in the Selected Systems list box. Once the transit systems of interest are selected, data can be retrieved and displayed either on the NTD forms or on some pre-defined standard reports by clicking the Forms and Reports buttons, respectively, at the bottom right of the screen. Figure 2 displays some service related data on the top portion of the S-10 form. Figure 3 shows a standard report that includes six years of trend data for MDT for a set of efficiency measures. The user can click on any

of the list boxes to get the same trend table for other system combinations, e.g., a directly operated heavy rail system for MARTA.

Transit Agency Service (S-10)										
Year	NTD ID	Location	State	Mode	Type of Service					
2009	4034	Miami	FL	MB	DO					
Item	Average Weekday				Average Weekday	Average Saturday	Average Sunday	Annual Total		
	AM Peak	Midday	PM Peak	Other	Total	Total	Total			
<b>Maximum Service Vehicles</b>										
01. Vehs operated in maximum service										716
02. Vehs available for maximum service										863
<b>Periods of Service</b>										
03. Time service begins	06:30	10:30	14:00		00:00	00:00	00:00			
04. Time service ends	10:30	14:00	19:00		24:00	24:00	24:00			
<b>Service Supplied</b>										
05. No. of veh/train in operation	687	448	716	448	716	341	290			
06. Total actual veh/train miles					113,983	70,239	72,742			37,092,499
07. Total actual veh/train hours					8,924	5,273	5,386			2,874,681
08. Total actual veh/train rev. miles					96,299	61,352	63,297			31,547,096
09. Total actual veh/train rev. hours					8,140	4,888	4,975			2,629,625

Figure 2. Retrieved data displayed on an emulated NTD form

Analysis:		Report:		System:		Mode Code:		Service Type:	
<input checked="" type="radio"/> Trend		<input checked="" type="radio"/> General Performance Indicators		4022 Metropolitan Atlanta Rapid Transit Authority		AG		DO	
<input type="radio"/> Peer	2009	<input type="radio"/> Effectiveness Measures		4034 Miami-Dade Transit		DR		PT	
<input type="radio"/> << Back	Page 1	<input type="radio"/> Efficiency Measures				HR			
						MB			
<b>Miami-Dade Transit</b>									
<b>Directly Operated Motorbus</b>									
EFFECTIVENESS MEASURES	2004	2005	2006	2007	2008	2009	% Change 2004-2009	% Change 2008-2009	% Change 1984-2009
<b>SERVICE SUPPLY</b>									
Vehicle Miles Per Capita	15.36	16.77	18.02	17.49	16.28	15.44	0.32%	-5.17%	n/a
<b>SERVICE CONSUMPTION</b>									
Passenger Trips Per Capita	32.03	32.25	34.30	34.74	35.71	31.47	-1.73%	-11.87%	n/a
Passenger Trips Per Revenue Mile	2.42	2.24	2.22	2.34	2.57	2.40	-0.8%	-6.67%	-22.95%
Passenger Trips Per Revenue Hour	29.63	28.09	27.67		31.17	28.75	-2.96%	-7.74%	-27.11%
<b>QUALITY OF SERVICE</b>									
Average Speed (RM/RH)	12.26	12.53	12.48	12.20	12.14	12.00	-2.18%	-1.15%	-5.4%
Average Headway (in minutes)	13.05	12.26	11.44	11.33	n/a	12.90	-1.12%	n/a	n/a
Average Age of Fleet (in years)	4.62	4.77	4.53	5.16	5.36	6.31	36.46%	17.7%	n/a
Number of Incidents	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Number of Vehicle System Failures	13,097	15,800	18,951	15,240	15,926	13,933	6.38%	-12.51%	-55.53%
Revenue Miles Between Incidents	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Revenue Miles Between Failures	2,374.63	2,164.89	1,943.19	2,338.30	2,097.66	2,264.20	-4.65%	7.94%	245.52%
<b>AVAILABILITY</b>									
Revenue Miles Per Route Mile	17,587.78	17,800.13	19,085.46	18,448.00	n/a	17,079.07	-2.89%	n/a	n/a
Weekday Span of Service (in hours)	n/a	24.00	24.00	24.00	24.00	n/a	n/a	n/a	n/a
Route Miles Per Square Mile of Service Area	6.16	6.28	6.31	6.32	n/a	6.04	-2.03%	n/a	n/a

Figure 3. Retrieved data displayed on a standard report

### Case Application 2: Quick Information Lookup for Unknown Agencies

Unlike the previous application, the user in this application is not looking for information for some specific systems in mind, but any systems that meet a specific profile. In this case, the user would apply the query capability of the system by clicking the Queries tab on the main screen to open the screen shown in Figure 4. This screen provides the user with a user-friendly query editor to quickly construct SQL (Structured Query Language) queries. The top half of the query screen allows the user to specify the data years, states, mode codes, and service types. The bottom half of the screen allows the user to construct SQL query conditions that help to narrow down the system selection.

