ESTIMATION AND VALIDATION OF MODE DISTANCES
FOR THE 1993 COMMODITY FLOW SURVEY

David P. Middendorf
Michael S. Bronzini
Bruce Peterson
Cheng Liu
Shih-Miao Chin

Center for Transportation Analysis
Oak Ridge National Laboratory

September 1995

Prepared by the
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37831
managed by
MARTIN MARIETTA ENERGY SYSTEMS, INC.
for the
U. S. DEPARTMENT OF ENERGY
under contract DE-AC05-840R21400

"The submitted manuscript has been authored by a contractor of the U.S. Government under contract No. DE-AC05-84OR21400. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes."

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED
SUMMARY:

The 1993 Commodity Flow Survey (CFS) collected shipment data from a sample of approximately 200,000 domestic business establishments. Each selected establishment provided information on origin, destination, commodity, shipment weight and value, and modes of transport for a sample of its outbound shipments. One data item not reported by CFS participants was shipment distance. This important piece of information was estimated by simulating probable routes using computer models of the highway, rail, air, waterway, and pipeline networks and their interconnections. This paper describes the nature of the shipment distance estimation problem, the procedures used to estimate mode-specific distances between origin and destination ZIP codes, and the techniques used to validate the results.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.
ESTIMATION AND VALIDATION OF MODE DISTANCES FOR THE 1993 COMMODITY FLOW SURVEY

ABSTRACT

The 1993 Commodity Flow Survey (CFS) collected shipment data from a sample of approximately 200,000 domestic business establishments. CFS participants provided information on origin, destination, commodity, shipment weight and value, and modes of transport for a sample of outbound shipments. One data item not reported by shippers in the CFS was shipment distance. This important piece of information was estimated by simulating probable routes using computer models of the highway, rail, air, waterway, and pipeline networks and their interconnections. This paper describes the nature of the shipment distance estimation problem, the procedures used to estimate mode-specific distances between origin and destination ZIP codes, and the techniques used to validate the results. Considerable computer resources were required to handle the huge number of ZIP code origin-destination (O-D) pairs involved. Existing representations of the highway, rail, and pipeline networks were inspected and enhanced, while representations of the waterway and air freight networks were developed from scratch. These network models were interconnected so that multimodal shipments could be simulated. Efficient network routing procedures were devised for six individual modes of transport and 31 mode combinations. Effective quality control procedures were implemented to monitor the performance of the network routing procedures and validate the results. Mode-specific O-D distances were estimated for approximately five million single mode O-D pairs and approximately 255,000 multimodal O-D pairs. Validating the results for these millions of O-D pairs required at least as much effort as producing them.
ESTIMATION AND VALIDATION OF MODE DISTANCES FOR THE 1993 COMMODITY FLOW SURVEY

INTRODUCTION

The 1993 Commodity Flow Survey (CFS) was the first attempt to gather nationwide data on the flow of goods and materials within the United States since the Commodity Transportation Survey of 1977. Conducted by the Bureau of the Census with additional funding and support from the U.S. Department of Transportation (DOT), the CFS collected shipment data from a sample of approximately 200,000 domestic establishments engaged in mining, manufacturing, wholesale, and selected retail and service activities. Each selected establishment provided information on a sample of its outbound shipments for a two-week period in each of the four calendar quarters of 1993. Among the data items reported for each shipment were the city, state, and ZIP code of the origin and destination; all domestic modes of transport used; the 5-digit Standard Transportation Commodity Classification (STCC) code of the commodity; the value and weight of the shipment; and an indication of whether or not the shipment involved a hazardous material, was containerized, or was an export.

One data item not reported by shippers in the CFS is shipment distance. This important piece of information had to be estimated. The Center for Transportation Analysis (CTA) at Oak Ridge National Laboratory (ORNL) was responsible for estimating transport mode-specific distances for each shipment by simulating probable routes over digital representations of the highway, rail, waterway, air, and pipeline networks and their interconnections. This paper describes the nature of the estimation problem, the procedures used to estimate mode-specific distances, and the techniques used to validate the results.
NATURE AND SCOPE OF THE PROBLEM

Two factors contributed to making the problem of estimating shipment distances for the CFS an especially challenging one. One centered around the requirement to maintain the confidentiality of individual responses to the survey. The other factor consisted of the huge numbers of shipments, origins, and destinations involved as well as the large number of possible transport modes and mode combinations which together presented some interesting data processing and storage problems.

By law, the information reported on each CFS questionnaire is strictly confidential and can be viewed by sworn Census employees only. Consequently, the Census Bureau could not provide data on individual CFS shipments to ORNL. This severely limited the kinds of information about each shipment that could be used to determine its likely route and distance. Some of the information reported by shippers on the CFS questionnaire such as type of commodity, shipment weight and value, and modes of transport would have been very useful in ascertaining routes and distances, especially for multimodal shipments where the characteristics of the commodity could aid in selecting the appropriate intermodal transfer points. Instead, to ensure complete data confidentiality, the only information which the Census Bureau could provide was a long list of 5-digit ZIP code origin and destination (O-D) pairs, including a large number of O-D pairs not reported by any shipper in the survey. For each O-D pair, ORNL had to determine a route and its corresponding distance for every single mode of transport serving both the origin and destination as well as individual mode distances for various combinations of transport modes connecting the O-D pair.

Conceptually, the problem was to construct a set of mode-distance lookup tables, one for
each individual mode of transport and one for each plausible combination of modes. During the early stages of planning for the CFS, each table was conceived as a matrix with a row for each origin ZIP code and a column for each destination ZIP code. Each cell entry in a single-mode table would contain the estimated distance between the corresponding origin and destination ZIP codes via the relevant transport mode. In the mode combination tables, each cell entry would be a vector of mode-specific distances for the various modes making up the combination. One or more other dimensions could also be added to each of these mode-distance tables, representing, for example, different types or classes of commodity or different shipment sizes. Even without these extra dimensions, however, the potential individual and collective size of the mode-distance O-D matrices posed a significant data storage and transfer problem.

