

2002 Commodity Flow Survey Processing Flow of Modal-Mileage Calculation



Presented by Michael Margreta & M. Adhi Dipo

michael.margreta@bts.gov

mohamad.dipo@bts.gov

Bureau of Transportation Statistics
U.S. Department of Transportation

Transportation Research Board
84th Annual Meeting
Washington DC
January 2005



BTS Responsibilities

per agreement with the Census Bureau

- Perform mileage calculations for each freight shipment sampled, by mode (airway, highway, railway, waterway, pipeline)
- Where necessary and if possible, correct missing / inaccurate / inconsistent data (usually destination Zip Code, mode) to obtain a routing that's reasonable and likely
- Where mileages are not obtainable, explain the problem (via coding system) to Census Bureau for possible correction
- Key shipment characteristics affected by modal-mileage calculation:
 - **Ton-Miles;**
 - **Average Miles per Shipment;**
 - **Mode of Transportation (editing capability);**
 - **Distance Shipped (in miles).**

Reasons to Use Routing Models Developed by Oak Ridge National Lab for 2002 CFS Mileage Calculation

- Proven expertise
- Maintain consistency with previous cycles (1993, 1997)
- Availability/adaptability of software
- Time/budget constraints for 2002 survey
- Training available
- Principal Investigator: Dr Frank Southworth
- ORNL Programmers: Dr Chin, B Peterson.

2002 Routing Models

What's New Since 1997



- USA Highway Network: minimal updates
 1. Added HWY penalty of 1 hour for border crossings into Canada & Mexico, due to "bottleneck" traffic analysis, not due to 9-11; Result: USA-only HWY routing more optimal.
 2. Reduced waterway (ferry) penalty for HWY shipments between USA West Coast & Alaska; Result: HWY routing more favorable thru Alaska than thru Canada.

2002 Routing Models

What's New Since 1997 (cont.)



- USA Railway Network: updated for mergers and foreclosures since 1997 to Class 1 railroads (2002 Annual Operating Revenue \geq \$272 Million).
- USA Waterway Network: updated Model of Potential Seaports based on commodity, value, weight.
- USA Airway Network: updated hubs based on Sept 2002 Official Airline Guide.
- Intermodal Transfer Points (terminal locations): updated as reported by FRA.
- For the first time ever, BTS personnel did the day-to-day mileage-calculation processing, which took place at facilities of the Census Bureau in Maryland.