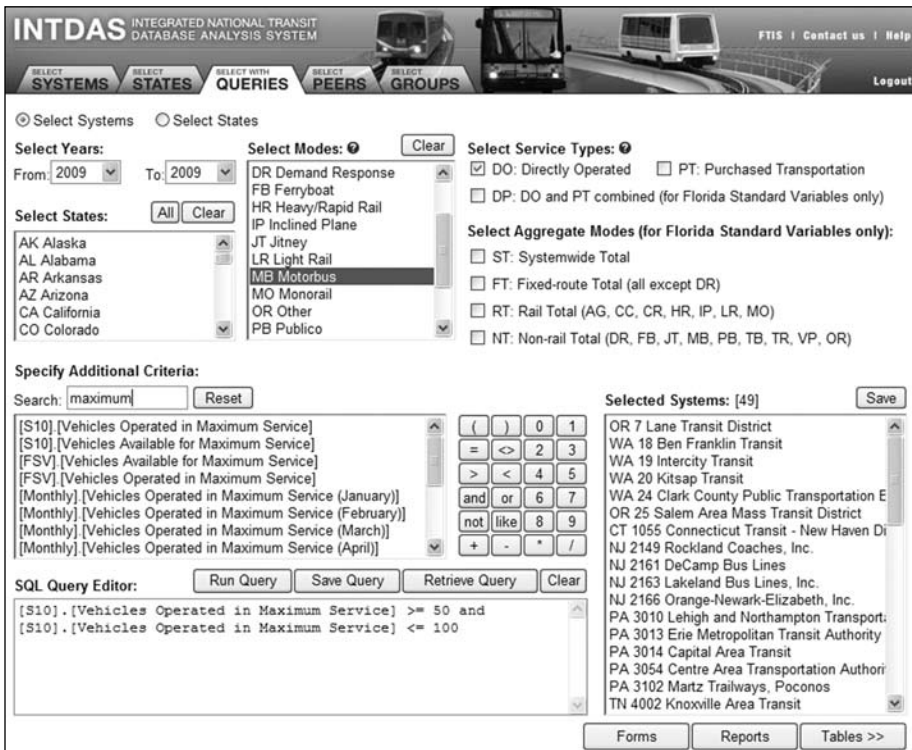


Figure 4. System selection by query

To add a query condition, the user first enters a keyword to narrow down the list of data variables. The user then clicks to select a variable from the variable list box and send it to the Query Editor box. A set of buttons are provided for selecting such common operators as “=” or “>=”. Using a combination of the variable list, the math operator buttons, and the keyboard, the user can construct a series of SQL conditions. For example, the query shown in the Query Editor box in Figure 4 is to

retrieve all directly operated (i.e., “DO”) bus transit systems (i.e., “MB”) in 2009 that operate a fleet size between 50 and 100 motorbus vehicles.

Once a query is constructed, the user clicks on the Run Query button to run the query to find transit systems that meet the query conditions. The output transit systems are listed in the Selected Systems list box. The user then looks up the information for these systems, as in the previous application, by clicking on the Forms or Reports button.

### Case Application 3: Peer Selection and Comparisons

In this case application, the user is a transit planner from MDT who is interested in comparing the performances of MDT’s bus system with those of its peer systems. As a first step of the analysis, the user needs to identify peer systems that are deemed comparable to MDT. To do so, the user first clicks the Peers tab on the main screen to open the screen shown in Figure 5. The screen allows the user to choose between two peer selection methods: TCRP or FTIS. The FTIS method was implemented prior to the development of the new TCRP method and is expected to be phased out. By default, the TCRP method, developed as part of the Transit Cooperation Research Program (TCRP) Project G-11 (Ryus et al. 2010), is selected.