As already stated, shipment origins and destinations in the CFS were recorded as 5-digit ZIP codes. Because there are nearly 43,000 5-digit ZIP codes, each mode-distance table could conceivably have over 1.8 billion cells. The final CFS sample was expected to include approximately 24 million shipments. Since each shipment would not have its own unique O-D ZIP code pair, it was clear that less than one percent of all possible ZIP code O-D pairs would be reported in the CFS. The number of unique origin ZIP codes reported in the CFS was approximately 23,200 and the number of unique destination ZIP codes was approximately 41,200. To account for every possible O-D ZIP code pair in the CFS, each mode-distance table would therefore have to contain almost 956 million cells, nearly 40 times the expected number of shipments in the CFS sample. Consequently, it was not practical to estimate mode distances from every origin ZIP code to every destination ZIP code reported in the CFS or to even reserve space in a mode-distance lookup table for every possible O-D ZIP code pair. A more space-efficient
mode-distance data structure was required.

An associated issue was the question of how many mode-distance tables to create or, equivalently, which modes and mode combinations to consider. The CFS questionnaire listed ten domestic modes of transport. CFS participants were instructed to indicate which one or more of these modes were used to complete each selected shipment. From these ten individual modes numerous plausible mode combinations could be identified. Creating a separate O-D distance table for each one was clearly not very practical. A decision was therefore reached to organize the various modes and mode combinations into three groups or tiers. Tier I consisted of the following modes of transport for single-mode shipments: truck, rail, inland water and/or Great Lakes, deep sea water, air, and pipeline. Tier II, shown in Table 1, included what were considered to be the most common combinations of modes for multimodal shipments. Grouped into Tier III were those three and four-mode combinations considered to be plausible but relatively rare. They are shown in Table 2.

Three mode-distance tables were therefore developed, one for each tier. The O-D matrix format was abandoned. Instead, a fixed-length record structure was specified for each table. Each record in the Tier I table consisted of fields for origin ZIP code, destination ZIP code, the great circle distance between the origin and destination ZIP code centroids, highway distance, rail distance, inland water and/or Great Lakes distance, deep sea water distance, air distance, and pipeline distance. Tier II and Tier III records also consisted of fields for origin, destination, and great circle distance as well as a distance field for each mode in a mode combination for each combination included in the tier. The Tier I table contained records for approximately 5 million O-D ZIP code pairs. A large but unspecified number of these O-D pairs were not reported in the
Table 1. The Tier II mode combinations for multimodal shipments.

<table>
<thead>
<tr>
<th>Combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>private truck — for-hire truck</td>
</tr>
<tr>
<td>rail — truck</td>
</tr>
<tr>
<td>rail — private truck — for-hire truck</td>
</tr>
<tr>
<td>deep sea water — truck</td>
</tr>
<tr>
<td>deep sea water — private truck — for-hire truck</td>
</tr>
<tr>
<td>deep sea water — rail</td>
</tr>
<tr>
<td>inland water and/or Great Lakes — truck</td>
</tr>
<tr>
<td>inland water and/or Great Lakes — private truck — for-hire truck</td>
</tr>
<tr>
<td>inland water and/or Great Lakes — rail</td>
</tr>
<tr>
<td>inland water and/or Great Lakes — deep sea water</td>
</tr>
<tr>
<td>air — truck</td>
</tr>
<tr>
<td>air — private truck — for-hire truck</td>
</tr>
<tr>
<td>pipeline — truck</td>
</tr>
<tr>
<td>pipeline — private truck — for-hire truck</td>
</tr>
<tr>
<td>pipeline — rail</td>
</tr>
<tr>
<td>pipeline — deep sea water</td>
</tr>
<tr>
<td>pipeline — inland water and/or Great Lakes</td>
</tr>
</tbody>
</table>
Table 2. The Tier III mode combinations for multimodal shipments.

<table>
<thead>
<tr>
<th>Mode Combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>rail — deep sea water — truck</td>
</tr>
<tr>
<td>rail — deep sea water — private truck — for-hire truck</td>
</tr>
<tr>
<td>rail — inland water and/or Great Lakes — truck</td>
</tr>
<tr>
<td>rail — inland water and/or Great Lakes — private truck — for-hire truck</td>
</tr>
<tr>
<td>deep sea water — inland water and/or Great Lakes — truck</td>
</tr>
<tr>
<td>deep sea water — inland water and/or Great Lakes — rail</td>
</tr>
<tr>
<td>deep sea water — inland water and/or Great Lakes — private truck — for-hire truck</td>
</tr>
<tr>
<td>pipeline — rail — truck</td>
</tr>
<tr>
<td>pipeline — rail — private truck — for-hire truck</td>
</tr>
<tr>
<td>pipeline — deep sea water — truck</td>
</tr>
<tr>
<td>pipeline — inland water and/or Great Lakes — truck</td>
</tr>
<tr>
<td>pipeline — deep sea water — inland water and/or Great Lakes</td>
</tr>
<tr>
<td>pipeline — deep sea water — private truck — for-hire truck</td>
</tr>
<tr>
<td>pipeline — inland water and/or Great Lakes — private truck — for-hire truck</td>
</tr>
</tbody>
</table>
CFS but were included by the Census Bureau to ensure data confidentiality. The Tier II and Tier III mode-distance tables contained records for approximately 250,000 and 5,000 O-D pairs, respectively. Only about one-fifth of these records represented ZIP code pairs actually reported in the CFS for multimodal shipments. The rest were added by the Census Bureau for the purpose of maintaining data confidentiality. In this way, the existence of any O-D entry could not be construed as evidence of an actual business relationship between firms in either ZIP code.

In general, the higher the tier, the more difficult it was both to estimate shipment distances and to validate the estimates, primarily because of the larger number of modes involved in a shipment.

MODE DISTANCE ESTIMATION

The approach taken to estimate mode-specific shipment distances for O-D ZIP code pairs was fairly standard. It basically involved measuring distances along minimum impedance paths simulated on digital representations of the various transportation mode networks and their interconnections. Each mode's network was modelled as a collection of links and nodes, the former representing portions of linehaul facilities and the latter representing terminals, intersections, interchanges, and other locations where some change in the movement of freight can occur. Each node's location was specified by its latitude and longitude. Nodes, therefore, served to define the shape or topology of the physical network. Each mode and mode combination, however, presented its own set of complications. Consequently, the detailed distance estimation procedures were highly mode-specific. The following sections describe some of the technical issues involved.
Local Access

The first step in finding a route for a shipment was to ascertain whether the relevant transport networks were accessible from within the origin and destination ZIP code areas. If so, the next step was to determine the points where the shipment would likely enter and exit the national modal or multimodal network. Because shipping and receiving establishments may not be located at a node on the national network or next to a linehaul link, some distance may be involved in accessing the national network at either or both ends of the shipment. This distance is known as the local access distance and was represented in the network model as a local access link. The procedures for determining local accessibility, access location, and access distance varied by mode.

