IMPROVED EMPLOYMENT DATA FOR TRANSPORTATION PLANNING

CTRE Management Project 97-11

OCTOBER 1998

CTRE

**Center for Transportation
Research and Education**

IOWA STATE UNIVERSITY

**Iowa Department
of Transportation**

IMPROVED EMPLOYMENT DATA FOR TRANSPORTATION PLANNING

Principal Investigator
Reginald Souleyrette
Associate Director for Transportation Planning and Information Systems
Center for Transportation Research and Education

Project Team
David Plazak
Transportation Policy Analyst
Center for Transportation Research and Education

Tim Strauss
Transportation Research Specialist
Center for Transportation Research and Education

Mark Imerman
Extension Program Specialist
Department of Economics, Iowa State University

Tim Johnson
Research Associate
Department of Economics, Iowa State University

**Center for Transportation Research and Education
Iowa State University**
Iowa State University Research Park
2625 North Loop Drive, Suite 2100
Ames, IA 50010-8615
Telephone: 515-294-8103
Fax: 515-294-0467
http://www.ctre.iastate.edu

October 1998

Contents

Figures

1 Abstract

Transportation planners typically use census data or small sample surveys to help estimate work trips in metropolitan areas. Census data are cheap to use but are only collected every 10 years and may not provide the answers that a planner is seeking. On the other hand, small sample survey data are fresh but can be very expensive to collect.

This project involved using database and geographic information systems (GIS) technology to relate several administrative data sources that are not usually employed by transportation planners. These data sources included data collected by state agencies for unemployment insurance purposes and for drivers licensing. Together, these data sources could allow better estimates of the following information for a metropolitan area or planning region:

- Locations of employers (work sites);
- Locations of employees;
- Travel flows between employees' homes and their work locations.

The required new employment database was created for a large, multi-county region in central Iowa. When evaluated against the estimates of a metropolitan planning organization, the new database did allow for a one to four percent improvement in estimates over the traditional approach. While this does not sound highly significant, the approach using improved employment data to synthesize home-based work (HBW) trip tables was particularly beneficial in improving estimated traffic on high-capacity routes. These are precisely the routes that transportation planners are most interested in modeling accurately. Therefore, the concept of using improved employment data for transportation planning was considered valuable and worthy of follow-up research.

However, the exercise revealed some significant problems with using the current unemployment insurance and drivers license databases as a starting point:

- The original source data are maintained in formats that proved to be very cumbersome to use for the purposes of the project. These problems could potentially be overcome if the project were institutionalized on a continuing basis.
- There are significant non-coverage and other problems with the databases; many businesses, work locations, and employees simply do not appear in the databases for various reasons. For example, about seven percent of all workers in the pilot area are self-employed and not covered by the unemployment insurance program.
- There are significant address-related problems with the original databases and with baseline Iowa street addressing information that could be used in geocoding; this led to very serious errors in geocoding data for use in a GIS environment. Whether the address matching problems are generated by a poor street network, by addressing problems in the original source data, or by a combination of the two, "hit rates" for address matching of firm and employee locations are very poor in rural areas and in smaller towns and cities. Hit rates of 10 to 25 percent are not unusual in these locations. On the other hand, the overall hit rate was about 60 percent and the rate for large cities often approached an impressive 80 to 90 percent level.
- Data will always need to be presented in an aggregate form to protect the confidentiality of both individuals and businesses. Confidentiality is a very serious concern with the new approach and must be dealt with in a systematic manner.

All of these problems combine to diminish the usefulness of the new approach. However, even given the inherent flaws and limitations, these data would be highly useful for planning highways and public transit systems, and also for non-transportation uses such as workforce and economic development planning. For this reason, it is suggested that further effort go into four areas:

- Institutionalizing the process of gathering, reducing, and merging the original source data to reduce development costs. This would be particularly helpful in the case of the Iowa Department of Transportation's drivers license database.
- Refining the original source data and street address utilities so many of the address-related and non-address related problems encountered in this research could be eliminated. This recommendation applies in particular to the Department of Workforce Development's databases.
- Overcoming problems with obtaining employment data for metro areas that cross state boundaries. This would be needed to implement improved employment data for transportation planning in several metro areas in Iowa, including Omaha/Council Bluffs, the Quad Cities, Dubuque, and Sioux City.
- Developing additional procedures so that the improved data may be more readily used in transportation planning and for other planning purposes.

2 Problem/Needs Statement

A recurring problem in the modeling of transportation flows is the acquisition of quality employment data. Employment data are used in transportation planning to model work trips. Employment data are rarely collected solely for transportation purposes because of the large expense involved in designing and administering surveys. Census data, which are of some use now, are threatened with budget cutbacks and are only collected every ten years. This project developed systems and protocols for utilizing already collected sets of detailed administrative employment data for a new purpose—transportation planning.

The availability of improved employment data would allow transportation planners in Iowa to better model and predict transportation needs from both a commuter and business viewpoint. A better understanding of travel patterns would in turn help to optimize investments in fixed physical infrastructure and public transportation. Once employment data sets have been refined for use in transportation planning, they would also very likely have valuable uses in other fields, such as workforce development planning and economic development. An example of such a use would be the analysis of the transportation and other needs of persons wanting to make the transition from welfare to work.

3 Project Background

This research project explored the utilization of primary non-farm employment data (part of which is generally called "ES202 data") as well as other non-traditional data sources as supplements for more traditional sources such as census data. The Department of Workforce Development (DWD) is Iowa's employment security agency. ES202 data are collected for various purposes, including unemployment insurance and wage and hour reporting. Separate databases include overall information on the employer (by location) and on its payroll by employee. This raises the possibility of having more up-to-date, detailed and accurate work trip data. At least one state, Utah, already provides annual small-area employment data generated from the ES202 database to its state department of transportation and metropolitan transportation

planners. DWD's data have been used for various purposes in Iowa previously, including by economics and transportation researchers at the universities and by transportation planners to identify employers.

4   Literature Review and Survey of Current Practice

4. 1 Published Literature on Employment Data
Published literature on the use of employment data in transportation planning—beyond literature about census data—is very limited. For all practical purposes, this project represents a subject that has not been written about previously. There is an extensive literature on methods for doing traditional, survey based origin-destination studies and replacements for them such as license plate surveys; however, these processes are ultimately very expensive per person surveyed. The result is that use of these data sources rests on a small sample size. This makes drawing conclusions or basing transportation modeling efforts on them risky at best. Estimates made from such small samples can be prone to very large errors.

4. 2 Survey of Other States Current Practice in Using Employment Data in Transportation
Several agencies were surveyed regarding their use of employment data in transportation planning. These agencies were identified through informal inquiries and through postings on Internet mailing lists that focus on issues related to transportation planning and state departments of transportation (in particular TRANSP-L and DOT-L).

Transportation planning agencies have had a wide variety of experiences with their use of employment data. These agencies differ in the specific types of employment data used and in their relationships with state employment security agencies. The Montana Department of Transportation, the Tri-County Regional Council in Lansing, Michigan, and the Utah Department of Transportation were notable for their use of ES202 data, an emphasis of this project. Interviews with these and other agencies revealed a number of difficulties in using such data for transportation planning:

- A relatively high percentage of employees—about 10 percent—do not show up at all in the database; the self-employed are excluded, as are employees working for non-covered and/or non-reporting employers, such as railroads;
- The proprietor of a shop may not be listed as an employee;
- The database often has the address of the payroll office instead of the workplace address;
- The files contain several addresses at post-office boxes, which cannot be assigned to traffic zones;
- There are reporting problems related to franchises and chains of stores;
- There are problems identifying in-region versus out-of-region employers;
- There are problems with the ways in which employers report the number of their employees;
- There are issues related to distinguishing between part-time and full-time employees;
- The Standard Industrial Classification (SIC) codes in the database are given for the entire company even though a company may employ several types of a employees that generate or attract traffic at different rates;
- Employment data are not designed for use in transportation modeling (e.g., there is no "retail" and "non-retail" category);
- The residences of contractors are often listed as their place of business even though most of their work is done elsewhere;

- Some companies report zero employees; this may be the result of highly seasonal operations;
- Inconsistency exists between reporting periods; e.g., some businesses show up in only two or three quarters (mainly due to seasonal workers).

Agencies also experienced several problems related to the assignment of employees to transportation analysis zones, either manually or through address matching. These address-related problems included:

- Address voids (sheer lack of address information);
- Inaccuracy of street names;
- Aliases for street names;
- Incorrect address ranges.

Agencies often augment their employment files with other data, including Dun and Bradstreet data, digital yellow pages, American Business Information (ABI) data, and building permits. One consultant found that there was only about a 50 percent cross-match between any two sources of data tested to date. Still, several respondents stated that, despite the flaws of the databases they use, they were better than nothing.

Following is a more detailed description of interviews with agencies at the national level and in other states.

U.S. Bureau of Labor Statistics
The Bureau of Labor Statistics (BLS) does not keep track of particular users of their data, as it currently does not have a good way to do so. Other sources of economic data were suggested, such as the Economic Census and County Business Patterns. Overall, there are several problems with existing data sets, and limits as to what they can be used for. For instance, it can be difficult to get a work site address from databases generated from tax information. This would limit their use in transportation planning. There are many ways data could be corrupted, and the disaggregation of data often uncovers flaws and anomalies.

Illinois
The Illinois Department of Employment Security (IDES) did, at one time, share data with Illinois Department of Transportation through a data sharing agreement. Employment Security would check documents and review reports before they were published to ensure confidentiality. The data sharing agreement has not been renewed and the DOT no longer has access to ES202 data. Instead, it accesses similar data from an extract of the "administrative" database that has undergone a cleaning process.

Illinois law prohibits the IDES from releasing ES202 data except for purposes stated in law (which do not include transportation planning and research). Thus, it is very difficult to access such data. Transportation agencies have requested data at the firm level. Employment Security says they do not need it.

Wyoming
In the state of Wyoming, employment data are collected and distributed by the Employment Resource Division of the Department of Employment. That office does not send the Wyoming Department of Transportation such data, and it is believed that the Wyoming Department of Transportation uses its own data. ES202 data do not include self-employed, farmers, elected

officials, and some students. It is estimated within the division that roughly 12 percent of the workforce is excluded from the database.

Utah
The Utah Department of Workforce Services (DWS) created software, now running in about 40 other states, used to aggregate ES202 data for the Bureau of Labor Statistics. The Utah DOT has had a contract for about 20 years with DWS to break out ES202 employment data by county and transportation analysis zones (TAZ), using census areas. The DOT pays a fee to have the TAZ number coded directly in the files. DWS gives the data annually, in May, to the DOT in dBase. Data are also provided to the Wasatch Front Regional Council. DWS does not use GIS technology to automate the process, although it may in the near future.

The person responsible for employment data at the Wasatch Front Regional Council recently retired. His replacement is currently learning his new responsibilities. The council uses the data to prepare reports of "surveillance data" of socioeconomic characteristics, employment, and population estimates by TAZ. The employment data come from the DWS, and employment is reported by SIC. Summary tables are available in paper form and electronically (including over the Internet). The tables include fields for record, county, zone, part-time employment, full-time employment, total employment, industrial employment, and retail employment.

Michigan
The Tri-County Regional Planning Commission in Lansing, Michigan, uses ES202 data and other data sources for its employment forecasts and travel demand models. A consultant, KJS Associates, is working with the commission. They consider employment data to be "the biggest headache with the travel demand modeling process." They have encountered several problems with the ES202 files (e.g., franchises, post office boxes, non-covered employees, work site address versus reporting address, invalid addresses, and in-region versus out-of-region employers). The state of Michigan makes the data available to the state DOT and other agencies.

The planning commission assigns the employment locations to TAZs, using GIS to do as much as possible. Unmatched locations are done by hand or are field checked. There are errors throughout the process, such as address voids, inaccurate street names, aliases for street names, incorrect address ranges, and conflation problems between layers which create left/side right side mismatches. Three sources of address ranges have been used, TIGER 90 and two CDs from Caliper CD (both basically based on TIGER). With TIGER, there were missing and incorrect addresses. There was no consistency in the errors over time, and more recent versions were not necessarily better. Major traffic generators often ended up in wrong place. In a small community, employment tends to be easier to track, but transportation planners in major metropolitan areas can quickly become overwhelmed by the number of employment records.

The planning commission has started to use other data to cross-match and verify employment. It has cross-matched the Dun and Bradstreet data and the Digital Yellow Pages and is in the process of adding ABI data. The consultant, KJS Associates, found only about a 50 percent cross-match between any two sources of data tested, which means that the secondary source can be used to crosscheck about 50 percent of the records in the ES202 file. It also means that the ES202 files underreport by about 50 percent of employers; the majority of underreported employees are in the small employer categories—firms with less than 10 employees.