INTDAS INTEGRATED NATIONAL TRANSIT DATABASE ANALYSIS SYSTEM

SELECT SYSTEMS SELECT STATES SELECT WITH QUERIES SELECT PEERS SELECT GROUPS

FTIS | Contact us | Help

Logout

TCRP Peer Selection  FTIS Peer Selection

Select Target Agency:

FL Florida [Search]

- 4034 Miami-Dade Transit
- 4035 Central Florida Regional Transportation Authority
- 4036 City of Tallahassee
- 4037 Board of County Commissioners, Palm Beach County, PalmTran, Inc.
- 4038 Escambia County Area Transit
- 4040 Jacksonville Transportation Authority
- 4041 Hillsborough Area Regional Transit Authority
- 4046 Sarasota County Area Transit
- 4050 Smyrna Transit System
- 4063 Space Coast Area Transit
- 4072 City of Key West - Department of Transportation

Select Benchmarking Level:

Agency-wide

Specific Mode: MB Motorbus

Select Data Year: 2006

Specify Data Source for Urban Area Population:

2006 American Community Survey (ACS)

User-specified: [ ]

Find Peers

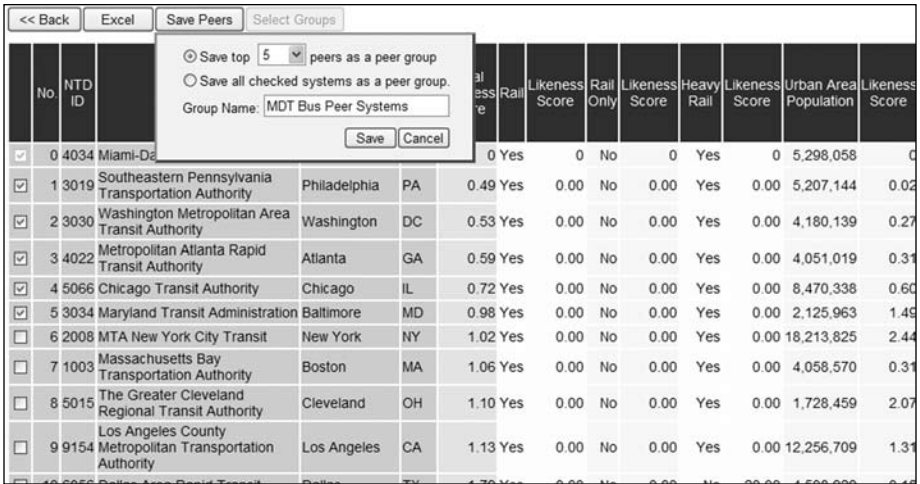
Figure 5. Peer selection input screen

The input screen for the TCRP method allows the user to first select Florida to shortlist the transit systems and then selects “Miami-Dade Transit” from the shortlist. The user can then choose to select peer systems based on agency-wide or mode-specific mode statistics for peer variables. In this application, the mode-specific option based on bus transit system (MB) is selected. The user then selects the data year for peer selection variables and the data source for the population size of the urban area in which the MDT is located. Next, the user selects the data year. As of the writing of this paper, only the 2006 data are available in the system for this peer selection method. Lastly, the user selects the data source for urban population. By default, the urbanized area (UZA) population from the American Community Survey (ACS) is used. However, the user may enter another population size. This allows the user to test different “what if” population scenarios. Once all of these inputs are specified, the user clicks the Find Peers button to search for peer systems based on their likeness scores, which are calculated from 3 screening factors and up to 14 peer grouping factors.

The three screening factors are whether an agency operates a rail system, only a rail system(s), and/or a heavy rail system. These variables ensure that potential peers operate a similar mix of modes as MDT. The peer grouping factors are used to determine which potential peer agencies are most similar to MDT in terms of service characteristics (e.g., vehicle miles operated, annual operating budget) and urban area characteristics (e.g., population density, percent low income). Based on these factors, a total likeness score is calculated for each potential peer system, as follows (Ryus et al. 2010):

$$\text{Total likeness score} = \frac{\text{Sum (screening factor scores)} + \text{Sum (peer grouping scores)}}{\text{Count (peer grouping factors)}}$$

Based on this equation, the lower the score of a potential peer system, the more similar it is to MDT. Figure 6 shows the output screen, which lists all potential agencies from the lowest to the highest scores. In general, a total likeness score under 0.50 indicates a good match, between 0.50 and 0.74 represents a satisfactory match, and between 0.75 and 0.99 represents a potential match that may be used with additional investigation to determine major differences that may make them unsuitable. Scores beyond 0.99 are considered undesirable. Readers are referred to TCRP Report 141 for details on the peer selection process implemented in INTDAS (Ryus et al. 2010).