Inter-ZIP code shipments that begin by truck were assumed to start at the origin ZIP code's geographic centroid. Likewise, inter-ZIP code shipments ending by truck were assumed to terminate at the destination ZIP code's geographic centroid. Access links to the national highway network, therefore, simulated the movement of trucks over local streets and secondary roads between the geographic center of the ZIP code area and nearby nodes on the national highway network. A ZIP code was considered to have access to the national highway network if at least one highway node was within 200 miles of the ZIP code's centroid. An access link was created from the ZIP code centroid to the nearest highway node in each quadrant, unless the straight line distance to the closest node in the quadrant was more than three times the straight line distance to the closest of all highway nodes. The length of a highway access link was defined as its straight line distance from the centroid to the highway node multiplied by an assumed circuity factor of 1.2.
For non-highway modes, the concept of a local access threshold circle was employed to ascertain local accessibility. The center of the circle was located at the ZIP code's centroid and its radius for a given mode $m$ was calculated from the following formula:

$$R_m = R_{MAX} + 2 \times \epsilon_m + \rho_m$$

where $R_{MAX}$ is the straight line distance from the centroid to the farthest point on the ZIP code's boundary, $\epsilon_m$ is the geographic error in the network model's representation of link locations for mode $m$, and $\rho_m$ is the maximum length of a local link not included in mode $m$'s national network.

In the case of rail, for example, the latter variable represented the length of an industrial rail spur.

For a ZIP code to have access to the rail network, part of at least one rail link had to lie within the local rail access threshold circle. If the rail network was accessible, rail access links were created connecting the centroid to nearby rail links in each quadrant. Because rail infrastructure is privately owned, track ownership and trackage rights were taken into consideration. A pair of access links was created for each railroad company's closest accessible rail link in each quadrant, each link of the pair connecting the centroid to one of the two end nodes of an accessible linehaul link. Unlike highway access links, however, the length of a rail access link was not defined as the straight line distance between the centroid and an end node multiplied by a circuity factor. This was because, unlike shipments that utilize trucks, rail shipments were not assumed to begin or end at a ZIP code centroid. In computing the length of a rail access link, use was made of the fact that linehaul links in the rail network were represented by a sequence of shape points rather than as straight lines. First, the point $X$ closest to the centroid $Z$ on the linehaul link was determined. The straight line distance between this point and the centroid, $d_{ZX}$, was then calculated. If this distance was less than the assumed maximum spur length, the link was considered to be accessible.
length, \( p \), the length of the rail access link to end node \( A \) was defined as \( d_{2X} + d_{X4} \), the latter being the distance along the linehaul link between point \( X \) and node \( A \). Otherwise, the length of the access link was computed as \( p + d_{X4} \).

Like rail, water transport was considered accessible to a ZIP code if any part of at least one waterway link was inside the ZIP code's local waterway access threshold circle.\(^6\) Pairs of access links were created from the centroid to each end node of the closest accessible waterway link in each quadrant. Like the linehaul links in the rail network, waterway links were delineated by shape points. For each accessible waterway link, the point \( X \) closest to the centroid was determined. The length of one access link in a pair was given by the length of the link segment between this point \( X \) and end node \( A \). Likewise, the length of the other access link was equal to the length of the link segment between \( X \) and end node \( B \). Waterborne shipments, therefore, were assumed to begin and end on the national waterway network rather than at ZIP code centroids.

In order for a shipment to begin or end as air freight without the need for some means of surface transport, an airport had to be located within the origin or destination ZIP code's local airport access threshold circle.\(^7\) Access links were created from the centroid to the closest accessible airport in each quadrant. However, unlike access links to the surface transportation networks, the length of air access links was zero, because air-only shipments were assumed to begin and end at the originating and terminating airports, respectively.

Local access to the national pipeline network was assumed if a ZIP code's local pipeline access threshold circle contained at least one pipeline node.\(^4\) Access links were created from the centroid to every pipeline node lying within the pipeline access threshold circle. The reason for this was twofold. First, individual pipelines are privately owned and operated. Second, various
companies' pipeline subnetworks often do not interconnect. Limiting the number of pipeline access links to no more than one per quadrant would have produced erroneous results for many O-D pairs. Pipeline access links were used to represent various gathering lines, field lines, or local distribution pipelines not included in the national pipeline networks. Consequently, the length of each pipeline access link was determined by the straight line distance between the ZIP code centroid and the pipeline node to which it connected.

**Single Mode Paths**

The methods used to estimate O-D distances for shipments involving only a single mode of transport were highly mode-specific. In general, however, the first step involved computing an impedance for each link in the relevant modal network. A separate impedance function was developed for each mode. The impedance function for highway links was fairly elaborate. In computing an expected link travel time, it considered several link characteristics such as divided or undivided roadway, degree of access control, rural or urban setting, type of pavement, functional classification, number of lanes, degree of urban congestion, traffic restrictions, and presence of toll facilities. The impedance on a rail network link was determined by multiplying the link length by the relative impedance factor for the link's line class, which is an indicator of line quality. Five rail line classes were defined, based on annual gross tonnage. For the water, air, and pipeline modes, link length alone was used as a measure of impedance.

The next step involved finding the path with the lowest cumulative impedance using a standard algorithm for finding the shortest paths between nodes in a directed graph with weighted edges. For the truck, water, and pipeline modes, only one minimum impedance path was possible for an O-D pair. In each of these cases, the O-D distance was found by simply adding up the
lengths of the links on the path.