KJS has a research contract with the Federal Highway Administration (FHWA) to work on the employment data problem and develop tools to assist MPOs in addressing problems with

employment data. KJS is finishing up the first phase of approach and is currently preparing a proposal for the second phase.

Montana
The Montana Department of Transportation uses ES202 data provided by the Montana Department of Labor and Industry. The data require cleaning before they can be used. Problems emerge from several sources, including the way some employers report their employment. Incorrect addressing (work site versus reporting address, such as when a central office reports for an entire chain of stores or when a school district reports all teachers as working in the district office) is another common problem. Further, some employers do not report at all because they are not required to, or for some other more questionable reason. In many ways, it is obvious that ES202 databases are not ideally designed for transportation analysis. ES202 files list one SIC code for the entire company, even though a company employs a variety of occupations that generate and attract traffic at different rates. Several types of employees, perhaps up to 20 percent of the workforce, are excluded, such as railroad employees and the self-employed. Contractors usually list their residence as their place of business, although most of their work is done elsewhere, and the proprietor of a shop in the mall may not be listed as an employee. The issue of part-time versus full-time employment is also a problem; part-time employees show up as full-time in ES202 data. However, despite all its problems, ES202 data are seen as a starting point simply because there is no other good alternative.

Building permits (from counties that require them) and assessor records are used to update census data. This can create problems from mixing data sources. Assessor records and building permits are often more accurate than census data, and updating data in this way may generate a higher than actual growth rate. The census has an occupancy rate that is difficult to verify, and which may skew the data. R.L. Polk city directories are also used, again with similar problems of consistency and mixing data sources. The directory often conflicts with the census. For instance, Montana has found places where the Polk has more houses than the census has occupied dwelling units. Phone and power companies have been hesitant to provide data on their customers, although this could be very useful if available to planners.

The Montana DOT uses QRSII for traffic modeling. "Retail" and "non-retail" employment are used, which is not entirely satisfactory since some non-retail employment tends to resemble retail employment (e.g., banks and schools). It also uses occupied dwelling units, which is difficult to update in the field. Employment is allocated to zones using a phone book and a map. The DOT does not use GIS to automate the allocation process through address matching but plans to do so in the future. The DOT starts by using the address in the employment file and updates it as necessary. It does this quarterly; this can cause problems of inconsistency, however, since some businesses show up in only two or three quarters. There are several reasons for this, such as seasonal workers. Also, some reports have zero employees listed.

The Montana DOT has been undertaking this effort for two years (previously a consultant was used). They have completed three cities so far: Billings, Butte, and Helena. A consultant did the rest and the Montana DOT will be updating these.

5 Project and Database Development Process

5.1 Partners Involved in the Project
This project was developed as a partnership involving several state agencies plus two academic units within Iowa State University and a metropolitan planning organization. The Iowa

Department of Transportation (Iowa DOT) provided funding for the project as well as access to a critical database of drivers license information. The Iowa Department of Workforce Development (DWD) provided two employment databases plus a considerable amount of computer programming time. The ISU Department of Economics provided expertise in working with very large administrative data sets in a mainframe environment. The Center for Transportation Research and Education (CTRE) at Iowa State University contributed expertise in GIS and transportation models. The Des Moines Metropolitan Planning Organization provided location-specific knowledge for the pilot area as well as current data and transportation model results for comparative purposes.

5.2 Project Advisory Committee

The following individuals provided guidance and input to the research team throughout the course of the study:

- Kevin Gilchrist, Des Moines Area Metropolitan Planning Organization, Des Moines
- Jennifer Kragt, Iowa Northland Regional Council of Governments, Waterloo
- Mike Lipsman, Iowa Department of Transportation, Ames (Iowa DOT project technical monitor)
- Craig Marvick, Iowa Department of Transportation, Ames
- Dick Sampson, Iowa Department of Workforce Development, Des Moines
- Bob Schutt, Iowa Department of Workforce Development, Des Moines

The advisory committee was selected to provide the perspectives of metropolitan transportation planners, the Iowa DOT, and data providers. Other members of the Midwest Transportation Users Group (MTMUG) also provided valuable input to the project throughout.

5.3 Pilot Region Selection

The selection of a pilot region for this project was subject to a number of important constraints. First, because of the way the administrative data sets used are constructed, only regions wholly contained within Iowa could be chosen. (A related constraint was that different states place different limits on the use of employment data.) This ruled out several metropolitan planning regions in Iowa that cross state lines, in particular the Dubuque, Omaha/Council Bluffs, Quad Cities, and Sioux City metro areas. Second, the size of the databases used meant that this project required a pilot region of manageable size. The research team and advisory committee agreed that an ideal size would be a regional planning affiliation (RPA); there are 16 of these in Iowa and they are responsible for coordinating state-level and local transportation planning. The average RPA in Iowa covers about six counties and 3,500 square miles of land.

RPA 11 was selected as the pilot area for the project based on a discussion held with the advisory committee. This eight-county region is located in central Iowa and contains both the cities of Des Moines and Ames. Des Moines is a metropolitan area with a population of about a half million. Ames is a near metropolitan area and is projected to achieve metropolitan status by the year 2000. The entire RPA includes well over 600,000 persons. It has roughly a half-million employees covered by unemployment compensation and roughly 20,000 business firms. It is a representative cross-section of Iowa in that it contains large cities, small towns, and rural areas. Both Des Moines and Ames have a mix of small and large employers and have a considerable amount of commuting activity, including some long-distance commuting.

5.4 Overview of Employment Database Development Methodology

The development of the new employment database was conceptually simple, but complicated in practice. First, the two halves of the Iowa Department of Workforce Development's (DWD) employment database—the ES202 employer database and the employee database ("wage records") —would be brought together to create a single new database which would contain records that included employees and their place of work.

Second, since the DWD employee data do not now contain an employee residence address, this would be added through the use of some other public or private sector database that contained address information; e.g., another state government administrative database or a telephone directory database. A number of potential candidates were identified for this purpose. Other databases were identified that might be used to improve either the employer or employee side of the new database; however, these were not used in the project since a state administrative database that met the basic requirements of the project was found.

The resultant database was then geocoded using GIS technology so that patterns of employment and residence as well as flows between them could be mapped. This also allowed the data to be used in a transportation modeling package, TRANPLAN. This is the most widely used transportation planning package in Iowa both at the state and metropolitan levels.

In practice, developing the new database and geocoding it proved rather difficult. This had to do with the large sizes of the original databases (in terms of both records and fields), their storage on mainframe computers, and above all, the lack of consistency and quality in street addressing in them.

(A more detailed description of the database development process is documented in Appendix A of this report.)

5.5 Iowa's Existing ES202 Database

ES202 data are maintained for two purposes by the DWD. First, employer files must be kept in order to maintain employer compliance with the employment security system, essentially for unemployment insurance purposes. Second, individual employee files ("wage data") are maintained to track individual eligibility for employment security benefits. In order to create an employment database for transportation planning, the data from these two systems must be brought back together and augmented with individual employee address information, which is not available in either. Individual employee address information must be brought in from another source, whether public or private.

The employer system includes the following complete fields necessary for the process:

- Employer site address;
- Employer SIC code;
- Employer ownership (federal, state, local, private);
- Federal tax I.D. (FEIN);
- Iowa employment security account number.

The employer file has also been set up with a field for the inclusion of a GIS location identifier. The DWD has not until this point had the resources or expertise to fill this field, however. A major interest in this project on the DWD's part is the ability to get some sort of geocoding included in the available information in the future.

The employee system includes the following fields:

- Employee Social Security number;
- Employment Security Account number (which allows a match to the employer system).

It is important to note that the employee system does not include an employee address, as the system was designed to function, unmodified, as employees moved from employer to employer or location to location. Files in both systems include information on payroll that is relevant to the particular system. Employer files have total payroll across all their employees. Employee files contain personal covered income information across all relevant employers.

Both systems are currently maintained as fixed-field flat database files on the DWD's mainframe computer system. Individual records are of reasonable length (up to about 750 characters) for working with in either a mainframe or a personal computer environment. The number of records, however, limits the ability of easily available personal software to do primary manipulations. Basic employer files for Iowa are approximately 70,000 records in length. Basic employee files for the state are well in excess of one million records since there are that many employees statewide.

5.6 Other Potential Employment Data Sources
This section provides a brief discussion of non-traditional data sources beyond those made available from the Iowa Department of Workforce Development, U.S. Department of Transportation, or U.S. Census Bureau that could be used to develop improved employment data for transportation planning. Some of these data sources would be useful on the employee side, some on the employer side, and some for both purposes. Based on a review of available data sources, the most promising sources for this project or similar efforts appeared to be the following:

- Iowa DOT drivers license records.
- City or county property tax assessor data.
- Various commercial sales and marketing databases, particularly from American Sales Leads and Dun and Bradstreet.
- R.L. Polk INFOTYME electronic city directories, which are the electronic version of its well-known city directories.
- Various telephone directory databases, including those produced by US WEST and various private database vendors.
- Iowa Department of Revenue sales tax databases and the Iowa Manufacturers Directory for identifying businesses.

For the purposes of this project, the Iowa DOT's drivers license database proved to be the most useful and most economical source of data to augment the DWD's original databases. The drivers license data were provided at no charge by the DOT and had the employee address data field that was needed to complete the basic database development process. It was the only non-DWD database that was used in the project. However, for information purposes, each potential data source explored is discussed below by general category.

5.6.1 State Administrative Databases

The Iowa DOT Motor Vehicle Division maintains a database on all the licensed drivers in the state. The Driver License Master File contains several fields with potential application to transportation planning, including name, address, and driver license number fields. In Iowa, the driver license number is often, although not always, the driver's Social Security Number (SSN). Drivers may choose not to use their SSN as their drivers license number in Iowa, although most do use the SSN. This database is available free of charge to government agencies and for university research. However, it should be noted that the drivers license database could only be made available in its entirety on a statewide basis and that it was made available in a mainframe tape format. This made for a huge data processing job to obtain the few needed fields and records for the pilot region.

The Iowa Department of Revenue and Finance (IDRF) maintains a large database of companies in Iowa that have Sales and Use Tax permits. The database includes address and firm name information as well as classification by standard industrial classification (SIC) code. This database includes essentially all retail companies and many manufacturers in Iowa. These data have been made available for university research projects free of charge in the past with some important restrictions to preserve confidentiality.

IDRF also collects the state income tax. Unlike the sales tax database, the income tax databases and all the data contained in them are considered confidential for both corporate and individual income taxpayers. IDRF is particularly concerned that Taxpayer Identification Numbers remain confidential. (TINs are the same as Social Security Numbers for individual taxpayers.)

The Iowa Secretary of State maintains a large voter registration database. It contains records on all the registered voters in the state, including address information and voter identification numbers (which are usually the voter's Social Security Number.) However, under Iowa law this database may only be used for election-related purposes. By law, it cannot be used for non-election related purposes such as research or transportation planning.

The Secretary of State also maintains information on corporations and limited partnerships registered in Iowa. Sole proprietorships and simple partnerships are registered at the county level in Iowa.

5.6.2 Local Government Data Sources

Many city and county assessors in Iowa maintain detailed property tax records databases. Although these databases can vary significantly from place to place, all contain address and owner fields for parcels. It may be helpful that many of the counties in Iowa now use one of two vendors for their land records software—CMS or Solutions. Property tax records are considered to be public records and could be used for transportation planning purposes. Because of a unique field that identifies homestead credit exemptions, these databases could be used with other data sources to identify home-based businesses.

According to a recent Iowa Department of Education survey, more than half of all local school districts in Iowa maintain some sort of detailed database on their student populations including address information. In general, the larger the district in terms of number of students, the more likely it is to have a student database. Small, rural school districts are less likely to have such databases. The nature of the databases (format, fields, etc.) and policies on their use by non-school organizations are entirely up to the local districts. No standards are set at the state level and none is likely to be set.

Many counties in Iowa for emergency response purposes develop enhanced-911 databases. These databases are public records. However, in order to be truly useful for emergency response purposes additional data such as names and telephone numbers are essential. Telephone companies (see below) add these. Once names and telephone numbers have been added, these databases become the property of telephone companies in Iowa according to Iowa law.

5.6.3 Private Sector Data Sources
American Business Information (ABI) through its subsidiary American Sales Leads (ASL) is one of two large companies that sells detailed information on businesses, including SIC codes and employment ranges. The ASL database also includes listings on educational and governmental institutions. ASL claims 98 percent or more accuracy on its database and can provide the actual number of employees for large employers (businesses with more than 100 employees). The cost is about 3.5 cents per listing, which would equate to about $700 for coverage of a study area with 20,000 businesses—roughly the size of the pilot area for this project.