**Figure 6. Peer output screen**

Figure 6 also shows that the user can select the desired peer systems and then click the Save Peers button to save the peer systems for later use as a peer group. By default, the five most similar systems are selected. The user may, however, select any number of desired peer systems by checking the box in front of each listed system. The user then enters a group name for the set of peer systems selected to save. After the systems are saved, the user clicks on the Select Groups button to open the screen shown in Figure 7. This screen can also be opened by clicking the Groups tab on the main screen directly. On this screen, the user selects the peer systems she just saved by clicking on the group name. The saved peer systems in the group are displayed in the Select Systems window. The user can then select the peer systems by clicking the All buttons or clicking on the specific systems. The user then clicks on the Tables button to select the calculated performance variables as described above. This opens the screen shown in Figure 8, which allows the user to select specific data variables for retrieval and analysis.



Figure 7. Screen for selecting saved peer groups

Figure 8. Screen for selecting data variables

INTDAS provides the following two selection options for variable selection:

1. Variable selection by list: This option allows the user to enter a keyword, or partial keyword, to search for variables that match the keyword. All variables that match the keyword are listed in the list box, as shown in Figure 8.
2. Variable Selection by NTD form: The second option for variable selection is through the emulated NTD forms. To select a variable, the user first selects the appropriate NTD form and then clicks on a check box to select a specific variable. Figure 9 shows that the users selected the total modal operating expenses on Form F-30.

Operating Expenses (F-30)																	
Year		<input type="checkbox"/>	NTD ID		<input type="checkbox"/>	Location		<input type="checkbox"/>	State		<input type="checkbox"/>	Mode		<input type="checkbox"/>	Type of Service		<input type="checkbox"/>
Expense Object Class	Function				Total Expenses for Period												
	Vehicle Operations 010	Vehicle Maintenance 041	Non-Vehicle Maintenance 042	General Administration 160													
<b>Labor (501)</b>																	
01. Operator's salaries and wages (01)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>												
02. Other salaries and wages (02)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>												
03. Fringe Benefits (502)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>												
04. Services (503)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>												
<b>Materials and Supplies (504)</b>																	
05. Fuel and lubricants (01)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>												
06. Tires and tubes (02)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>												
07. Other materials and supplies (99)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>												
08. Utilities (505)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>												
09. Casualty and Liability Cost (506)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>												
10. Taxes (507)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>												
<b>Purchased Transportation (508)</b>																	
11. In report (01)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>												
12. Filing separate report (02)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>												
13. Miscellaneous Expenses (509)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>												
14. Expense Transfers (510)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>												
15. Total Modal Expenses	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>												

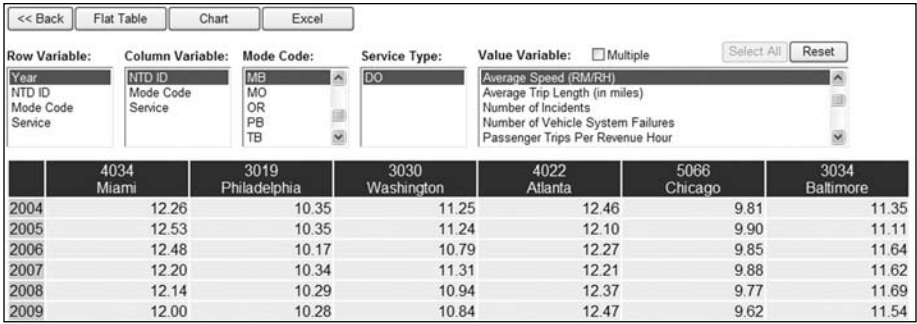
Figure 9. Selecting variables from NTD forms

Once the variables of interest are specified, the data for the selected systems can be retrieved and displayed in a Flat Table, Cross Table, or Chart. Figure 10 shows an example of a flat table. It can be seen that the tenth column gives the total operating expenses after they are adjusted for inflation. The adjusted values could be quickly calculated in INTDAS by applying the Inflation function and specifying a target year to adjust to.