Required Input for Mileage-Calculation Routing

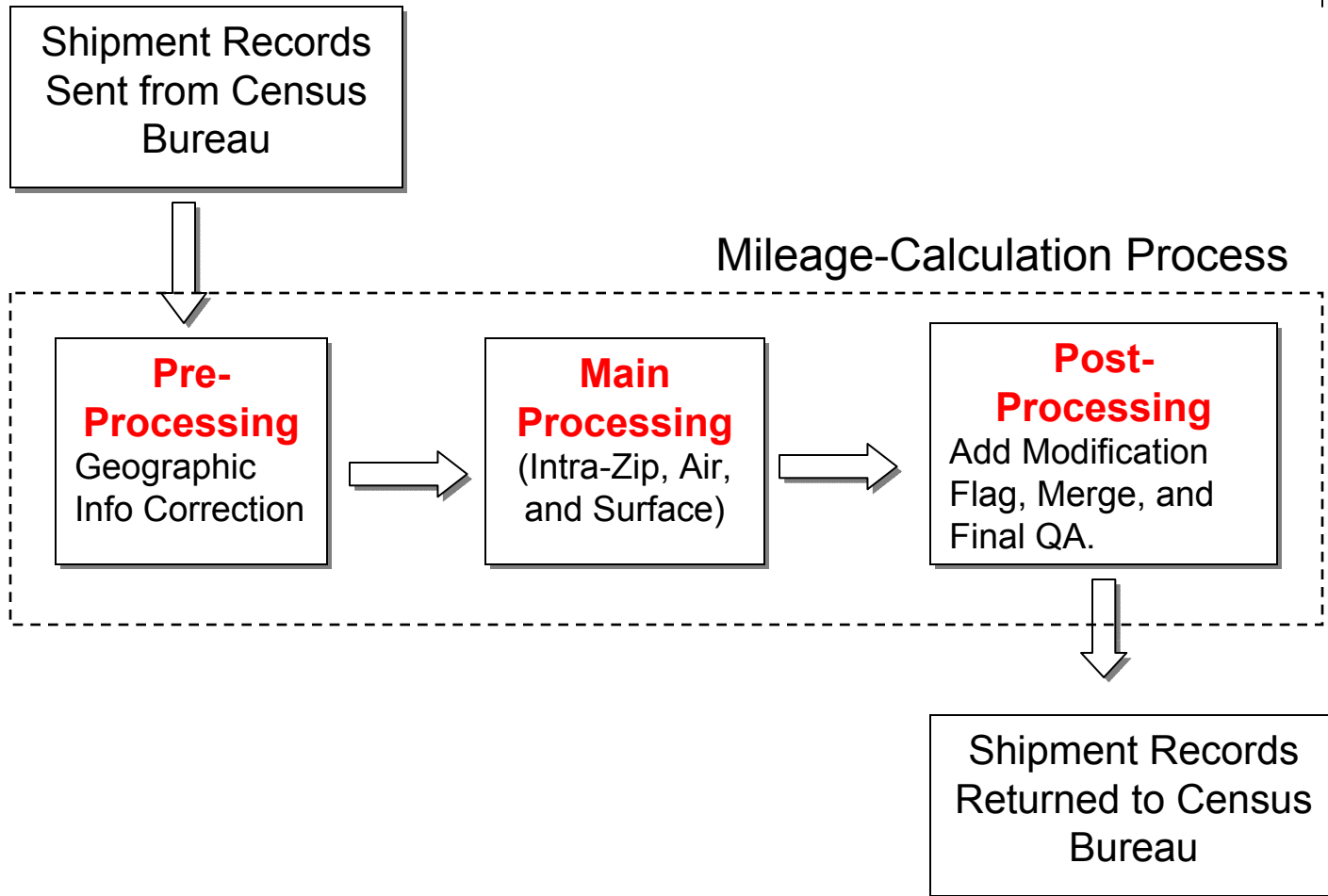
(from responding shipper, as keyed by collection agency)

- Valid Origin Zip Code
- Valid Destination Zip Code; if an export, valid Country Name (valid City Name for Canada and Mexico)
- Mode or Modal Sequence.

Modes of Transport

- Parcel Delivery, Courier, US Postal Service
- Private Truck
- For-Hire Truck
- Railroad
- Shallow-Draft (Inland Water) Vessel
- Deep-Draft (Ocean) Vessel
- Great Lakes Vessel (generated by Routing Models)
- Air
- Pipeline (mileage = Great Circle Distance)
- Unknown = Don't Know
- Other
- Multi-Mode = any combination of the above.

Simplified Data Flow for Mileage Processing



How Good Are the Data Supplied by the Responding Shippers ?

(as keyed by collection agency)



- Of the 2.7 million records processed for the 2002 CFS, about 300,000 records (11.3% of total) required some type of mechanized or manual correction by BTS analysts to produce an acceptable routing.
- About 45,000 records (1.7%) were corrected by Census Bureau analysts, sometimes by means of call-backs to the shippers.

What Resources Are Available To Aid in Correction of Problematic Shipment Records



- Feith Document – a snapshot of the completed survey form (available electronically from Census Bureau);
- DeLorme Street Atlas USA – map software with national network of highways, railways, and waterways;
- DeLorme Earth 'A' Global Explorer – map software with international network of highways, railways, and waterways;
- Freight-transportation experts at Oak Ridge – advice, past experience;
- Pre-processing software from Oak Ridge to identify all locations (State – Zip Code) of a given U.S. city name;
- Support staff at Census Bureau – Internet searches, call-backs to the respondent (last resort).

Pre-Processing Step - What Could Possibly Go Wrong ?

1. Absolutely No Destination Information (missing city, state, zip code) on about 15,000 records with domestic shipments.

Solution:

- a. Call-Back almost always required.
- b. Correct typing omission during keying by collection agency (infrequent).
- c. Limited ability to impute (from earlier quarter in 2002 or 1997).



Pre-Processing Step (cont.)

What Could Possibly Go Wrong ?

2. Destination Zip Code is missing, but City and (sometimes) State are provided on input.

Solution:

Develop a Domestic Place Name File (now containing 32,000+ entries) to insert Zip Code when City – State match the input, which corrected about 64,000 records (2.4% of total).