The other major business information provider is Dun and Bradstreet Information Services (D&B). D&B collects and compiles information from a variety of public and private sources and then resells it. D&B's data cost considerably more than ASL's, about 75 cents per listing, because they contain business credit information. It would cost about $15,000 to purchase a complete D&B database for the RPA 11 pilot study area.

The R.L. Polk Company produces numerous city directories in Iowa that contain information on both households and businesses. Polk enumerators, who go door-to-door, collect data for these directories. Until very recently, these directories have only been available in hard copy. However, in the last several years, Polk has begun producing electronic versions of its directories called INFOTYME databases. Polk directories are now available in electronic formats (diskette or CD-ROM) for several cities in Iowa; more are being produced this year, including one for Des Moines. The cost of an exportable city directory database is approximately $400 to $500 per medium-sized city (based on prices for the recently published Iowa City and Waterloo directories). This is about twice the price of a printed city directory. Obtaining all the INFOTYME directories needed to cover a large study area could easily cost tens of thousands of dollars and might not include some unincorporated areas.

US WEST Data Products Group, based in Denver, produces very current Marketing Resource Lists of both households (Consumer Lists) and businesses for Iowa and 13 other Western states. These include names, addresses, phone numbers, and (usually) "Zip + 4" zip coding information. The Consumer List databases also include about 40 demographic variables that can be used to refine the list. These are most often used for direct marketing purposes. The address listings incorporate the E-911 data produced by counties. The US WEST lists are moderately priced. For example, the lists for Johnson County (including Iowa City) have about 27,300 residences and 2,900 businesses and cost about $1,560 and $165, respectively, or a little less than 6 cents per listing. However, they are updated daily and US WEST indicates that they are over 95 percent accurate. A business list with 20,000 listings would likely cost about $1,200.

Several companies produce national telephone directories on CD-ROM with street prices under $100 for all the white pages listings in the United States. These databases are developed using local exchange telephone carrier (LEC) directory databases (such as those originally produced by US WEST). An example is Pro CD, Inc. Select Phone, which takes up 9 CD-ROMs. American Business Information, Inc. (ABI) produces a more specialized and higher quality product with listings of 10 million U.S. businesses on a CD-ROM (American Yellow Pages).

This product costs about $600. Other telephone listing products are also coming on the market. The data in any of these listings can be assumed to be at least a few months old.

The Harris Publishing Company produces the Iowa Manufacturers Directory along with directories for the nation and many other states. This product provides detailed listings for all the manufacturing companies in Iowa, including addresses and number of employees. This product is also available in an electronic database format (on diskette or CD-ROM) for $325 for the entire state. It is updated every year.

5.7 Data Confidentiality Issues

The DWD employment data used in the completion of this project is heavily access restricted to protect the privacy of both the employers and the employees involved. The availability of the new database created for this project will also be constrained due to the confidentiality requirements of the data. There is no way to assure confidentiality with data records that include specific location data on both the place of residence and the place of work for individuals. The risk of confidentiality problems increases as the final working database is made available for direct use by transportation planners and others.

Simply aggregating data in the new database by geographic region such as traffic analysis zones (TAZs) could considerably improve confidentiality. Because the database being constructed could ultimately be released in such a zone-to-zone format, confidentiality problems with respect to employees could be largely eliminated. There will always be enough employees in a zone to protect their privacy. Major employers within a zone, however, could continue to stand out as obvious destinations within the system, particularly where they are located in a sparsely developed TAZ. This is a problem that could not be easily resolved. As a result, some operating guidelines would need to be developed for distributing and using the new database.

Ideally, any final employment data produced would be available on a World Wide Web site such as Iowa PROfiles. This would certainly be possible from a technical standpoint, and confidentiality issues could be addressed through restricted access to selected parties. There are really two sets of confidentiality issues involved:

- Confidentiality of the raw data;
- Confidentiality of the final product—TAZ aggregated data..

Within the final product data, there are issues of confidentiality with respect to

- Employers;
- Employees.

Data confidentiality restrictions are a very serious issue from the standpoint of the employers, employees, the DWD's ability to make data available, and our ability to get the data. There are varying opinions concerning who owns the ES202 data. The current position of the Bureau of Labor Statistics at the federal level is that the data are federal property. The BLS does not authorize our access, and would prevent it, if reasonably possible. Iowa is one of the few states that now allows direct access to these data for continuous research.

Under these conditions, system management staff at DWD would prefer that the raw data not be made available beyond the immediate university researchers. Providing raw data access to

COGs or RPAs through CTRE would take individual employee and income security considerations out of DWD hands and put DWD at considerable risk.

However, DWD has a strong interest in model development, and is very interested in the continuous provision of data for planning. The DWD is considering whether data developed and available within the zone-to-zone model, while still containing restricted content, could be made more easily accessible through researchers to local planners. Given that the confidentiality issues within the new database (if aggregated by zone) are primarily employer rather than employee issues, strong consideration is being given to this option. Under this scenario, the researchers would be responsible for the acceptable distribution of the zone-aggregated data to qualified users with explicit agreements to maintain the confidentiality of the content of the data.

5.8 New Employment Database Specifications
The specifications for a usable employment analytical database were very simple. It must be small enough to be reasonably used on a current model personal computer and be in a format that could be readily imported into common transportation modeling (e.g. TRANPLAN), GIS (e.g. MapInfo or ArcView), and database management (e.g. Access) software.

In terms of database fields, the resultant new database was fairly simple. For records that could be fully matched, it contained:

1-9	Social Security Number
10-15	Unemployment insurance number
16-19	SIC Code
20-22	FIPS Code
23-57	Firm Name
58-92	Firm Address
93-120	Firm City
121-122	Firm State
123-131	Firm Zip Code
132	Address Type 1-Physical 2-Mailing 3-Administrative HQ 9-Unknown
133-152	Employee Last Name
153-167	Employee First Name
168-182	Employee Middle Name
183-185	Employee Suffix (Jr, Sr etc)
186-193	Employee Birthdate
194	Sex
195-218	Address
219-234	City
235-236	State
237-245	Employee Residence Zip Code
246-247	County
248-255	Process Date

However, because the number of records is very large, the resultant database for the pilot region was at times a challenge to use, especially in a GIS environment. The overall process used to create and test the new database is illustrated in Figure 1.

Improve Databases Using Additional Data Sources and Other Applicable Techniques

Input Original Data (DWD ES202 and Iowa DOT Drivers License)

Develop Initial Trip Generation and Origin/ Destination Tables

Identify Best Opportunities For Improvement of Data and Costs Of Improvement

Assess Initial Overall Database Errors And Accuracy

Database Issues

DB/ Matching Issues

Sample Error Assessment

Use Trip Generation and Origin/Destination Tables In Pilot Applications and Models

Validate Models And Identify Application-Specific Problems With Tables

Are Data Adequate As Is?

No

Quantify Potential Benefits of Improving Databases

Yes

Figure 1: Employment Database Development Process

5.9 Database Address Matching and Geocoding Test

Some initial testing was done on the employer file system to determine the ability to geocode the information using ESRI Arc/Info GIS' geocoding capabilities. Just under 19,000 records were identified with a FIPS code matching one of the eight pilot area counties (Polk County). This test area comprised just under 25 percent of the businesses in the state that report unemployment insurance information.

In a sample of 1,000 records in this test area, over 90 percent indicated the address given was the physical address where employees worked. This defines the population that could then be tested for geocoding accuracy. Including the remaining 10 percent (where a physical work location was not given, but instead something like a post office box number was supplied) in a journey-to-work model would require locating physical addresses for these employers. This would potentially be an expensive and time-consuming task.

Geocoding success on this sample varied from about 50 percent in rural areas to 75 percent in urban areas. Most errors were due to the addresses contained in the original employer file. Rectifying this situation would also require the location of matchable physical addresses for these records. With this success rate, there may be as many as 6,000 addresses that would need to be checked and corrected. Other geocoding errors were due to the lack of address ranges in the ESRI Street Map geocoding reference program itself. These omissions were often concentrated in smaller cities and rural areas. For these addresses the geocoding reference ranges are blank. Without these numbers, ESRI Arc/Info (the GIS program used as the platform for geocoding) ignores the address in the employer file.

This test helped to identify a number of serious weaknesses that would crop up when the new employment database was used in transportation planning and other applications.

5.10 Database Quality Considerations

5.10.1 Non-Address and Geocoding-Related Issues

Considerable non-address related data quality issues exist with respect to employee non-coverage and business reporting and work location. Some large population groups are excluded from coverage by the unemployment compensation system. These include the self-employed, independent contractors, sole proprietors of small businesses, and certain groups of employees (such as railroad workers) covered by their own unemployment insurance system. Although the percentage of non-covered workers will vary from state to state, a reasonable estimate of non-coverage for Iowa employees would be roughly 12 to 15 percent.

A number of potentially more serious problems exist on the employer end of the ES202 databases. These problems include the following:

- Sole proprietorships may not report at all since they consider themselves to have no employees.
- Franchises, chain establishments, and branches may report from a central location that is not the actual work site. Out-of-region employers may report from that address.
- Diversified businesses may report under a single SIC code which is not totally descriptive of their lines of business.
- Start-up and bankrupt businesses may report zero employees some quarters because they are filing tax returns.
- Seasonal businesses may show up in some quarters and not others.

- Part-time workers may appear to be full-time employees, giving an inaccurate picture of employment at a location.
- Independent contractors may give their residence as their work location rather than the location at which work is actually done.

Up to 50 percent of all potential business work locations in Iowa may in fact not show up at all in the ES202 database. Most of these unreported work locations will be for very small businesses with fewer than 10 employees, although some significant problems may also exist for very large employers with multiple work locations in an area, such as supermarket chains.

More serious database errors are likely in terms of the location of businesses than in terms of the locations of employees. It appears likely that errors will be large (potentially in excess of 50 percent) when other data sources—either public or private—are related to the ES202 databases. This essentially means that the employment database cannot be anywhere near the "population" that was originally hoped for. It must instead be thought of as a very large sample. However, this should not be seen as an insurmountable drawback to this approach since a very large sample can in fact represent an improvement to the existing situation.

5.10.2 Address- and Geocoding-Related Issues
Another set of problems that was identified in the development of the database had to do with street address matching or geocoding the data. The major apparent failure in geocoding appeared in rural areas where counties are responsible for maintaining street and road addresses. In these areas the lack of specific emergency response (911) addressing or of up-to-date geocoding utilities (utilities that had already incorporated recently instituted emergency addressing) made address matches highly sporadic. About the best that could be accomplished in these areas was a 50 percent geocoding "hit rate."

The second consistent problem was currency and accuracy of street databases in incorporated areas. There turned out to be considerable variation in hit rates within urban areas in the pilot region.

Both of these problems would argue that more attention be paid to developing and maintaining a useable interjurisdictional baseline of GIS street addressing information for the state of Iowa. The major beneficiaries of this would probably be in the transportation and emergency management communities, but other agencies, businesses, and organizations would benefit as well.

The institution or improvement of addressing conventions for such things as street representations (St., Street, etc.), abbreviations, unit numbers, etc., within both the Iowa DOT (for the drivers license database) and the DWD (for the employer and employee databases) would, undoubtedly, increase the ability to successfully geocode the final database.

There is also some question regarding the mobility of workers and the currency of their drivers license addresses. Many mobile young adults maintain a hometown address on their licenses because it is convenient and routes many recurring transactions through a long-term stable address—their parent's. It may be speculated that this may not be the case with regard to corresponding automobile registrations, as registration is automatic with the county in which it is accomplished (imposing a marginal transaction cost on the maintenance of a historical address). A routinely scheduled match of drivers license addresses against vehicle registration addresses might provide the Iowa DOT with the means to increase the currency of its drivers license

information. This could also provide a means for improved administrative auditing and enforcement.

In addition to "no match" on Social Security Number, this error rate includes employees that carry out-of-state drivers licenses or have no valid drivers licenses. Youth can be and are regularly employed in covered positions at age 14. In a college town like Ames, or, to a lesser extent, Des Moines the number of out-state drivers licenses associated with part-time employment is probably significant. There is no way to know how large, but it might be expected to be predominantly in the service and retail sectors. A statistical evaluation of the populations that matched and did not match by area and SIC codes might provide valuable insights into the incidence of these types of problems.