<< Back   Cross Table   Chart   Excel   Sort   Inflation   Summation   Statistics   Regression   Calculate   Reset												
No.	Year	NTD ID	Company Name	Location	State	Mode	Service	Total Modal Expenses	Average Age of Fleet (in years)	Average Headway (in minutes)	Average Speed (RM/RH)	Average Trip Length (in miles)
1	2004	4034	Miami-Dade Transit	Miami	FL	AG	DO	\$18,672,871	13.93	3.33	10.20	1.02
2	2004	4034	Miami-Dade Transit	Miami	FL	HR	DO	\$61,437,722	22.00	5.30	23.58	7.79
3	2004	4034	Miami-Dade Transit	Miami	FL	MB	DO	\$229,427,318	4.62	13.05	12.26	3.95
4	2004	3019	Southeastern Pennsylvania Transportation Authority	Philadelphia	PA	CR	DO	\$186,242,753	28.90	11.40	27.05	14.32
5	2004	3019	Southeastern Pennsylvania Transportation Authority	Philadelphia	PA	HR	DO	\$125,380,076	11.70	4.42	19.56	4.46
6	2004	3019	Southeastern Pennsylvania Transportation Authority	Philadelphia	PA	LR	DO	\$46,088,287	23.59	3.79	9.45	2.51
7	2004	3019	Southeastern Pennsylvania Transportation Authority	Philadelphia	PA	MB	DO	\$400,367,435	6.64	13.20	10.35	2.90
8	2004	3030	Washington Metropolitan Area Transit Authority	Washington	DC	HR	DO	\$525,516,163	17.27	3.97	25.17	6.01
9	2004	3030	Washington Metropolitan Area Transit Authority	Washington	DC	MB	DO	\$395,725,481	8.82	12.43	11.25	2.99
10	2004	4022	Metropolitan Atlanta Rapid Transit Authority	Atlanta	GA	DR	DO	\$10,094,433	2.97		17.89	10.32

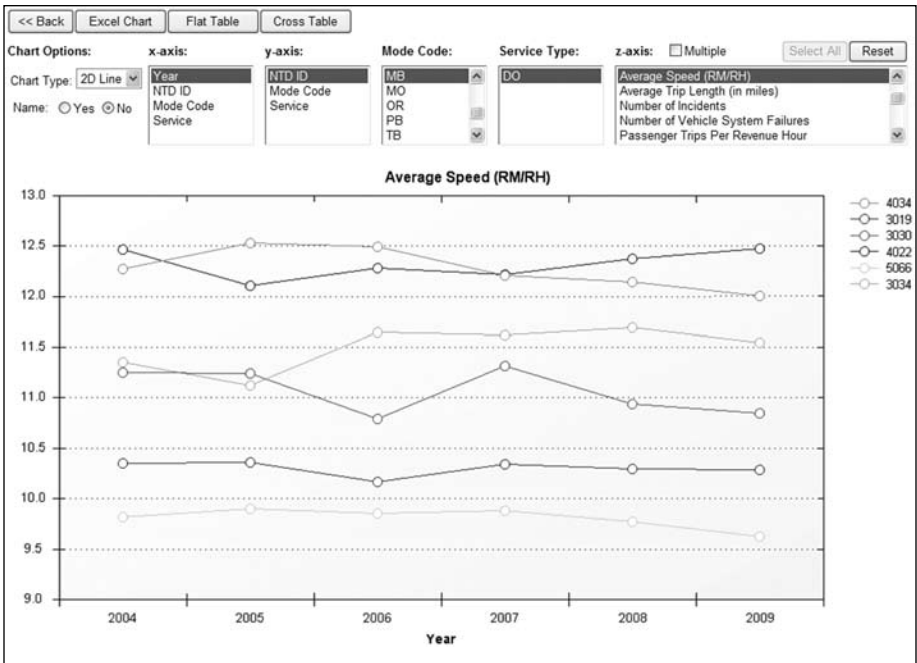
**Figure 10. Retrieved data displayed in a flat table**

It can be seen from Figure 10 that the tabulated data with multiple systems and multiple years are somewhat difficult to read. As such, INTDAS provides the ability to generate cross-table view that is cross-classified by two system variables at a time. Figure 11 shows such an example. The figure shows six years of average speeds for MDT and its five selected peer systems. The five list boxes on top of the cross-table screen allow the variables to be selected in different manners. The first two list boxes define the row and column variables, respectively, while the third and fourth list boxes list all of the possible data items for each of the remaining two system variables. The last list box lists all of the data attributes selected for analysis. The value for the selected data attribute is displayed on the cross table shown under the list box. The cross table is immediately updated as soon as a change occurs in any of the list boxes; this allows different data combinations to be examined quickly.