In the case of rail, multiple paths were possible for many O-D pairs. A separate minimum impedance path was found for each combination of originating and terminating railroad serving an O-D pair. This was accomplished by decomposing the rail network into a separate subnetwork for each railroad company. The various subnetworks were connected at established transfer points by means of notional interline links used to simulate the exchange of freight between interconnecting carriers. Impedances on these links represented interline transfer penalties. The default transfer penalty was set at 300 mainline miles, which is roughly equivalent to the distance traveled in one day by unexpedited rail freight. The weighted average rail distance for an O-D pair connected by multiple rail paths was then determined from the following equation:

\[ d = \frac{\sum_{j=1}^{p} d_j e^{-\theta t_j}}{\sum_{k=1}^{p} e^{-\theta t_k}} \]

where \( p \) is the number of alternative rail paths between the origin and destination, \( d_j \) is the distance using path \( j \), and \( t_j \) is the total impedance on path \( j \), including interline transfer penalties. The value of the parameter \( \theta \) was determined to be 1/200.

The procedure for determining O-D air distances made use of an airport-to-airport distance matrix. Each entry in the matrix represented the average distance of alternative air routes between a pair of airports. The five shortest routes were found for each carrier providing air freight service between a given pair of airports. Any route whose distance was greater than 2.5 times the great circle distance between the airports was excluded from further consideration. The
air distance between the two airports was then computed as the average distance of the remaining
routes. If, however, no one air carrier served the pair of airports, or if the lowest circuity of all
the single airline routes was greater than 2.5, then the five shortest routes involving no more than
one interline transfer were determined. Again, any route with circuity greater than 2.5 was
excluded. Moreover, only routes with interline transfers occurring at FAA-defined hubs were
considered. The average distance of these routes was then entered into the airport-to-airport
distance matrix. The next step was taken only if no routes involving only one interline transfer
and a circuity less than 2.5 were found in the previous step. In this case, the search was extended
to finding the five shortest routes involving two or more interline transfers. The air distance was
computed as the average distance of these routes, excluding any route whose distance was more
than 2.5 times greater than the great circle distance between the airports.

In determining the air freight distance between a pair of ZIP codes, the first step was to
ensure that there was at least one airport within the local airport access threshold circle for each
ZIP code in the pair. If so, the average air route distance for each airport pair serving the origin
and destination was looked up in the airport-to-airport distance matrix. The shortest airport-to-
airport distance was then chosen as the air shipment distance for the O-D pair.

Finding paths and determining distances for shipments involving the parcel delivery-
courier-U.S. Postal Service mode presented an interesting technical problem. This mode consists
of delivery services that carry letters, parcels, packages, and other small shipments typically
weighing under 100 pounds. These services may use one or more means of transport to complete
a shipment, including truck, bus, airplane, and rail. The initial assumption was that parcel
shipments between ZIP codes less than 200 miles apart would utilize the highway network while
longer parcel shipments would utilize air transport. Further investigation, however, revealed that
different package delivery carriers will use a different mix of ground and air transportation
depending not only on the distance involved but also on the class of service provided, day of the
week, existing labor agreements, and other contractual and legal requirements. The initial
assumption, therefore, could not be supported. Because of the complexity and wide variety of
parcel delivery services and the lack of sufficient information upon which to develop a more
sophisticated routing model, the decision was made to use highway distances for this mode
regardless of the geographic distance between the origin and destination ZIP codes. In other
words, Tier I truck distances were used as a surrogate for parcel delivery distances.

**Multimodal Paths**

Determining O-D distances for shipments involving more than one mode of transport
requires having a multimodal network. This network was created by connecting the individual
modal networks at nodes representing truck terminals, trailer-on-flatcar/container-on-flatcar
(TOFC/COFC) facilities, river ports, seaports, airports, refineries, tank farms, and other places
where intermodal transfer can occur. The multimodal network, therefore, consisted of access
links between ZIP code centroids and the various individual modal networks, links and nodes on
the mainline networks, and intermodal transfer links between the individual modal networks.

Table 3 indicates how various pairs of modes were connected. In most cases, the
intermodal connections consisted of mode-specific transfer links between nodes representing
intermodal transfer facilities and nearby nodes on a modal network. For example, intermodal
transfer links between airports and the national highway system were created to simulate the
movement of trucks over local streets and roads between an airport and nearby nodes on the
Table 3. Transfer connections between modes.

<table>
<thead>
<tr>
<th>Mode Pair</th>
<th>Transfer Connections</th>
</tr>
</thead>
<tbody>
<tr>
<td>private truck — for-hire truck</td>
<td>Centroids of ZIP codes containing truck terminals where freight can be interchanged between motor carriers</td>
</tr>
<tr>
<td>truck — air</td>
<td>Links representing local streets and roads connecting airports to nearby nodes on the national highway network</td>
</tr>
<tr>
<td>truck — rail</td>
<td>Links representing local streets and roads connecting rail nodes with TOFC/COFC ramps to nearby nodes on the national highway network</td>
</tr>
<tr>
<td>truck — water</td>
<td>Links representing local streets and roads connecting ports to nearby nodes on the national highway network</td>
</tr>
<tr>
<td>truck — pipeline</td>
<td>Links representing local streets and roads connecting refineries, tank farms, and other pipeline terminals to nearby nodes on the national highway network</td>
</tr>
<tr>
<td>rail — water</td>
<td>Links representing rail spurs connecting ports with nearby nodes on the national rail network</td>
</tr>
<tr>
<td>rail — pipeline</td>
<td>Links representing rail spurs connecting refineries, tank farms, and other pipeline terminals to nearby nodes on the national rail network</td>
</tr>
<tr>
<td>shallow draft vessel — deep draft vessel</td>
<td>Designated ports where freight can be interchanged between shallow draft and deep draft vessels</td>
</tr>
<tr>
<td>pipeline — water</td>
<td>Links representing local pipelines connecting ports to nearby nodes on the national pipeline network</td>
</tr>
</tbody>
</table>
national highway network. Likewise, truck-rail intermodal links represented unspecified local streets and roads used by trucks to access TOFC/COFC facilities. Transfer links between the rail and waterway networks were used to model the movement of railcars over industrial spurs to and from river and ocean ports. Many of these mode-specific intermodal transfer links were notional. Instead of representing an actual physical facility, they represented either a group or class of transport facilities such as local streets and roads or an unknown actual facility not included in the national modal network such as an industrial rail spur or a gathering pipeline. This was a consequence of the fact that the various national modal networks were independently developed or adapted by separate groups and organizations. As improvements are made to the national modal networks and more information is obtained on intermodal transfer facilities and connections, it is hoped that most if not all of these notional transfer links will be replaced by links representing actual physical transportation facilities.