Examples:

	<u>City</u>	<u>State</u>	<u>Zip Code</u>
[missing state]	Los Angeles		90001
[abbreviations]	L A	CA	90001
	LAX	CA	90045
[common misspellings]			
	Las Angeles	CA	90001
	Los Angels	CA	90001
[place name]	Los Angeles International		90045

Examples of Challenging Cases with Missing Zips



Domestic

Destination (from Shipper) Investigation

Loop LA OCGS	Louisiana Offshore Oil Port \\ both near Morgan Outer Continental Gulf Shelf // City, Zip = 70380 .
SEATAC	Seattle-Tacoma International Airport, Zip = 98158 .
Prtg MI	Portage MI, Zip = 49002 .
SC WA	Snohomish County WA, Zip = 98223 .
Rising Sun AZ	Check DeLorme: no Rising Sun in AZ but in MS; check Feith: sloppy writing, AZ -> AR; revisit DeLorme: Rising Sun on border of MS & AR; use MS Zip = 38930 .



Pre-Processing Step (cont.)

What Could Possibly Go Wrong ?

Examples of Destination Zip Code on input that is invalid.

Note: A Zip Code is considered invalid if it cannot be found in the 2002 Zip Code File purchased from Geographic Data Technology (GDT), which was updated with valid U.S. zips through January, 2002.

3. Transposed Zip Code numbers:
 - Invalid zip for Keller TX **76428** – corrected to **76248**;
 - Invalid zip for Cleveland OH **44163** – corrected to **44136**.

4. Less than five digits for Zip Code:
 - Invalid zip for Danbury CT **6810** – corrected to **06810**;
 - Invalid zip for Selma AL **3670** – corrected to **36701**.

Pre-Processing Step (cont.)

What Could Possibly Go Wrong ?



5. One zip digit miscopied / misprinted / miskeyed:
 - Invalid zip for Woodston KS **64675** – corrected to **67675**;
 - Invalid zip for Brooklyn NY **11200** – corrected to **11201**.
6. Made-up guesses for zip codes:
 - Invalid zip for Manhattan NY **99999** – corrected to **10021**;
 - Invalid zip for Hutchinson MN **55555** – corrected to **55350**.
7. Zip Code not found in 2002 Zip file but valid during 1997:
 - Invalid zip for Chantilly VA **22021** – corrected to **20151**.
8. Zip Code and State not compatible (invalid Zip for given State):
 - Inglewood **CO** 90307 → corrected to Inglewood **CA** 90307 [CO zips are 80+++ or 81+++].

Pre-Processing Step (cont.)

What Could Possibly Go Wrong ?

9. Missing or improperly provided data on about 70,000 records of export shipments (export records were 4½% of total but ⅔ of them required correction).

Solution - Manually correct typical mistakes:

- a. Foreign city / country / mode provided on input as domestic city / state / mode.
- b. Misspellings of foreign country (country abbreviations not acceptable).
- c. Shipments to U.S. possessions (Puerto Rico, Guam, U.S. Virgin Islands) are usually reported as domestic, but are considered exports for CFS.

Pre-Processing Step (cont.)

What Could Possibly Go Wrong ?



10. U.S. Destination Zip Code is missing, but Foreign Country and (sometimes) Foreign City are provided on input. Note: Once an export reaches the U.S. port of exit (POE), be it airport or seaport or highway border crossing into Canada/Mexico, the POE is considered the final domestic destination, the domestic route is finished, and any following mileage is considered international mileage and as such, is not counted from the POE. The Routing Models locate a POE, when missing, depending on foreign destination and commodity shipped.

Solution:

Develop an Export Place Name File (now containing 300+ country codes; 5,700+ Canadian cities/towns; 26,000+ Mexican cities/towns) to provide latitude and longitude positions for foreign routings.

Examples:

	<u>City</u>	<u>Country</u>
[popular cities]	Hong Kong Singapore	
[abbreviations]		ENG G B U K
[common misspellings]	Winnipeg Winnepeg Veracruz Vera Cruz	Canada Canada Mexico Mexico
[multiple names]		England Great Britain United Kingdom

Examples of Challenging Cases of Exports with Missing Zips



Foreign

Destination (from Shipper)

Investigation

Tijuana Mexico

Check spelling (near San Diego).

Mississauga Canada

Check spelling (truck terminal near Toronto).

Johnston Island UM

U.S. main outlying island between Hawaii & Marshall Islands.

Seneffe Europe

Found in Belgium.

IOM

Isle of Man in Irish Sea.