It was noted that the ability to geocode employee addresses in the final database was slightly higher than the success rate on employer addresses. We expect that much of this is accounted for by the fact that employee information was prescreened through matching DWD with DOT data prior to geocoding. The fact that an individual's data were consistently available in both sources is a probable eliminator of a high proportion of problematic records.

The overall geocoding success rate was only about 60 percent and was similar on both employer (firm) and employee addresses. Good matches (based on a 75 percent score in ArcView) were slightly better for employee addresses than for employer addresses. So, more than one third of all the database records are "lost" just in the geocoding phase of the development process. Using ArcView on a personal computer, about 2,000 items could be geocoded in an hour. To geocode data for the entire state of Iowa would take approximately 30 days at this rate.

(See Appendix A for more details on the geocoding process and results.)

5.10.3 Analysis of Spatial and Other Biases in the New Database
Several analyses were performed on the employment database to determine whether any significant spatial biases exist. The biases that were found are, unfortunately, quite large. The most important of these are described below.

Although the non-coverage rate for the database due to self-employment was found to be about 6.5 percent overall for RPA 11, this rate varies significantly by location. The figure for rural counties is much higher. For instance, the self-employment rate for Madison County (which is the most rural county in RPA 11) was over 13 percent, while the figure for highly urbanized Polk County was only five percent. This suggests that the employment data must be used with much greater caution in rural areas and smaller towns and cities.

The "no-match" percentage generated when DWD employee codes could not be matched to drivers license numbers varied both by industry of employment and the origin of employees' Social Security Number. In particular, no-match rates are higher for employees who work in retailing (a high-turnover industry both in terms of employees and firms) and for employees who have relocated to Iowa at some point. This suggests that churn in both employees and employers is a significant concern in terms of bias and error. The database is much more accurate for long-term, stable residents and for more stable industries.

Geocoding success rates were also found to vary greatly by location. For example, The "hit rates" for firms and employees for address matching were 67.4 percent and 66.4 percent, respectively, in Polk County, which is the most urban county. On the other hand, the geocoding "hit rates" for firms and employees in more rural Madison County were only 10.4 percent and

25.9 percent, respectively. This represents a very high error rate and appears to be due to a combination of poor addressing in the base street map used for geocoding and poor address data quality in the original source databases used to build the improved employment database.

Geocoding hit rates by city also illustrate this large source of spatial bias. The hit rate for the best city for employees (Des Moines) was almost 90 percent, while the rate for the worst city (Madrid) was less than 30 percent. The results were even more extreme for geocoding of firms. In Winterset, which is a small city located in Madison County, only a 13 percent match rate was achieved. Again, the use of the new database in the more rural portions of the pilot region (RPA 11) would be subject to very large errors. On the other hand, the accuracy rate inside the urbanized area of Des Moines would be very high.

(See Appendix B for additional detail on spatial and other statistical biases introduced during the creation of the database.)

5.11 Database Development Costs
Database development costs were a significant part of this project. As with any pilot project, the high costs incurred here reflect the need to invent and learn many new processes and procedures. The costs could be reduced substantially in a statewide expansion, especially if the two participating state agencies would play an active part in the process. Database development costs fell into four categories for this project:

- Acquisition costs
- Reduction costs
- Merging costs
- Geocoding costs

If the project effort were to result in the institutionalization of employment data development, most of the acquisition costs could be eliminated and data merging costs could be significantly reduced. Geocoding costs are probably the most scalable of the costs encountered in the pilot and are the most relevant with respect to a continuation of the effort. They would basically go up with the number of employees and employers contained in the final database.

Data acquisition consisted of finding out what the necessary databases contained, how they could, potentially, be matched, and whether the matches would actually provide a success rate that would justify continuation. This process with the DWD took about 20 hours of researcher time working through what was available, how it was available, and what the DWD would require to produce the data in a format that could be used. The researchers and DWD finally settled on acquiring the complete state quarterly employer and employee files from the ES202 system and reducing/matching the files at Iowa State University. DWD could have done some of the programming, if necessary, but scheduling its participation was not necessary to complete this portion of the project. Statewide DWD data on an individual employer and employee basis came on two magnetic tapes, one for employers and one for employees. DWD tapes were shipped almost immediately following the determination of which files would be used.

Some researcher time was also involved in testing small subsets of the data to see if consistent employee/employer matches were possible and whether the employer addresses could be geocoded with reasonable success.

The data acquisition phase also involved acquiring the drivers license data from the Iowa DOT. This required no investigation of content since excellent documentation was provided. Actually obtaining that data, however, was more problematic than dealing with the DWD employer and employee files. First a detailed request for data had to be prepared and submitted. In response to the detailed request, the researchers were informed that the information was only available as a complete set of multiple tapes (in series) for the whole state with all variables. After a considerable number of fits and starts, the data eventually delivered turned out to fill 10 tapes. About 40 hours of researcher time was involved trying to fill this request for publicly available data before a set was delivered that could be worked with. Nearly all of this time and expense could be eliminated if the database development effort were institutionalized within the Iowa DOT.

Data reduction was also a very time consuming task within the pilot project but would also be nearly nonexistent if the Iowa DOT were to take an active part in any institutionalized effort to continue this project. The employee file from DWD and the drivers license file from Iowa DOT both included more than 1.5 million records. These files were too large to manipulate using standard user accounts on the Iowa State University mainframe computer. This was not due to a hardware/software problem, as the system could easily handle the files. It was an access rights problem in a time-shared environment. In order to provide widespread access to the public system, individual ability to monopolize the system must be limited. In this case, file sizes exceeded the capacity limits.

This problem could have been easily avoided with the DWD data, as they would have scheduled the programming if asked to reduce record sizes to include only those fields necessary for this project and select only the counties in the pilot area. The Iowa DOT, however, indicated that they would not reduce the sizes of files coming out of their system, so work-around was necessary.

Reduction of the two files required expanding the DWD packed decimal file formats, querying for the records matching the pilot county region, and reducing the record size to include only necessary fields. For the Iowa DOT file, the process required consolidating the multiple tape series on hard disk drive space, querying for the records matching the pilot county region, and reducing the record size to include only necessary fields.

This process required approximately two days of working time by a researcher and a consultant in the Iowa State University Computation Center. Using a consultant accomplished two things. First, it provided unlimited access to mainframe computing resources. On the other hand, it added yet another individual to the process who was completely unfamiliar with the data sources, necessitating significant oversight time that would have been unnecessary if the database work were institutionalized.

Data merging consumed about 60 hours of researcher time and 40 hours of computer consultant time. Much of this time was spent checking the resulting data files, determining how match variables might be reformatted or reread to improve matches, and reiterating the process. Within the context of a pilot project, this effort will have to be recreated each time a new data handler is found or each time a legacy system is replaced. In this case, the system used for data operations is not available after June 30, 1998, since Iowa State University is retiring its mainframe computer system and adopting a more distributed computing model. The alternative system at ISU, the Project Vincent distributed research network, faces an uncertain future as a centrally supported system.

Database geocoding is the only part of the database project that does not stand to be significantly less expensive within an institutionalized environment. The merged data files need to be geocoded for both employer and employee addresses.

There was interest at the outset of the project in DWD institutionalizing a geocoding effort for its employment files. This is probably not a realistic expectation. While recognizing advantages from participating in this effort, the DWD has no direct mandate or incentive to fund such an effort, and would be hard-pressed to maintain and update a geocoding system. Since the primary (but by no means only) beneficiaries of improved employment data would be the Iowa DOT, it might make sense if the DOT committed to geocoding all the data for this effort. DWD would expect that a copy of the geocoded employer and employee files be provided back to them as a byproduct of the effort, and as a return on the DWD's participation.

The geocoding effort for the pilot area consumed nearly two weeks of researcher time and 64 hours of student help in the Iowa State University GIS lab. In addition, the GIS Support Facility director, Kevin Kane, invested about 10 hours in arranging and supervising the student effort in this process. These costs would be almost directly scalable in expansion to a statewide effort. Variables involved would include the quality of the geocoding databases/utilities used and the speed of the computers utilized in the process.

## 6 Illustrative Applications of Improved Employment Data

6.1 Transportation Applications

6.1.1 Potential Regional Transportation Planning Applications
There are many potential applications of an improved employment database such as has been created through this project in transportation planning. These include:

- Bypass analysis;
- New interchange justification;
- Gauging the impact of a new, major employer on the transportation network;
- Other types of traffic impact analyses;
- Estimating the extent of reverse commuting and commuting within the suburbs;
- Estimating the potential impacts of telecommuting on transportation patterns;
- Visualizing journey-to-work patterns by socioeconomic status and type of employment;
- Estimating the impact of variations in inputs (employment levels, spatial scale, boundaries, etc.) on results of transportation models;
- Air quality analysis.

Several of these concepts are illustrated in maps below. Figure 2 is a map depicting all of the gasoline stations and eating and drinking establishments in Indianola, Iowa that could be successfully geocoded. These businesses are nearly all located along the primary highway routes (US 65, US 69, and IA 92). These are types of businesses that could potentially be impacted the most by a bypass of the city along any of these routes or by a major change along any of these routes. Using the GIS, the individual businesses can be identified very quickly and compared with various bypass scenarios. The employer database behind the GIS map allows the business names and addresses to be accessed immediately.
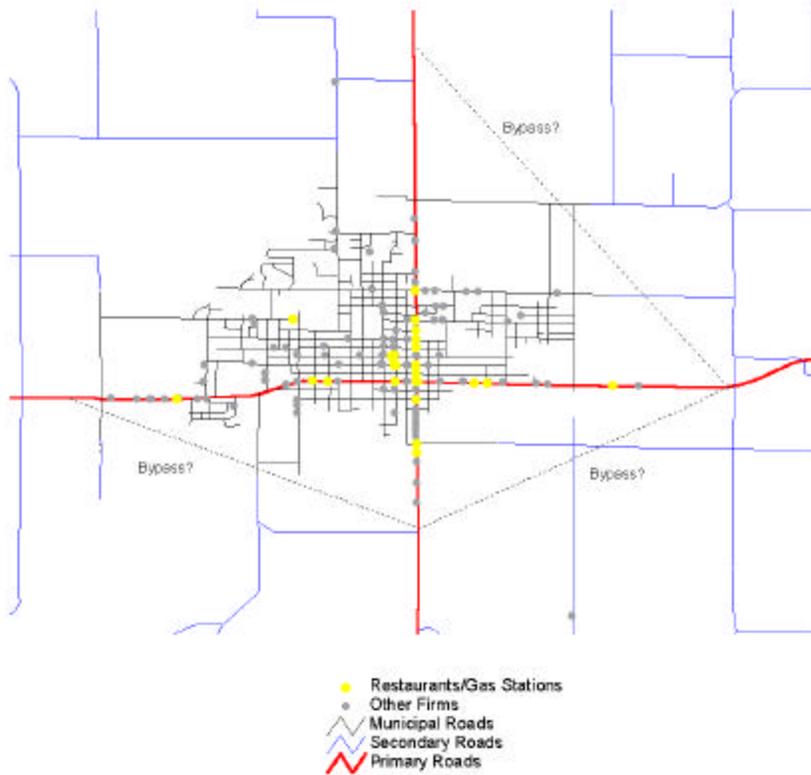
Figure 2: Indianola, Iowa Business Locations by Type and Bypass Scenarios

Figure 3 depicts all of the businesses that could be geocoded in Ames by major type (for example, retail, service, and other). The pattern of location clearly varies by type of business. Almost all the retail businesses are located in clearly identifiable clusters (e.g. downtown, shopping malls, etc.) along major arterial streets. On the other hand, some of the service businesses are located in residential neighborhoods, indicating home-based businesses. The heavy retailing areas may be good candidates for attention regarding such issues as access management or parking studies.