One of the intermodal connections in Table 3 was modeled as a zero-length transfer link. This was the connection between shallow draft and deep draft vessels at designated ports linking the inland, Great Lakes, and ocean waterway networks.

Table 3 also includes an intramodal transfer connection. For shipments involving both private and for-hire trucking but no other mode, it was assumed that private trucking is used to convey the goods from the origin to a truck terminal and that for-hire trucking is used to transport the freight from the truck terminal to the destination. The American Motor Carrier Directory was consulted for the locations of truck terminals where freight may be interchanged between motor carriers. This publication provided ZIP codes for many but not all truck terminals. A set of nearly 800 terminals, operated either by major transcontinental motor carriers such as Roadway
Express and Consolidated Freightways or by large regional carriers, was carefully selected from this directory. They were incorporated into the national highway network as intramodal transfer nodes located at the centroids of their respective 5-digit ZIP codes.

Another aspect of multimodal shipments that had to be considered was the sequencing of the relevant transport modes. Although shippers indicated which modes were involved in a shipment sampled for the CFS, they were not required to specify in which order the modes were used. For some mode combinations, more than one realistic mode sequence is possible. For combinations that included trucking, it was assumed that trucks were used to either pick up the freight at the origin or deliver the freight at the destination or both. In other words, it was assumed that trucks would not be employed as the principal mode. When more than one mode sequence was possible for an O-D pair, a minimum impedance path was found for each. Mode distances were then reported for the path having the lowest total impedance.

To find a multimodal path for an O-D pair, the link impedances on each mode network had to be adjusted to account for the relative utility of each mode. Usually only one of the modes in a mode combination is the principal one, used to cover the bulk of the distance between the origin and destination. The reasons for the shipper choosing this mode as the principal one may have something to do with cost, travel time, reliability, safety, capacity, convenience, or a combination of these and other considerations. The other modes in the mode combination are mainly used to get the cargo to and from the principal mode. Presumably, from the shipper's standpoint, these other modes are more costly or, in some other way, less desirable than the principal mode and are needed to complete the shipment because the principal mode is not directly accessible at either or both ends of the shipment. A realistic multimodal routing model,
therefore, must seek a path that favors the principal mode as much as possible without unduly increasing the total shipment distance or the route's circuity. One way of accomplishing this for a minimum impedance path model is to develop and apply relative impedance factors for the various modes.

As a starting point, published information on annual operating expenses and ton-miles by mode was examined. These data suggested that the cost per ton-mile in 1992 was about $0.22 for tractor-trailer, $0.03 for rail, and $0.01 for domestic water. Thus, the cost impedance of rail would appear to be about 1/7th and that of domestic water about 1/22nd that of trucking. Direct operating cost, however, is only one consideration in determining relative impedance factors, although it provided a very useful starting point. Since the objective was to find a path for a specific mode combination rather than determine the best mode combination and its path for an O-D pair, it was necessary to adjust the relative impedance factors based on operating costs alone so as not to overly penalize the more costly modes in a mode combination. The following relative impedance factors were therefore postulated:

- Truck — 1
- Rail — 1/4
- Air — 1/10
- Barge — 1/8
- Ocean vessel — 1/10
- Pipeline — 1/12

These factors represent the impedance of a mile of movement over the most important traffic arteries of a mode's network relative to one mile of movement over a rural Interstate highway.
Thus, for the truck-rail-inland water mode combination, rail link impedances were factored by $\frac{1}{4}$ and waterway link impedances were factored by $\frac{1}{8}$.

Intermodal transfer impedances also had to be specified. These penalties had to be large enough to prevent transfers from occurring back and forth between a pair of modes, but they could not be so large as to prevent any intermodal transfers from occurring at all. Table 4 shows the intermodal transfer penalties that were assumed. These impedances can be loosely interpreted as the generalized cost of transferring a ton of freight to another mode relative to the cost of moving the ton of freight one mile in a truck on a freeway.

**VALIDATION OF RESULTS**

To validate every estimated mode distance was clearly impractical. Tier I, for example, required finding shipment distances between approximately 5 million O-D pairs for each of six individual modes for a total of 30 million O-D mode distances. Tier II involved 17 mode combinations for each of approximately 250,000 O-D pairs, thereby adding another 4.25 million O-D distances to be confirmed. Moreover, because of the huge amount of mass storage space required, it was possible to store only a very small sample of the paths found by the network routing procedures. Therefore, a quality assurance (QA) plan had to be developed for methodically choosing a sample of O-D paths for validation.

The devised QA plan incorporated two strategies. The first was used to identify estimated O-D mode distances which satisfied certain conditions that, in turn, could indicate possibly erroneous, illogical, or unrealistic results. The second strategy consisted of checking the results for O-D pairs involving the most frequently occurring origin and destination ZIP codes in the CFS. Although these two strategies were employed for single mode as well as multimodal
Table 4. Intermodal transfer impedances.

<table>
<thead>
<tr>
<th></th>
<th>Air</th>
<th>Rail</th>
<th>Barge</th>
<th>Ship</th>
<th>Pipeline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truck</td>
<td>300</td>
<td>300</td>
<td>500</td>
<td>700</td>
<td>200</td>
</tr>
<tr>
<td>Rail</td>
<td>—</td>
<td>—</td>
<td>600</td>
<td>800</td>
<td>200</td>
</tr>
<tr>
<td>Barge</td>
<td>—</td>
<td>600</td>
<td>—</td>
<td>600</td>
<td>300</td>
</tr>
<tr>
<td>Ship</td>
<td>—</td>
<td>800</td>
<td>600</td>
<td>—</td>
<td>300</td>
</tr>
<tr>
<td>Pipeline</td>
<td>—</td>
<td>200</td>
<td>300</td>
<td>300</td>
<td>—</td>
</tr>
</tbody>
</table>
distances, there were some differences in how they were implemented.