SMD

Specialty Materials Division in Niagara Falls, Canada.

Pillotte Canada

Pillette Road in Windsor, Ontario, Canada.

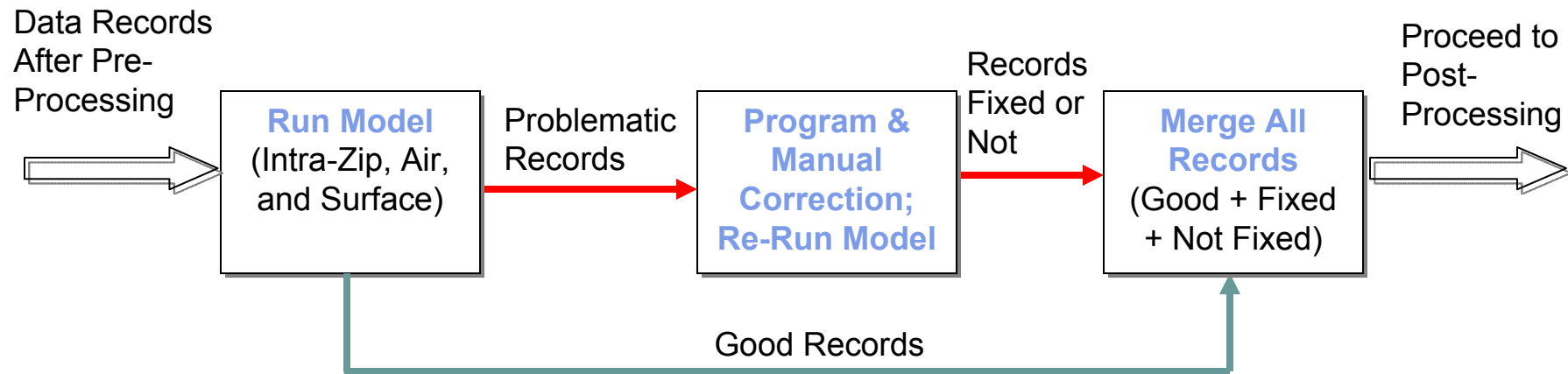
Klaubbe FN

Not located.

Lost Cove Canada

Not located.

Main Processing Step in Data Flow for Mileage Calculation



Main Processing Step for Mileage Calculation



- Run Routing Models from Oak Ridge:
 1. Intra-Zip (areas where one Zip Code area is embedded, either entirely or nearly so, within another Zip Code area, so the shipment routing is relatively short, usually < 20 miles);
 2. Airway (Highway & Air);
 3. Surface (highway, railway, waterway, pipeline).
- Investigate and, if necessary, try to fix any problematic record to obtain a routing that's reasonable and likely.

Examples of Problematic Routings Requiring Investigation / Correction



1. Problematic Freight Shipment via Airway:

Origin Zip: 93901 (Salinas CA)
Destination Zip: 20111 (Manassas VA)
Mode: Airway Only

Mechanized Modal Correction in Airway Routing Model:

Private Truck (Salinas → SFO):	104 miles;
Air (SFO → ORD → IAD):	
2,146 + 701 =	2,847 miles;
<u>For-Hire Truck (IAD → Manassas):</u>	<u>15 miles;</u>
Total Domestic Miles:	2,966 miles.
Great Circle Distance (GCD):	2,268 miles.

Background for Problematic Freight Shipment via Airway



1. (cont.) [Mechanized correction for about 32,000 airway records, 72% of airway total] For domestic routings with a respondent-provided single mode of airway, the Routing Model has been mechanized to automatically add a mode of private truck to the beginning of the route (from origin zip to the sending airport), and then add a mode of for-hire truck to the end of the route (from the receiving airport to destination zip). This same methodology was in use during the mileage calculations for the 1997 CFS, and hence, there was no change for 2002 CFS processing. There was one exception to this methodology in 2002 processing: If a newly manufactured airplane needed transportation from origin zip to location of customer, it was probably flown directly from a private airfield to another airstrip, possibly private. In these cases, the BTS analyst manually adjusted the airway mileage to equal GCD, with no truck (highway) mileage at all in the routing.

Illustration of Mechanized Correction for Problematic Shipment via Airway



Examples of Problematic Routings Requiring Investigation / Correction (cont.)