Firms
⚡ Retail
▲ Service
• Other (manuf., agric., transp., etc.)
Primary Roads
Secondary Roads
Municipal Roads
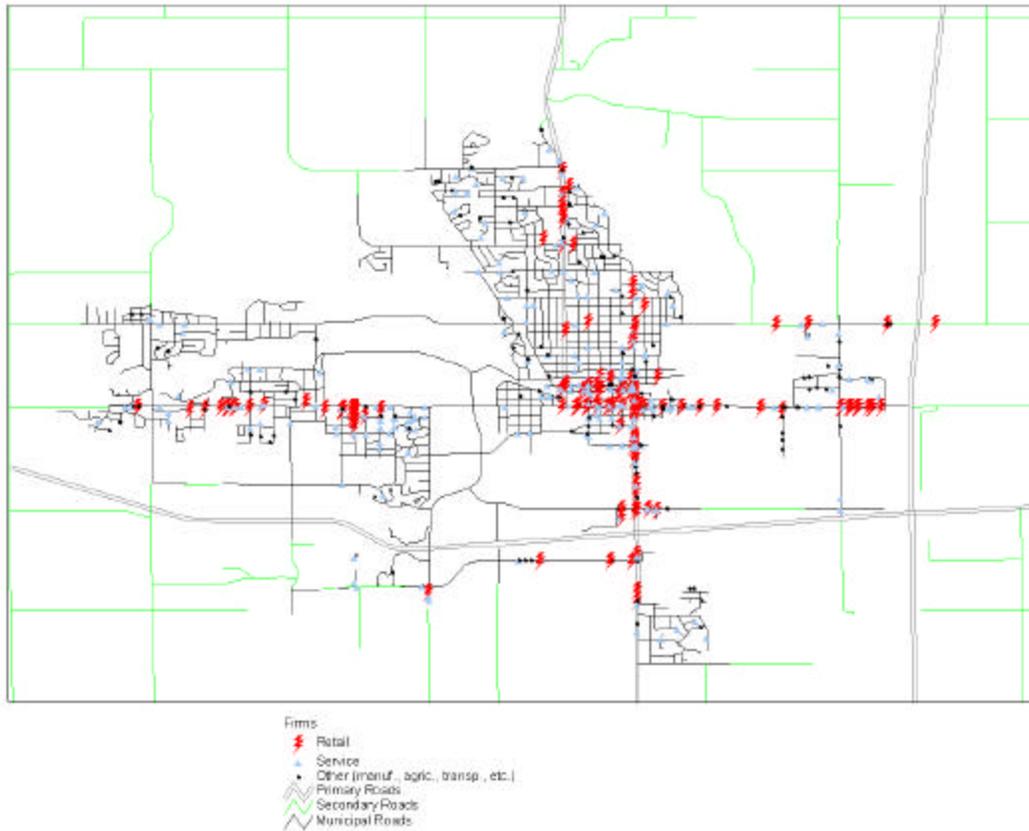
Figure 3: Ames, Iowa Business Locations by SIC Category

There are also a number of potential uses of improved employment data in the fields of Transportation System Management (TSM) or Transportation Demand Management (TDM). These include the following:

- Impact of TDM—research has been done that links response to TDM with family status, income, and employment status;
- Impact of transportation control measures (TCMs) and transportation systems management (TSM) measures;
- Flex time impacts;
- Impact of local laws requiring a traffic impact analysis for employers generating over a certain number of trips per day.

A number of public transit-related applications would also be possible. These might include the following:
- Mapping public transit routes versus work trip flows;
- Vanpool market analysis;
- Evaluation of potential rideshare programs (there are ride matching GIS programs in use by some agencies around the country);
- Paratransit/demand responsive transit planning.

**6.1.2 Specific Applications in Transportation Model Improvement**
There are many specific applications of improved employment data in transportation modeling. These include improving trip production equations, improving trip attraction equations, recalibrating friction factors (used to calibrate gravity models), developing "K factors" (socioeconomic adjustment factors), and developing relationships between friction factor and distance that depend upon socioeconomic characteristics of the from and to zones.

Transportation modeling generally involves a four-step process: 1) trip generation, 2) trip distribution, 3) modal split, and 4) traffic assignment. Although improved employment data could potentially be of some use in all four steps of the modeling process, it is of particular benefit in terms of trip generation and trip distribution. This is because the employment database can give us much more accurate information than census data or a small sample database on employees' home-based work trips—their commuting activity. Since commuting tends to be concentrated in time and subject to large peaks, it is often the single most important consideration in transportation planning, modeling, and forecasting.

For this project, the employment database was used to synthesize new home-based work (HBW) trip tables for the Des Moines metropolitan area transportation model. The end result of this work (described in great detail in Appendix C of the report) was an "improved" set of HBW trip tables. An evaluation of the systhesized data indicates that the use of employment data as in this project can lead to small, yet strategically important, improvements in transportation model results.

The coefficiant of determiniation (R^2) for the original Des Moines base1990 model was 0.901 in terms of the ability of the model to estimate actual ground counts of traffic. The synthesized trip tables produced via the employment database produced modest improvements in the R^2 comparing the estimates produced by the model with the actual ground counts for links. In particular, the variable percentage method of producing the HBW trip tables produced an R^2 of 0.911, or a one percent improvement in accuracy. (See Appendix B for a detailed explanation.) The base1990 model had 564 links with errors beyond the maximum level recommended by the Transportation Research Board, while the "improved" model produced via the employment data project reduced this number to 539 links, a more significant 4.4 percent improvement.

What is more important than the one to four percent improvement in model accuracy is the fact that the improvements in link volume estimates were not random across the network. Instead, the model using the synthesized HBW trip tables yielded improvements that were especially concentrated on high traffic capacity links such as freeways, expressways, and major arterials; these are the links that regional transportation planners are most interested in accurately estimating and forecasting travel for. The improvement is particularly noteworthy for the highest capacity links, where the variable model improves conformance with tolerance by 58 percent for links over 20,000 average daily traffic (ADT) and 63 percent for links over 30,000 ADT. These high capacity links tend to carry an inordinate amount of commuting trips, so having more accurate HBW trip tables is most beneficial for estimating traffic on them. Figure 4 illustrates the improvement in estimates.
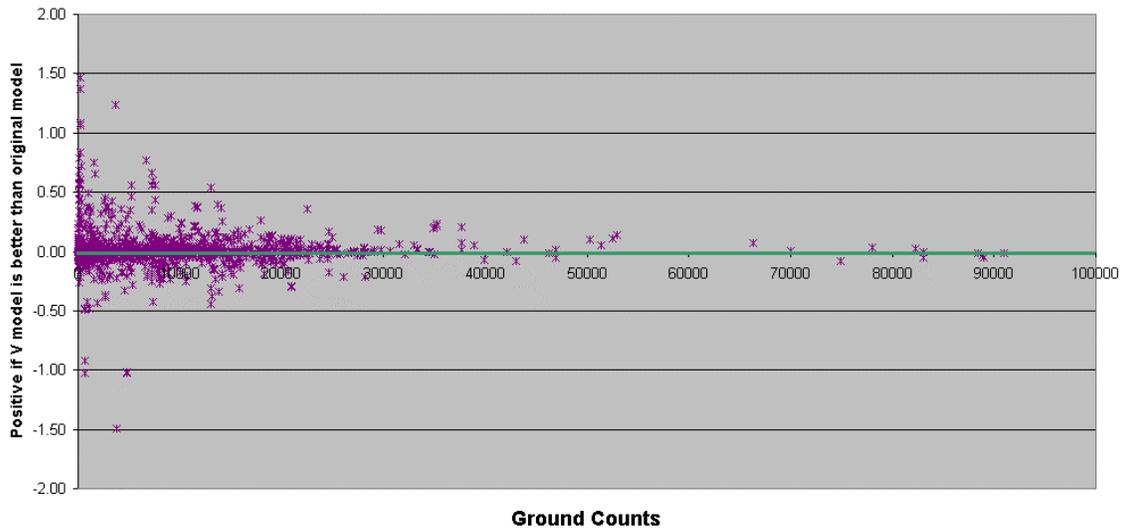
Figure 4: Improvement in Traffic Estimates with   Variable Percentage   Model

6.2 Economic Development and Workforce Development Applications

The institutionalization of a continuous statewide journey to work modeling effort as piloted here would also have many uses in economic analysis and development. Some of these uses might include the following:

- Identifying welfare-to-work transportation needs;
- Locating workforce development centers and employment information kiosks;
- Locating new industrial and commercial park sites (e.g. places with large amounts of out-commuting);
- Identifying spatial mismatches such as jobs-housing imbalances;
- Analyzing locations for new child day care centers or services.

Currently, most impact models assume that the entire labor force affected by an economic event reside in the locality of the impact (ignore commuting altogether). Better models incorporate gross flows, utilizing data available from the decennial census, but these give no indication of industry, occupation, or wage rate (job quality). These are significant issues within the ongoing debate as to the push or pull nature of commuting in rural areas and also in many discussions of state and local economic and community development policy.

Increasing the specificity of commuting data (geographically, occupationally, industrially, etc.) would allow a substantial increase in the ability of researchers and planners to conduct local economic analysis. This would increase the understanding of workforce choices regarding places of employment and residence, as well as the tradeoff between continuing commuter expenses and household capital costs. As a result, more accurate fiscal and economic impact models could be constructed. Better predictions of the impact of public infrastructure investments on land values could also be made.

An example of the use of the database for welfare-to-work planning is shown below. This map shows the location of matched and geocoded computer and data processing services businesses

in the Des Moines metropolitan area. According to DWD labor market forecasts, businesses in this SIC code (737) around Des Moines will be among the fastest growing industry groups, creating several hundred net new jobs over the next five years. These will partly be entry-level jobs that could be filled by current welfare recipients, who are often females heading single parent households. The female, working aged population living below the poverty line is concentrated in and just east of the central city, according to 1990 census data. However, as the map also shows, about half of the businesses in SIC 737 are scattered around the periphery of the metropolitan area (especially in the northwest suburbs). The other half is located in and near the Des Moines central business district, many on and near one thoroughfare (Grand Avenue). There is a partial spatial mismatch between the location of jobs and the location of a potential workforce. This mismatch could potentially be addressed through transportation initiatives, for instance a circulator transit system in the western suburbs.

The map in Figure 5 was created by combining employer data created with the ES202 files with census demographic data from the Census CD in ArcView GIS. An even more useful analysis tool could be created by introducing a database containing information on the residential locations of potential welfare-to-work clients.
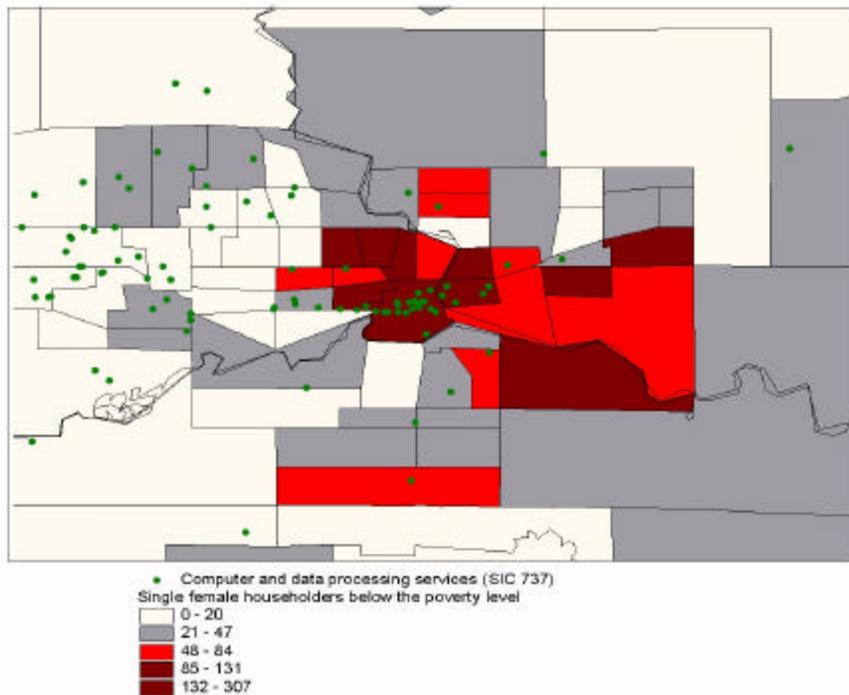
Figure 5: Comparison of Potential Employment Locations with Location of Low-Income Single Female Headed Households in Des Moines, Iowa Metro Area

7 Database Evaluation and Recommendations

7.1 Costs and Benefits of Statewide Expansion and Replication

As noted above in the discussion of database costs, the most important requirement for statewide expansion is the active participation of the Iowa DOT in database development. This would reduce many of the database development costs and routinize them so that an annual reinvestment in the process would not be necessary. A major problem faced by the research team was the huge size and cumbersome nature of the drivers license database.

The only database development cost that directly scales with an increase in project area is the geocoding component. The pilot area included over 33 percent of the state's work force. It is estimated that statewide geocoding required could be completed given about 400 hours of work (0.2 FTE) on the part of a technician. With this one notable exception, the cost of going statewide with the current databases and processes would be essentially the same as the cost of doing one pilot region.

Although the cost involved in developing the employment database, even on a statewide basis, would not be overly large, development would represent a real benefit for practicing transportation planners. The benefits of having the improved employment database would come in terms of the ability to improve transportation model results in places that already have models, e.g., all of the MPOs in Iowa. Based on experience in Des Moines, these improvements would appear to come largely in terms of the ability to predict traffic volumes on major, high-capacity routes. Since this is a main goal of transportation modeling in metro areas, the techniques and databases explored in this project could be highly valuable. The process could be even more valuable for smaller urban areas that have yet to develop transportation models. Such places could potentially develop much more accurate models from the beginning.