**Single Mode Shipment Distances**

The primary means of detecting questionable O-D distances for single mode shipments consisted of an analysis of route or path circuity. For each combination of O-D pair and mode, the reported mode distance, if any, was divided by the great circle distance (GCD) between the origin and destination ZIP code centroids to yield a circuity factor for the minimum impedance route or path. To increase the likelihood of detecting unreasonable or illogical mode distances, the O-D pairs were categorized into five GCD groups:

- $0 < \text{GCD} < 100$ miles;
- $100 \leq \text{GCD} < 500$ miles;
- $500 \leq \text{GCD} < 1000$ miles;
- $1000 \leq \text{GCD} < 2000$ miles; and
- $\text{GCD} \geq 2000$ miles.

Within each GCD group, a cumulative frequency distribution of circuity factors was produced. The O-D pairs associated with the highest circuity factors (generally above the 95th to 99th percentile, depending on the mode) were then selected for manual checking. The latter involved locating the origin and destination on a map and noting the presence of any geographic barriers such as lakes, rivers, bays, inlets, mountain ranges, and so on as well as the topology of the network for the specific mode of transportation under consideration. In most cases this was enough to confirm the high circuity. In a few cases it was necessary to view the path chosen by the network routing algorithm. This was accomplished by rerunning the routing program on the O-D pair for the given mode, storing the sequence of links and nodes comprising the minimum
impedance path, and displaying or plotting the results using geographic information system (GIS) software.

The circuity analysis did not reveal any systematic errors in the Tier I results. Average circuity factors, shown in Table 5, generally agreed with expectations. The vast majority of the highly circuitous paths were explained either by the presence of some kind of geographic barrier between the origin and destination or by the configuration or shape of the relevant modal network.

The circuity analysis did expose a few isolated or localized problems. For example, although ferry links were included in the highway network, because of their relatively high impedance, the minimum impedance paths tended to avoid them. This resulted in some highly circuitous paths around Puget Sound as well as a few questionable paths between places in the Alaska panhandle. The impedance function for ferry links was subsequently modified to rectify these problems.

Although the rail network routing procedures performed quite satisfactorily, the circuity analysis uncovered two kinds of problems in the rail network itself. The first one consisted of missing connections between rail carriers, usually between short-line and Class I railroads. The result was an overly circuitous route for some O-D pairs and no rail path at all for others. The other rail network problem involved incorrectly coded connections between railroads. For example, the rail network database indicated an interline connection between the Soo Line and the Duluth & Northeastern Railroad at Detroit Lakes, MN, whereas the connection should have been between the Soo Line and the Burlington Northern. This network data error was exposed by a highly circuitous path between two ZIP codes in western Minnesota. Although the rail circuity
Table 5. Circuity factors for single mode O-D paths.

<table>
<thead>
<tr>
<th>GCD Group</th>
<th>Truck</th>
<th>Air</th>
<th>Rail</th>
<th>Inland Water&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Pipeline</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Avg.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 100 mi</td>
<td>1.4</td>
<td>1.3</td>
<td>2.3</td>
<td>2.2</td>
<td>1.5</td>
</tr>
<tr>
<td>100 - 499 mi</td>
<td>1.2</td>
<td>1.6</td>
<td>1.6</td>
<td>2.9</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>1.7</td>
<td>2.3</td>
<td>4.8</td>
<td>6.2</td>
<td>2.4</td>
</tr>
<tr>
<td>1.3</td>
<td>1.4</td>
<td>2.3</td>
<td>1.4</td>
<td>2.2</td>
<td>1.2</td>
</tr>
<tr>
<td>1.4</td>
<td>1.6</td>
<td>2.3</td>
<td>3.8</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>500 - 999 mi</td>
<td>1.2</td>
<td>1.4</td>
<td>1.7</td>
<td>2.2</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>1.3</td>
<td>1.9</td>
<td>3.8</td>
<td>2.6</td>
<td>1.5</td>
</tr>
<tr>
<td>1.3</td>
<td>1.6</td>
<td>1.5</td>
<td>1.5</td>
<td>2.6</td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>1.4</td>
<td>1.5</td>
<td>1.4</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>1.9</td>
<td>1.7</td>
<td>1.5</td>
<td>1.4</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>2000 + mi</td>
<td>1.2</td>
<td>1.1</td>
<td>1.3</td>
<td>—&lt;sup&gt;c&lt;/sup&gt;</td>
<td>—&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>1.3</td>
<td>1.3</td>
<td>1.4</td>
<td>—&lt;sup&gt;c&lt;/sup&gt;</td>
<td>—&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Notes:

<sup>a</sup> Includes Great Lakes.
<sup>b</sup> 95th percentile circuity factor.
<sup>c</sup> None of the O-D pairs with great circle distances of 2000 miles or more were connected by paths on the inland water or pipeline networks.
analysis did not uncover all instances of these kinds of errors in the rail network, it did reveal their existence. As a result, special diagnostic programs were written to find all occurrences of these errors in the rail network so that they could be corrected.

Questionable inland waterway distances were obtained for O-D pairs in which the path began and ended on the same waterway link. These cases were characterized by large rural origin and destination ZIP codes with short great circle distances between their centroids and a waterway link whose length was considerably longer than the O-D great circle distance. In these situations, the path would begin on the link at the point $P_o$ closest to the origin ZIP code centroid, proceed to one of the end nodes of the link, and then go back on the same link to the point $P_d$ closest to the destination ZIP code centroid. A more realistic path would have been from point $P_o$ directly to point $P_d$. This problem was not considered serious for the simple reason that virtually none of the affected O-D pairs appeared to be important from the standpoint of waterborne commerce. One way to solve the problem would be to insert additional nodes into the inland waterway network, thereby breaking up long waterway links into two or more shorter ones.

The analysis of pipeline route circuitry revealed a few weaknesses in the pipeline network and in the original method for determining pipeline access. Originally, a pipeline access link was created between a ZIP code centroid and the closest pipeline node within the pipeline access threshold circle for each quadrant around the centroid. This approach ignored the fact that pipelines are owned by private companies and that the various companies' pipelines often do not interconnect. Under the original procedure, if two or more companies' pipelines were accessible in a given quadrant, access was provided only to the one with the closest pipeline node. As a consequence, shipment by pipeline was incorrectly reported as impossible for some O-D pairs.
because the appropriate pipeline could not be accessed. In other cases, the result of the minimum impedance pathfinding algorithm was a convoluted and highly circuitous path. To remedy this problem, the original pipeline access procedure was revised to connect the ZIP code centroid to all pipeline nodes lying within the pipeline access threshold circle.