**Figure 11. Retrieved data displayed in a cross table**

Instead of displaying data in a cross table, the user may also choose to display the data graphically, as shown in Figure 12. In this example, the user can quickly examine 25 years of average speed trends for MDT and compare them with those of the peer systems. Similar to the cross table, the plot is also refreshed immediately after any of the variables in the list boxes are changed, allowing the user to plot charts of all possible variable combinations.



**Figure 12. Retrieved data displayed in a chart**

## **Summary and Further Developments**

The National Transit Database (NTD) is an important database for the transit industry. These NTD data, however, were not easily accessible. Using three case applications, this paper described a web-based system designed to facilitate the retrieval and analysis of NTD data. The system is useful for both practitioners and researchers who use the NTD data to improve transit performance and services. The system provides user-friendly functions that allow multiple years of NTD data to be quickly retrieved and analyzed for multiple transit systems. The data retrieval functions allow the user to quickly select transit systems and NTD variables of interest, retrieve data for the selected systems and variables, and display the data in forms, tables, charts, and reports. The system also includes a peer selection process that allows one to quickly identify peer systems for peer comparisons.

The development of INTDAS remains an ongoing process. The original NTD data and the calculated performance measures will continue to be added to the system annually. While the system is quite user-friendly, an online web training system is being developed to help the users get started quickly and allow them to use the system more effectively. Current planned activities include the addition of rural NTD data, where were first collected in 2006, and the research and implementation of data mining tools to discover patterns and relationships from the vast amount of NTD data collected over the years.

## **Acknowledgements**

The development of the INTDAS system was sponsored by the Public Transit Office of the Florida Department of Transportation. The peer selection feature in the system was implemented as part of Transit Cooperative Research Program (TCRP) Project G-11.

## **References**

- Center for Urban Transportation Research (CUTR). 2000. "Performance evaluation of florida's transit systems." Final Report Prepared for the Florida Department of Transportation.
- Gan, A., I. Ubaka, and F. Zhao. 2002. "Integrated National Transit Database Analysis System (INTDAS)," *Transportation Research Record* 1799: 78-88.

- Gan, A., I. Ubaka, and J. Zheng. 2004. "Automated transit peer selection and analysis," *Proceedings of the 2004 Annual Meeting of the Transportation Research Board*, Washington, DC.
- Lyons, W. M., and E. R. Fleischman. 1992. "New future for the Federal Transit Administration Section 15 Program," *Transportation Research Record* 1349: 18-27.
- Ryus, P., K. Coffel, J. Parks, V. Perk, L. Cherrington, J. Arndt, Y. Nakanishi, and A. Gan. 2011. *A Methodology for Performance Measurement and Peer Comparison in the Public Transportation Industry*, TCRP Report 141, Transit Cooperative Research Program, Transportation Research Board.
- Ryus, P., K. Coffel, J. Parks, V. Perk, L. Cherrington, J. Arndt, Y. Nakanishi, and A. Gan. 2010. "Development of the TCRP G-11 transit agency peer-grouping methodology," *Proceedings of the 90th Annual Meeting of the Transportation Research Board*.

## **About the Authors**

**Albert Gan** (*gana@fiu.edu*) is an Associate Professor with the Department of Civil and Environmental Engineering at Florida International University. He received his Ph.D. in Civil Engineering from the University of Florida in 1996. His areas of research have included traffic simulation, intelligent transportation systems (ITS), transit planning, demand modeling, and highway safety.

**Feng Gui** (*feng.gui@gmail.edu*) has been a System Programmer with the Lehman Center for Transportation Research at Florida International University (FIU) since 2006. He is the chief programmer of the INTDAS system and received his Ph.D. in Computer Engineering from FIU in 2009.

**Li Tang** (*ltang001@fiu.edu*) is a Graduate Research Assistant with the Department of Civil and Environmental Engineering at Florida International University. He received his B.S. and M.S. in Computer Science from Northeastern University of China and is working on his Ph.D. dissertation on transit data mining.