Even though the process explored in this project has proven valuable, it could be far more valuable with additional refinements. This is because the various errors and biases in the final database are, when combined, rather large. Significant reductions in the original databases would greatly improve their accuracy and usefulness for planning purposes. Even so, having what amounts to a large sample of data is still an improvement over the much smaller employment data samples or outdated census data that are used now.

7.2 Barriers to Statewide Expansion and Replication

Statewide expansion of the improved employment data project would be feasible given current database availability, processes, procedures, and technologies. A major barrier to statewide expansion of improved employment data would undoubtedly occur in the four Iowa metropolitan areas for which employment data cross state boundaries. These are Council Bluffs/Omaha (Nebraska), Sioux City (Nebraska, South Dakota), Dubuque (Wisconsin, Illinois), and the Quad Cities (Illinois). Although ES202 database systems are somewhat standardized, administrative regulations regarding their use are not. Negotiation would be necessary with each state and it is not at all clear that all would be cooperative, since the U.S. Bureau of Labor Statistics does not encourage sharing of the ES202 databases. From conversations with other states, it would appear that some might not choose to be as cooperative as Iowa.

7.3 Recommendations for the Iowa Department of Workforce Development

This project has indicated, perhaps as expected, that using the ES202 and related databases for a purpose for which they were not originally intended and designed is a difficult proposition. DWD could benefit from more rigid content controls on the name and address fields within its database system. This will be considerably easier to implement as DWD continues the current

migration to network database platforms from mainframe, flat file databases. The entire process would also be much easier if geocoding could be done at DWD as part of the original development of the ES202 and related databases rather than later on.

Two potential approaches to solving this dilemma are possible. One involves the development of additional processes along the lines already developed through this project to clean up and adjust the current database to make it less biased, more complete, and more usable by transportation planners.

The other approach, which would be preferred, is to start back at the beginning and work with the Iowa Department of Workforce Development to redesign and clean its ES202 and wage record databases so they can more easily be used for planning purposes. This would involve such steps as obtaining "clean" street addresses, using physical locations rather than mailing addresses, and identifying actual work locations of employees for firms with multiple locations. This work would be done with geocoding and GIS use in mind from the very beginning.

Ideally, this philosophy would also be extended to the Iowa DOT drivers license database. Working with the very large drivers license database proved to be a very costly part of the pilot study process. A major help in this regard would be the development of a personal computer compatible database of some of the most basic fields of the extremely unwieldy database that could again be used for planning purposes.

8 References

The following were contacts for the survey of other states' employment data practices:

- Mike Buso, U.S. Bureau of Labor Statistics
- Nancy Brennan, Wyoming Department of Employment, Employment Resources Division, Research and Planning Section
- Scott Festin, Wasatch Front Regional Council (Utah)
- Jim Hall, Department of Workforce Services, State of Utah
- Paul Hamilton, Tri-County Regional Planning Commission (Lansing, MI)
- Susan Hendricks, KJS Associates (Bellevue, WA)
- Alan Horowitz, Center for Urban Transportation Studies (University of Milwauke)
- Henry Jackson/Rita Lee, Illinois Department of Employment Security
- Janet Oakley, National Association of Metropolitan Planning Organizations
- Bob Parry, Utah Department of Transportation
- Alan Vander Wey, Montana Department of Transportation

Sources of information on various public and private employment data sources include the following:

- American Business Information, Inc., Omaha, Nebraska.
- Steve Boal, Iowa Department of Education
- Harris Publishing Company, Twinsburg, Ohio
- KJS Associates, Inc.
- Pro CD, Inc., Danvers, Massachusetts
- R.L. Polk & Company, Manhattan, Kansas
- David Miller, Iowa Emergency Management Division
- Jerry Musser, Johnson County, Iowa Assessor

- Kathy Harpole, Iowa Department of Revenue and Finance
- Terry Dillinger and Bruce Shuck, Iowa Department of Transportation
- US WEST Data Products Group, Denver, Colorado

APPENDIX A: Detailed Description of Database Development Process

This Appendix summarizes the processes undertaken to determine the feasibility of obtaining data from the Iowa Department of Workforce Development and the Iowa Department of Transportation to create a database with ability to provide data that may be geocoded to present a geographical display of work travel patterns for the Central Iowa pilot area.

A.1 Working with Magnetic Tapes
Reading the original magnetic tapes was done with Wylbur on the Iowa State University mainframe computing system. Because Wylbur is being phased out and will not be available after July 1, 1998, some of the details reported here will no longer be relevant. All programming will need to be converted to use Project Vincent resources in the future.

Three data sources were used for the initial information:

Quarterly Unemployment Insurance (QUI) File – Supplied by the Iowa Department of Workforce Development. This provided information on businesses including the unemployment insurance account number, SIC Code, FIPS Code, and addresses.

Wage Record Information -- Supplied by the Iowa Department of Workforce Development. This provided information on employee's Social Security Numbers, and unemployment insurance account number.

Drivers License Records – Supplied by the Iowa Department of Transportation. Drivers License Numbers (Generally but not always equal to social security numbers), home Addresses and other information was retrieved from this dataset.

The test area selected was RPA 11, which represents the following counties: Boone, Story, Dallas, Polk, Jasper, Madison, Warren and Marion.

The following is a list of data files created during the manipulation process:

1. Selected Firms in Central Iowa Counties
2. Uncompressed DWD Wage Files
3. Sorted File 2 by Unemployment Insurance Number
4. Matched Files 1 and 3
5. Sorted File 4 by Social Security Number
6. Matched Files 4 and 5
7. Unmatched records from File 4 during attempted match of File 5

Each of the following steps represents a unique program in Wylbur.

Step 1: Creates FILE 1. Select counties in the study area by FIPS location code from the QUI file. This procedure also selects the other data that will be used: FIPS code, SIC code, unemployment insurance number, and address information. Then, the selected records are sorted by unemployment insurance number.

Number of Selected Records: 19,243

Step 2: Creates FILE 2. The DWD wage record file is compressed, so it must be uncompressed to be useable.

Step 3: Creates FILE 3. Sort wage record file by unemployment insurance number. This is done to reduce CPU time on the mainframe during the matching process.

Step 4: Creates FILE 4. Match unemployment insurance numbers from the wage record file and the QUI file.

Number of Matched Records: 603,134

Step 5: Creates FILE 5. The matched records are sorted by Social Security Number.

Step 6: Creates FILE 6 and FILE 7. Match Social Security Numbers from the output of step # 5 to the driver license number from the driver license database. The resulting process creates two separate files, the records that matched a driver license number and the records that did not.

Number of Matched Records: approximately 550,365
Number of Unmatched Records: approximately 52,769

It might be inferred from this that about nine percent of all drivers in Iowa choose not to use their Social Security Number on their drivers license; however, some of the unmatched records are created by clerical and other errors.

If repeated again, the completion of these steps would only take 4 to 6 hours. If it is done under a system other than Wylbur, the programming would need to be done over and that would take about 10 to 20 hours for an expert programmer to complete it.

The No Match file contains the following information:
(Social Security Number is from the DWD Wage file and the remainder is from the DWD QUI file)

1-9	Social Security Number
10-16	Employment insurance number
16-20	SIC Code
20-23	FIPS Code
23-58	Firm Name
58-93	Firm Address
93-121	Firm City
121-123	Firm State
123-132	Firm Zip Code
132	Address Type 1-Physical 2-Mailing 3-Administrative HQ 9-Unknown

The Matched file contains the above fields in addition to the following:
(All of the additional information comes from the DOT Driver's License File)

133-153	Employee Last Name
153-168	Employee First Name
168-183	Employee Middle Name
183-186	Employee Suffix (Jr, Sr etc)
186-194	Employee Birthdate
195	Sex
195-219	Address
219-235	City
235-237	State
237-246	Zip
246-248	County
248-256	Process Date

A.2: Geocoding Firm and Employee Locations

Using a very unscientific sample of both firm and employee locations, an attempt at geocoding produced the following results:

(984 items sampled)

Firms:
Good Match (75-100)	573	58.2%
Partial Match (60-74)	66	6.7%
No Match (< 60)	345	35.1%

Employees:
Good Match (75-100)	609	61.9%
Partial Match (60-74)	36	3.6%
No Match (< 60)	339	34.5%

The numbers in parenthesis are based on a system used by Arc/Info to score the likelihood of a correct match. Sixty was set as the minimum match score, so any record that scored under 60 was considered a "no match". The Arc/Info manual considers any score 75 or above as a good match.

Both firm and employee data was taken from the same record. This causes some identical addresses, mainly with the firms. Removal of the "identicals" would not change the success rate.

The limiting factors in the success rate of the geocoding are the accuracy of both the street addresses in the state files and the addresses in the street database. The street database poses the larger problem, but some of the DWD and DOT addresses also contain only post office box numbers and rural route designations. The purchase or development of a more reliable (and more expensive) street address database could increase the successful match rate.

Using a Pentium Pro-200 with 32Mb of RAM and ArcView 3.0a, the geocoding process took approximately 1 hour for each 2,000 records. The output created must be divided into 500 records per group to avoid taxing the system. To geocode the entire matched set of 550,000 records for RPA 11 would take about 275 hours and create 1100 separate output files. As the power of the computer used is increased, the time required and number of output files required both decrease.

APPENDIX B   Statistical and Spatial Biases Introduced During the Project Database Conversion/Management Process

The database used in this study was created through a multi-step process of generating, linking, and manipulating several files. At each step in the process, a number of employee and employer records were excluded and thus did not appear in subsequent steps. The resulting database, therefore, is a detailed but potentially biased sample. Specifically, the process used to create the final database may have introduced biases related to location, industry classification, and other factors. This has serious implications for transportation and employment applications that rely in this database. This section outlines the potential biases introduced by the database creation process and assesses their importance.

The first potential bias relates to the data universe of the original unemployment insurance files used in this study. They exclude several categories of workers, such as agricultural workers, sole proprietors, and railroad pension system participants. About 7 percent of Iowa's workers are classified as agricultural and thus are not in the files. Table B1 provides estimates, from the 1990 Census, of the importance of agricultural employment in the 8-county Des Moines region and for the state as a whole. (The Census data used combines farming, forestry, and fishing employment; the latter two are assumed to be negligible.) Figure B1 shows the statewide pattern. Not surprisingly, such workers are likely to be located outside urban areas. The average county has 12.4 percent of its workers in agriculture, higher than all 8 counties in the Des Moines region. In contrast, Polk County, has the lowest percentage in the state.

Similarly, Table B2 and Figure B2 provide estimates of the self-employed workforce, here used as a rough estimate of the presence of farmers as well as sole proprietors. Again, the Des Moines area has much smaller values than the rest of the state, and Polk County has the lowest value in Iowa.