The analysis of circuity factors was useful for identifying possibly erroneous O-D distances by focusing on highly circuitous paths. The other strategy used to validate O-D distances for single mode shipments involved checking the distances between the most frequently occurring origin and destination ZIP codes. Nearly 3000 O-D pairs were examined under this strategy. Circuity factors were computed and analyzed for each selected O-D pair by mode. Where possible, highway and rail distances were estimated from maps or atlases, from published intercity highway mileage tables and rail route profiles, or from commercial travel planning software and road atlas databases available on CD-ROM. These estimated distances were then compared to the distances produced by ORNL's network routing procedures. When there was significant lack of agreement, the minimum impedance path found by the network routing algorithm was plotted and examined to determine the reason for the discrepancy.

In general, the analysis of reported mode distances for the major O-D pairs did not reveal any significant or systematic problems in the results. Circuity factors were consistent with expectations. Cases in which shipment by a particular mode was reported as impossible were visually confirmed. Although reported highway distances sometimes differed from distances obtained from maps or published highway mileage tables, the minimum impedance paths were always reasonable and, in many cases, were more likely for truck shipments than the shortest distance or time paths suggested by the other sources.
Multimodal Shipment Distances

Because multimodal shipping is inherently more complex than single mode shipping due to intermodal connections, the estimated O-D distances for each of the Tier II and Tier III mode combinations were much harder to validate. A considerable amount of judgment was required. For example, in evaluating the minimum impedance path for a specific O-D pair and mode combination, a judgment had to be made on the plausibility of the intermodal exchange locations chosen by the multimodal routing procedures. Furthermore, in judging the validity of a path, some consideration also had to be given to the likelihood that the particular mode combination would be used to transport freight between the specified origin and destination. For example, the combination of rail and water is not likely to be used to ship freight between Denver, Colorado, and Tucson, Arizona. A minimum impedance rail and water route for this O-D pair would necessarily be long and circuitous but, nevertheless, it could be valid although unlikely.

Circuity factors were calculated for each multimodal path. For mode combinations not involving water transportation, the most highly circuitous multimodal paths were typically between origins and destinations less than 200 miles apart. For example, of the O-D pairs where the estimated total O-D shipment distance was more than five times the GCD, the percentage involving origins and destinations less than 200 miles apart was 97 percent for truck and rail shipments, 99.7 percent for truck and air shipments, and 99 percent for truck and pipeline shipments. For mode combinations including inland water transportation, of the O-D pairs where the estimated total O-D shipment distance was more than five times the GCD, the percentage involving origins and destinations less than 500 miles apart was 95 percent for truck and inland water shipments, 93 percent for rail and inland water shipments, and 78 percent for pipeline and
inland water shipments. In virtually all of these cases, the reason for the high degree of circuity was the fact that the origin and destination ZIP code centroids were closer to each other than they were to the nearest access point for the principal non-highway mode. In general, the most circuitous multimodal paths tended to involve an unlikely mode combination for the particular O-D pair. Consequently, the path circuity analysis was only moderately useful in evaluating the performance of the multimodal routing models.

Also singled out for further scrutiny were paths where the truck mileage accounted for more than half of the O-D distance. The assumption was that the primary function of trucks in a multimodal shipment is to haul cargo to and from the non-highway modes. The latter were considered to be the principal linehaul modes and, therefore, should account for the bulk of the shipment distance. For most of the mode combinations involving truck transport, trucks were the dominant mode in less than a majority of the cases. For example, truck mileage accounted for more than half of the total O-D distance in only two percent of the truck and rail shipments, two percent of the truck and air shipments, and 23 percent of the truck and inland water shipments. In the case of truck and pipeline shipments, however, trucks were the dominant mode in nearly three quarters of the O-D pairs where this mode combination was physically possible. In all cases where trucking was the primary mode, the relevant non-highway modes were far away from both origin and destination. For that reason, multimodal paths consisting mainly of highway links tended to be highly circuitous. In most cases, they indicated that the mode combination was highly unlikely for the O-D pair. Like the circuity analysis, this check on quality control was only moderately useful in evaluating the performance of the multimodal routing procedures.

The most effective but also time-consuming strategy for validating the multimodal O-D
distances was to visually inspect the results for carefully selected O-D pairs. The most frequently occurring origin and destination ZIP codes among the Tier II and Tier III O-D pairs were identified. A sample of O-D pairs involving these frequent origins and destinations were selected. GIS software was then used to display and plot the multimodal paths for each major O-D pair along with the underlying networks and intermodal transfer points. Each multimodal path was studied and possible alternative paths involving different intermodal transfer points and mode sequences were identified. In those cases where one of the modes in a mode combination was excluded from the generated minimum impedance path, an effort was made to manually find at least one realistic path that included all of the pertinent modes and to determine why the multimodal routing procedure failed to produce it. As a result of this validation process, some of the relative impedance factors and intermodal transfer penalties were adjusted and the routing models for some of the mode combinations were modified.

SUMMARY AND CONCLUSIONS

Estimating mode-specific O-D distances for the 1993 CFS was a major undertaking that presented a number of significant technical challenges. Distances were determined using computer simulation to find the "most likely" routes between origin and destination ZIP codes over a network representation of the transportation system. Considerable computer resources were required to handle the huge number of ZIP code O-D pairs involved. Existing computer models of the highway, rail, and pipeline networks had to be inspected and enhanced, while computer models of the waterway and air freight networks had to be developed from scratch. All of these networks had to be interconnected so that multimodal shipments could be simulated. Efficient network routing procedures had to be devised for six individual modes of transport and
31 combinations of modes. Effective quality control procedures were needed to monitor the performance of the network routing procedures and to validate the estimated mode-specific O-D distances for approximately five million single mode O-D pairs and approximately 255,000 multimodal O-D pairs.

The single mode routing procedures produced highly satisfactory results. Average circuity factors for each mode were consistent with expectations. Most of the problems uncovered by the validation process were easily resolved by correcting errors in the modal networks, adding links or nodes, or adjusting the mode-specific impedance functions. The most questionable distances generally involved non-highway routes between large, rural origin and destination ZIP codes where the distance between the centroids was less than the lengths of nearby links on the relevant modal network. However, the likelihood that much, if any, non-highway freight traffic flows between such places was considered too small to justify spending much time in trying to produce more reasonable distances.