2. Problematic Freight Shipment via Waterway:

Origin Zip: 55768 (Mountain Iron MN)

Destination Zip: 15056 (Leetsdale PA outside Pittsburgh)

Mode: Waterway Only

Surface Routing Model Output:

Great Circle Distance (GCD): 779 miles;

Commodity: Iron ore, weighing 350,000 pounds;

Error Flag: No access from origin to water (that is, no body of water in the Zip Code area);

Investigation (by BTS freight-mileage analysts):

Check DeLorme Map: Waterway route is reasonable on Great Lakes; determine how to get shipment from origin to a Great Lakes port and then into the destination; railway network is available from origin (Mountain Iron) and into destination (Leetsdale).

Manually Correct Modal Sequence to Rail – Water – Rail.

(cont.)

Examples of Problematic Routings Requiring Investigation / Correction (cont.)



2. (cont.) Problematic Freight Shipment via Waterway:

Investigation (cont.):

Re-run Surface Routing Model.

Routing is plausible – the Model finds Duluth MN as the sending port and Cleveland OH as the receiving port, with the following mileages resulting from the Manual Modal Correction prior to re-run of Surface Routing Model:

Railroad (Mountain Iron —▶ Duluth):	54 miles;
Great Lakes Vessel: (Lake Superior —▶ Lake Huron —▶ Lake Erie)	879 miles;
<u>Railroad (Cleveland —▶ Leetsdale):</u>	<u>132 miles;</u>
Total Domestic Miles:	1,065 miles.
Great Circle Distance (GCD):	779 miles.

Illustration of Manual Correction for Problematic Shipment via Waterway



Examples of Problematic Routings Requiring Investigation / Correction (cont.)



3. Problematic Freight Shipment via Highway:

Origin Zip: 49442 (Muskegon MI)
Destination Zip: 53204 (Milwaukee WI)
Mode: For-Hire Truck Only

Surface Routing Model Output:

Great Circle Distance (GCD):	89 miles;
For-Hire Truck (Muskegon → Milwaukee):	284 miles;
Circuity = (Truck Mileage) / GCD =	3.2

Investigation (by BTS freight-mileage analysts):

Check DeLorme Map: Highway route around geographic barrier (southern tip of Lake Michigan via U.S. Route 31 to Interstate Route 196 to Interstate Route 94).

Routing is plausible – No corrective action necessary.

Illustration of Investigation for Problematic Shipment via Highway



Considerations for Process Improvement of Modal-Mileage Calculation

[subject to available resources (time and money)]



1. Investigate the use of a Geographic Information System (GIS) network to perform mileage calculations for CFS shipments. The GIS network will have an updated highway system that is kept current by the Federal Highway Administration, an updated railway system from the Federal Railroad Administration, an updated waterway system from the Army Corps of Engineers, and updated intermodal transfer points (truck-rail-waterway terminal locations). The use of a GIS network is expected to greatly improve the performance, quality, and reliability of the mileage calculations in the following areas:
 - **Processing Speed.** For the 2002 CFS data, the Surface Routing Model processed at a rate of 1,500 records per minute and the Airway Routing Model at a rate of 300 records per minute.
 - **Consistency of Output.** The output and notification of problematic shipments from the 2 Models were not consistent.
 - **Uniformity of Programming Code.** For the 2002 mileage processing, Fortran (Zip replacement programs, Surface Routing Model, and final merge of sub-files), FoxPro (Zip validity checks, Airway Routing Models, and Export Routing Model), and Visual Basic (file clean-up) were all used to process the same shipment, thereby affecting file management due to continual importation of files to accommodate non-uniformity of programming input protocols/requirements.

Considerations for Process Improvement (cont.)



2. **Integrate map capabilities during the correction process to help visualize problematic routings that require investigation. For processing 2002 data, additional software (off-the-shelf DeLorme Highway) was purchased separately to aid the Freight-Mileage Analysts in visualizing a problematic routing (for example, a respondent-suggested mode of river/rail where no waterway/railway network appears to be accessible in the area).**
3. **In partnership with the Census Bureau, develop editing specifications for additional "cleaning" of the input data to assure that every record sent to BTS has a valid Zip Code on input, or at least destination information (city-state or foreign country) that will allow a Zip Code to be reasonably determined.**
4. **Develop a more systematic approach to debugging problematic records:**
 - **accumulate and segregate records with similar problems for correction by a subject-matter expert;**
 - **develop more mechanized corrections, as done for for shipments with reported modes of airway only that require truck delivery before and after air transportation.**