Table B1: Percent of workers in farming, forestry, and fishing occupations, 1990

County	Percent
Boone	6.8%
Dallas	6.8%
Jasper	7.2%
Madison	10.1%
Marion	5.5%
Polk	1.0%
Story	3.9%
Warren	4.6%
Total – Region 11	2.9%
Total – State of Iowa	7.1%

Source: 1990 Census

Percent of workers in farming, forestry, and fishing occupations
- ☐ 0 - 0.04
- ☐ 0.04 - 0.08
- ☐ 0.08 - 0.12
- ☐ 0.12 - 0.16
- ■ 0.16 - 0.23
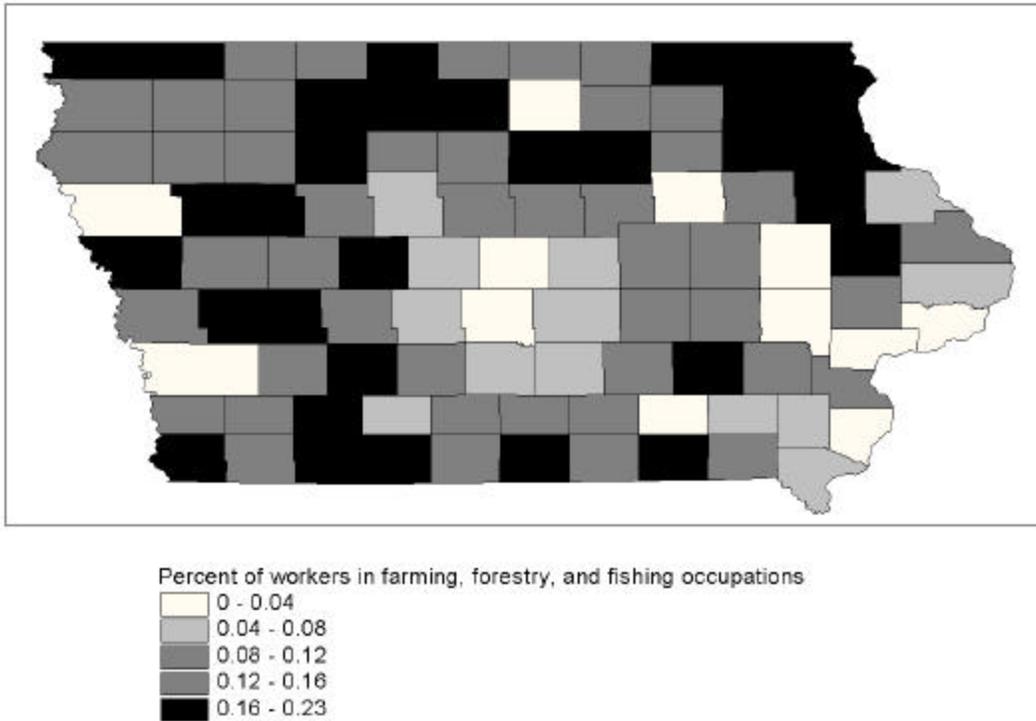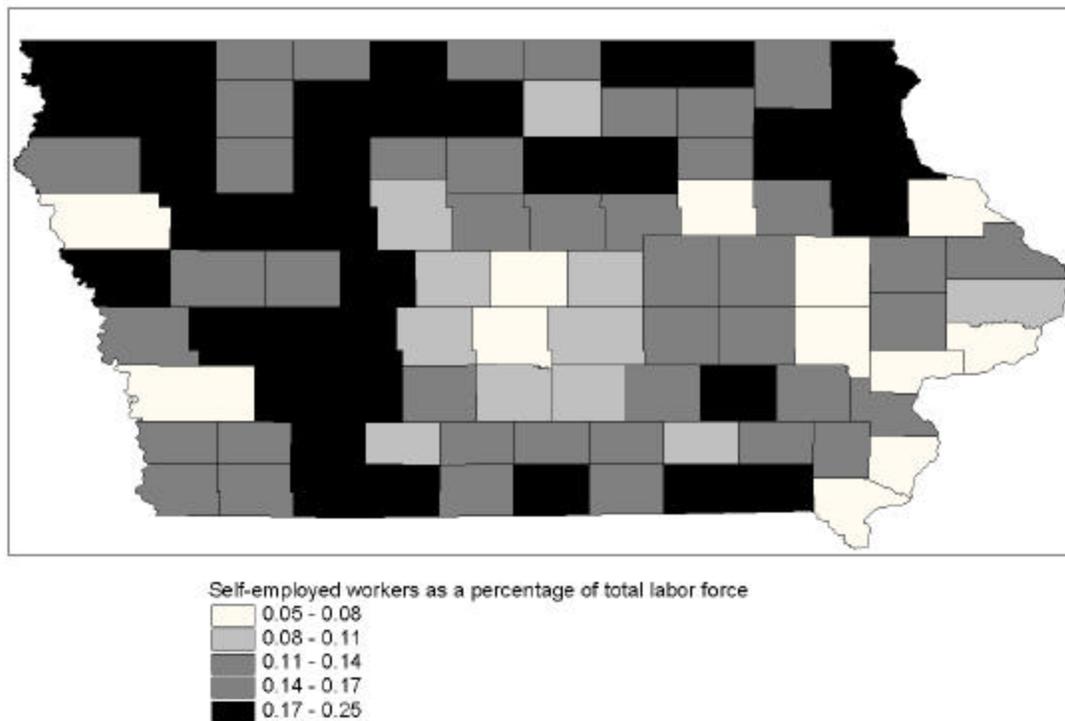
Figure B1: Agricultural Employment in Iowa

Table B2: Percent of the labor force that is self-employed, 1990

County	Percent
Boone	8.8%
Dallas	10.1%
Jasper	9.5%
Madison	13.3%
Marion	9.2%
Polk	5.1%
Story	6.6%
Warren	8.6%
Total – Region 11	6.5%
Total – State of Iowa	10.4%

Source: 1990 Census

Self-employed workers as a percentage of total labor force

	0.05 - 0.08
	0.08 - 0.11
	0.11 - 0.14
	0.14 - 0.17
	0.17 - 0.25

Figure B2: Self-employment in Iowa

Excluding much of the agricultural workforce obviously means that rural areas will tend to be underrepresented in any derived database. Excluding other self-employed individuals and other portions of the workforce may similarly introduce biases. The implications of this spatial bias depend on the application of the data and the scope of the analysis performed. Statewide or regional analyses may be skewed in favor of urban areas. On the other hand, transportation modeling in metropolitan areas will be less affected. Moreover, the excluded workforce is likely to work at home, either on the farm or in a home office, and thus is not likely to greatly skew analyses of travel demand patterns.

In the next step in the database creation process, the employer and employee files maintained by the Department of Workforce Development are joined using the firm unemployment insurance number as the common identifier. This is a relatively straightforward match and we can safely assume that no bias is introduced in this step. Unfortunately, however, the actual place of employment cannot be determined from this process because a firm may have a single unemployment insurance number but several locations. In this case, an employee record would simply match to the first instance of the correct employer, regardless of location. This creates difficulties, as discussed below.

The second potential source of bias is the joining the DWD joint employer/employee files with the DOT's drivers license file by employee social security number. This was done to include the employee address information from the DOT file. Of the 597,541 records in the employee/employer file, 52,009, or 8.7%, did not match with a record in the drivers license file. There are three possible explanations for a non-match. First, the employee may not have an Iowa drivers license. Persons likely to be in this category include those living in other states, newcomers, and young people. Second, individuals now have the choice not to have one's social security number be used as an identification number on one's drivers license. Third, there are data entry errors in the employment file or the drivers license file. Employment

file errors are likely to be detected and fixed, because they affect one's wage and benefit records. Drivers license file errors may be detected, but neither the DOT nor the driver has an urgent need to correct this item.

It is difficult to estimate the spatial bias introduced by this stage in the process because the employee's workplace location is not necessarily the correct one. However, an assessment of bias can be made by SIC code. As shown in Table B3, there is not a great deal of variation in match rate according to industry type. (The retail vs. non-retail split is particularly important for transportation analysis.)

Table B3: Match rate of employment file to drivers license file, by SIC (employees)

Industry Type	SIC	Matched	Total	Match Rate
Retail trade	5200-5900	145,804	161,205	90.4%
Services – FIRE, health, education, etc.	6000-9000	211,756	231,860	91.3%
Other 1 – Agriculture, Mining, Construction, Manufacturing, Transportation, Wholesale	100-5100	161,513	176,573	91.5%
Other 2 – Public Administration, and non-classifiable	9100-9900	26,459	27,903	94.8%
TOTAL		545,532	597,541	91.3%

The match rate can also be assessed according to the first three digits of the employee social security number, which roughly correlates to a person's location at the time one received a social security card. The prefixes of 478 to 485 are those identified by the Social Security Administration as generally originating in Iowa. Table B4 shows that the match rate is much higher for employees with Iowa social security number prefixes. Iowa locations with many individuals having non-Iowa prefixes, likely to be border regions (due to cross-border commuting) and metropolitan areas (due to migration), may be underrepresented in the matched sample.

Table B4: Match rate of employment file to drivers license file, by employee social security number prefix

Prefix Category	Matched	Total	Match Rate
Iowa (478-485 prefixes)	445,433	466,742	95.4%
Outside Iowa (all other prefixes)	100,099	130,799	76.5%
Total	545,532	597,541	91.3%

The third source of bias occurs during the address matching process to identify the x,y coordinates of each record. The records in the joint employment/drivers license file were address matched according to the home address of the employee. A separate file containing the firm addresses was also address matched.

Several factors affect the address matching hit rate (i.e., the percentage of records for which an x,y coordinate is successfully identified). The street database can have missing or incorrect streets and street names. In addition, records in the attribute database (here, the joint employment/drivers license file and the firms file) can have incorrect or non-standardized addresses. In particular, the use of P.O. boxes and rural routes, addresses for which no actual street name is identified, is perhaps the most important cause of poor hit rates. In the Des Moines region, the overall hit rate was 69.9% for employee residences and 59.8% for firms. Not surprisingly, the hit rate is much worse in rural areas than urban areas. Table B5 illustrates this at the county level for the Des Moines region.

Hit rates were also calculated at the city level, for both firm locations and employee residence locations. In general, the larger the city, the better the hit rate, as shown in Table B6.

Table B5: Address matching hit rates of employee residence and firm location, by county

County	Employee Residence Hit Rate	Firm Location Hit Rate
Boone	54.4%	44.3%
Dallas	37.4%	28.9%
Jasper	47.7%	48.2%
Madison	25.9%	10.4%
Marion	42.1%	44.1%
Polk	79.7%	67.4%
Story	62.0%	53.8%
Warren	61.2%	49.7%
Total – Des Moines Area	69.9%	59.8%
Total – State of Iowa	65.3%	-

Table B6: Address matching hit rates of firm location and employee residence, by city (representative sample)

City	Employee Residence Hit Rate	Firm Location Hit Rate
Des Moines	87.4%	78.2%
West Des Moines	66.6%	56.0%
Ames	70.8%	63.8%
Ankeny	77.0%	76.1%
Newton	68.1%	67.2%
Urbandale	80.6%	65.5%
Indianola	58.8%	57.9%
Boone	64.6%	51.0%
Pella	63.9%	65.7%
Winterset	38.5%	13.0%
Nevada	63.2%	52.2%
Altoona	76.7%	78.2%
Adel	38.1%	45.1%
Madrid	28.7%	27.5%
Huxley	50.9%	20.6%
Polk City	49.4%	19.3%

At finer level of analysis, Figure B3 shows the hit rate by zip code employee residences in the state of Iowa. (This includes employees who work for firms with at least one location in Region 11.) Again, urban areas tend to have much better hit rates. Similarly, Figure B4 shows the hit rate for firm locations in the Des Moines area by zip code, and the same pattern appears.
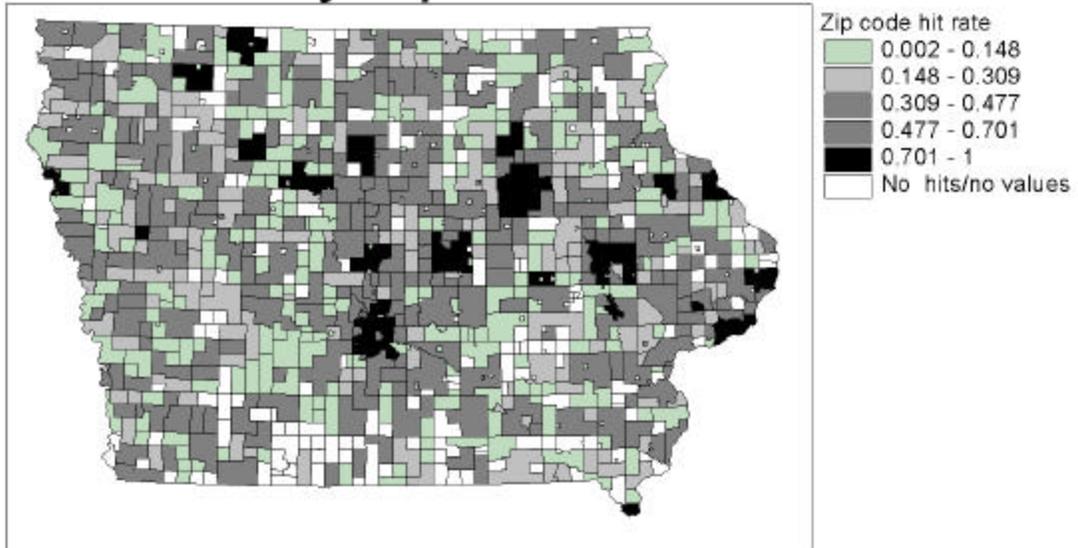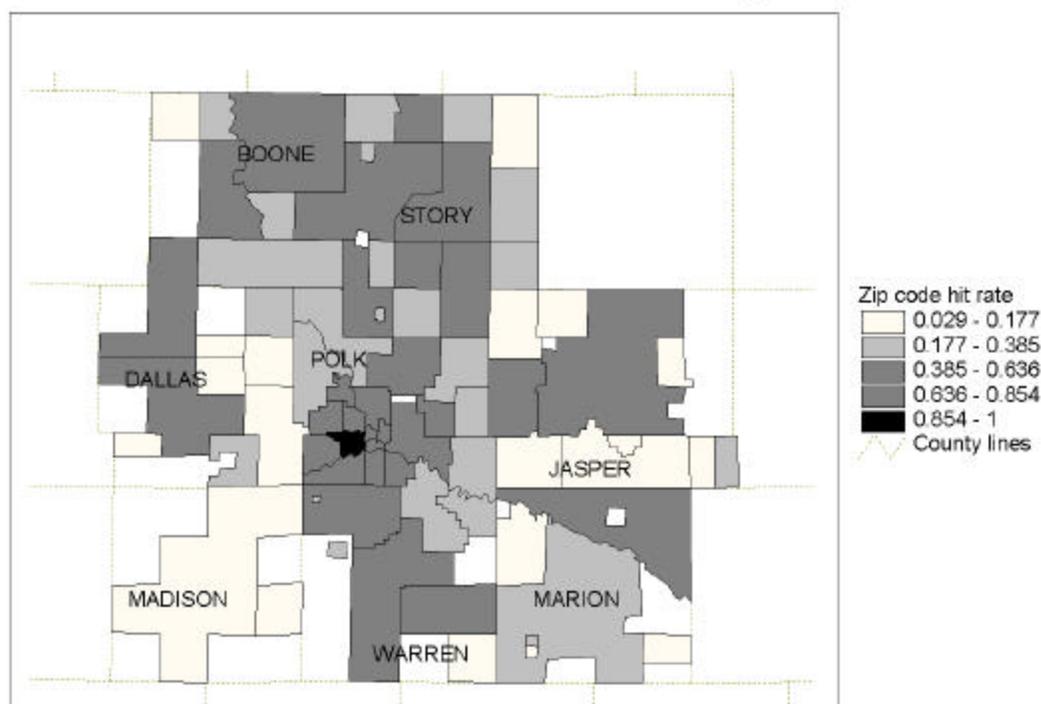
Figure B3: Residence Location Hit Rates by Zip Code, State of Iowa