As expected, the multimodal results were generally adequate but not as satisfactory as the single mode distances. The multimodal routing procedures were usually able to find a minimum impedance path involving all modes in a mode combination for O-D pairs where such a path reasonably existed. Whether this path was the most likely one was debatable in many cases. Other intermodal transfer locations perhaps better than the ones selected by the routing program could often be identified. The choice of intermodal transfer facility depends on commodity type and shipment size. Some ports, for example, are equipped to handle large quantities of grain but not coal. The route taken by a rail-barge shipment would therefore depend on which type of commodity was involved. The multimodal routing procedures were clearly hampered by the lack
of commodity-specific information.

Improvements in the individual modal networks and the connections between them would also enhance the ability to determine multimodal routes and shipment distances. Most of these improvements require additional attribute data for non-highway links and nodes such as typical speed or travel time and functional class. This would enable the development and application of better generalized cost functions to replace the relative impedance factors. More detailed information on the functional characteristics and relative attractiveness of individual intermodal transfer facilities is also needed. This would result in facility-specific intermodal transfer impedances. Work is currently being undertaken to improve the existing network models and their interconnections as part of the development of a National Transportation Atlas Database.12

Validating the mode-specific distances for millions of O-D pairs required at least as much effort as producing them. Fortunately, GIS software was available to facilitate the validation process. Commercial geographic information systems were used to display selected single mode and multimodal paths along with the underlying networks, intermodal connections, and ZIP code boundaries. Nevertheless, a considerable amount of manual effort was still required. It is hoped that future improvements in existing GIS software, particularly in the area of transportation network analysis, will reduce the manual burden even more.

ACKNOWLEDGEMENTS

The work described in this paper benefited from the contributions of many people and organizations. The authors are grateful to the Bureau of Transportation Statistics (BTS), U.S. Department of Transportation, which provided funding and technical support for the CFS and particularly to Dr. Rolf Schmitt, Associate Director, for providing valuable technical guidance.
throughout the project. The authors also wish to thank the Bureau of the Census, which conducted the CFS in an expedited manner, and especially Mr. James Aanestad, Mr. John Fowler, Ms. Wanda Dougherty, Mr. Bob Crowther, and Mr. Jerry Ankers who worked closely with ORNL in testing the network routing procedures and transmitting data between ORNL and the Census Bureau. Mr. David Clarke of The University of Tennessee Transportation Center produced the Canadian highway network which was attached to the National Highway Planning Network (NHPN) and also modified the FRA rail network by adding mainline classification data.

At ORNL many people assisted in the development or enhancement of the various transport mode networks and in the validation of the mode-specific shipment distances. The authors thank the following staff members for their technical contributions: Dr. Ingrid Busch, Ms. Tai-Lun Chiang, Ms. Angela Gibson, Mr. Robert Gibson, Mr. Kerry Hake, Dr. Ho-Ling Hwang, Dr. Frank Southworth, and Ms. Xiaoming Zhu.

ENDNOTES


2. Each establishment participating in the Commodity Flow Survey was asked to report on a sample of its outbound shipments for a specified two-week period in each of the four calendar quarters of 1993. The sample size for a two-week period depended on the total number of shipments commencing within that period. Establishments with 40 or fewer shipments were asked to report on each one, while those with over 40 shipments were asked to select every 2nd, 5th, 10th, 20th, or 40th shipment depending on the total number involved. The sampling plan was designed to produce an expected sample size...
between 20 and 40 shipments per two-week period per establishment. With 200,000 establishments reporting on about 30 shipments in each of four two-week periods, the total expected sample size was 24 million shipments.

3. The ten domestic modes of transport listed in the CFS questionnaire are: (1) parcel delivery, courier, or U.S. Postal Service; (2) private truck; (3) for-hire truck; (4) railroad; (5) inland water and/or Great Lakes; (6) deep sea water; (7) pipeline; (8) air; (9) other mode; and (10) unknown.

4. The transport mode networks used to estimate distances for the CFS were modified versions of datasets published on CD-ROM by the Bureau of Transportation Statistics (BTS). The highway network was version 1 of the National Highway Planning Network (NHPN) included on the Transportation Data Sampler CD-ROM (BTS-CD-01, January 1993) with the addition of a matchstick Canadian highway network integrated at the border and the Alaska Marine Highway through the Inside Passage. The rail network was a slightly modified version of the 1:2,000,000 scale network produced by the Federal Railroad Administration (FRA) and included on BTS-CD-01. The network was modified by adding a mainline classification link attribute roughly based on traffic volume classes. Minor corrections in the original network were also made. The waterway network was an earlier version of the network published on the National Transportation Atlas Data Bases: 1995 CD-ROM (BTS-CD-06, Bureau of Transportation Statistics, U.S. Department of Transportation, January 1995). The version used in the CFS had fewer shape points on some inland waterways but additional inland waterway ports where intermodal transfers are allowed. The pipeline network was the 1989 National Petroleum
Council products pipeline network included on BTS-CD-06.

5. The locational uncertainty, $e$, of the rail network was estimated to be approximately one mile. The maximum length of an industrial rail spur, $p$, was assumed to be 1.5 miles.

6. The locational uncertainty, $e$, of the waterway network was estimated to be two miles because of the width of many waterways around their centerline representations. The value of $p$ was 0.

7. The value of $e$ was assumed to be 0.5 mile, while 0 was used for $p$ in the local access distance threshold formula.

8. The locational uncertainty, $e$, of the pipeline network was estimated to be 5 miles. The estimated maximum length of a gathering line or local distribution pipeline, $p$, was also 5 miles.


11. The deep sea water mode is not shown in Table 5. Most paths involving deep sea water transport included an inland water segment. Because deep sea and inland water distances had to be reported separately for a shipment, most Tier I deep sea water shipments were promoted to the Tier II deep sea and inland water mode combination.

12. Hancock, Kathleen L., *Spatial Data and Geographic Information Systems within the Bureau of Transportation Statistics (BTS)*, prepared for the Center for Transportation