Firm Location Hit Rates by Zip Code

Figure B4: Firm Location Hit Rates by Zip Code, Des Moines Region

Table B7 shows the firm location address matching hit rates by industry type. As noted above, variation in match or hit rates by industry can have serious implications for transportation modeling. As Table B7 illustrates, the address matching hit rates are consistent for retail vs. service employment, while the "other" categories, especially public administration, have lower hit rates.

Table B7: Address matching firm location hit rates, by SIC

Industry Type	SIC	Matched	Total	Hit Rate
Retail trade	5200-5900	2,520	3,919	64.3%
Services – FIRE, health, education, etc.	6000-9000	5,555	8,654	64.2%
Other 1 – Agriculture, mining, construction, manufacturing, transportation, wholesale	100-5100	3,172	6,138	51.7%
Other 2 – Public administration, non-classifiable	9100-9900	199	432	46.1%
TOTAL	-	11,446	19,143	59.8%

Although the origin state of the employee's social security number greatly influenced the match of the employment file to the drivers license file, it has little impact, not surprisingly, on the address matching hit rate (see Table B8).

Table B8: Address matching hit rate by social security number of employee

SSN Prefix Origin	Hit Rate
Iowa SSNs (478-485 prefix)	65.9%
Non-Iowa SSNs (all other prefixes)	63.1%
Total	65.3%

The address matching hit rate even varied by year of birth: individuals born in the year 1950 or before had a hit rate of 67.2%, while those born between 1951 and 1965 had a hit rate of 63.1%, and those born 1966 or after had a hit rate of 66.3%. This result is likely related to the others outlined above.

The fourth source of potential bias was created during the preparation of an origin-destination table for the Des Moines area. This was done to determine the usefulness of the database for transportation modeling applications (see the main body of this report). First, any firm whose address in the database was not the physical location of employment (e.g., the address was an administrative headquarters) was taken out of the same. Figure B5 shows the spatial pattern of these firms. They tend to be located in downtown Des Moines and to the west. Second, the process excluded firms having more than one location in the Des Moines area. This was done to delete records for which an employee was potentially linked to the incorrect location of a multi-location firm. The spatial pattern of these excluded firms is shown in Figure B6. Finally, the process selected only records for which an individual both worked and lived within the boundaries of the Des Moines MPO transportation analysis zones. This was done to construct an internal-internal trip table.

Percent of firms not using physical address
- 0 - 0.05
- 0.05 - 0.1
- 0.1 - 0.3
- 0.3 - 0.7
- 0.7 - 1
- Zero or one observation

Figure B5: Percent of Firms Not Using Physical Location as the Firm Address in the Employment File

Figure B6: Percent of Firms Having More Than One Location in the Des Moines Region, By TAZ

Percent multi-location firms, by TAZ
0 - 0.1
0.1 - 0.25
0.25 - 0.4
0.04 - 0.6
0.6 - 1

Finally, we can estimate the "total" hit rate of the process to create the trip table for the transportation model. This summarizes the total impact of biases introduced by address matching, eliminating non-physical addresses, eliminating multi-location firms, and keeping only those records for which an individual lives and works within the area covered by the TAZs. Figures B7 and B8 display the total hit rates for employee residences and firm locations, respectively. In both case, areas near the central part of the region tend to be over-represented. These hit rates were used in the "variable percent" method noted in the main body of this report. It was the most successful method in enhancing the data used in the current model.
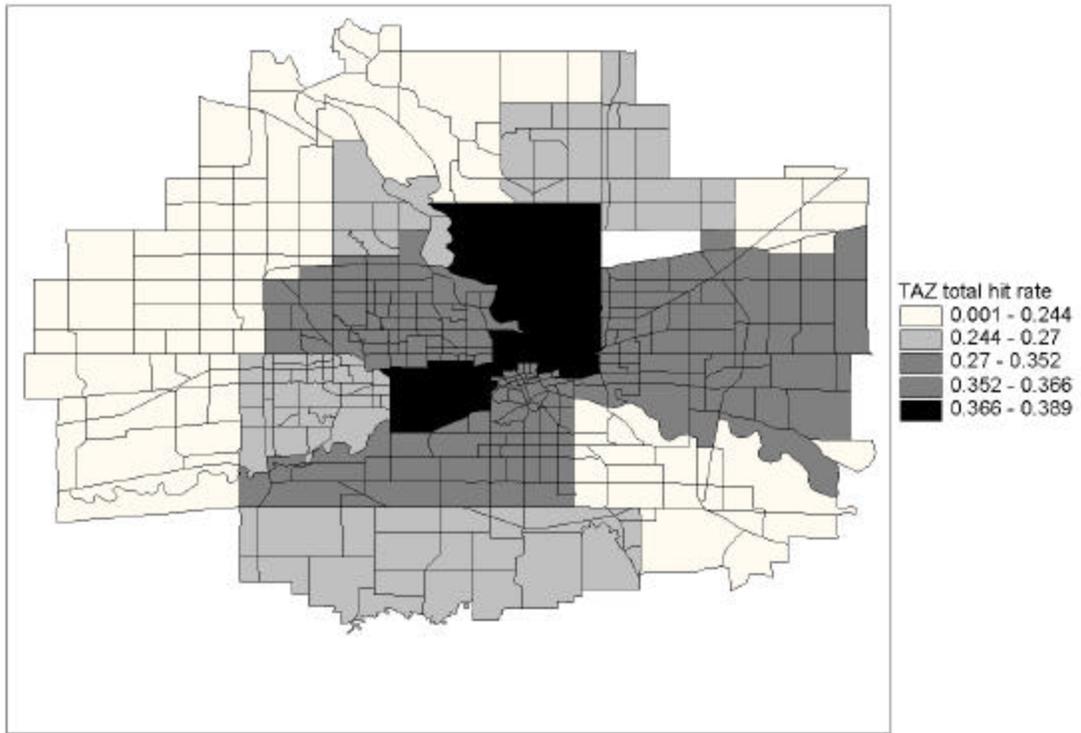
Figure B7: Total Employee Residence Hit rate by Des Moines Model Transportation Analysis Zone
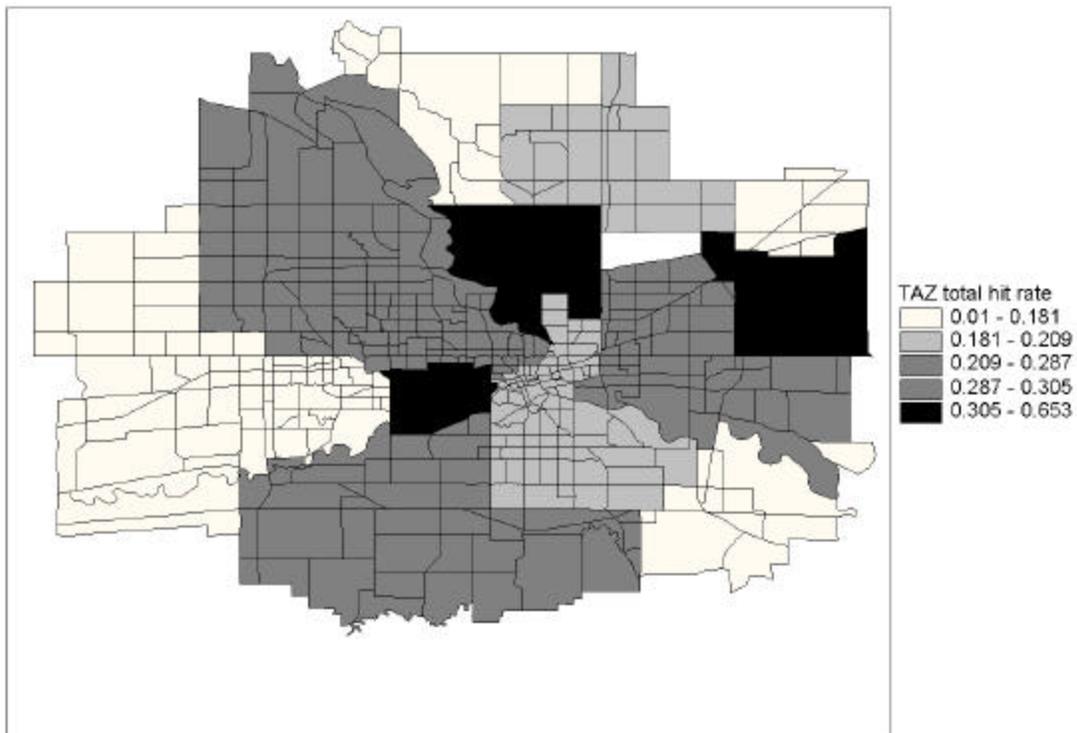
Figure B8: Total Firm Location Hit rate by Des Moines Model Transportation Analysis Zone

**APPENDIX C: Using IDWD Employer and Employee Files
to Improve Travel Demand Modeling in Des Moines**

The following is a methodology for trip distribution calibration (part of the conventional traffic forecasting methodology) using employment data for a case study of the Des Moines MPO 1990 base model (can be generalized for other areas). Three techniques are applied to produce either synthesized (methods 1 and 2) or "improved" (method 3) home based work (HBW) trip tables. All three begin with a sample OD (origin-destination) database, derived from ES202 and drivers license data. The sample represents a spatially biased subset of home based work trips in Des Moines. The sample has type I errors (many trips that take place are not sampled), but is not likely to contain type II errors (representing trips that in reality do not take place).

The first technique synthesizes a new HBW table by applying a fixed percentage to factor up the sample OD trip table to a control total. The control total is the total number of trips in the original 1990 base HBW, Des Moines area trip table, produced by a gravity model (hereafter known as original HBW table). Although the sample provides trip pairs which can be used to calibrate the results of the gravity model, we have more confidence in the original HBW productions (Ps) and attractions (As), than the factored-up Ps and As provided by the sample. Therefore, a growth factor matrix manipulation technique (FRATAR method) is applied to constrain row and column totals (Ps and As) in the synthesized OD table to the Ps and As of the original HBW table.

The second technique synthesizes a new HBW table by applying a variable percentage to factor up the sample OD table to the control total. Variable percentages are determined by spatial "hit" (sampling) rate for each zip code (all zones within a given zip code are assumed to have the same hit rate). Zip code "hit" rates are determined as, the number of employee residences or jobs in the sample file (row and column totals) divided by the total number of trips that "should" have originated or terminated in each zip code (we have a complete, non-address matched, listing of jobs and residences in each zip code). Sample trips do not include many trips (those that were not geocoded properly or that represent trips to or from employers with multiple locations within the model area - see previous section on model sample bias.) As in the first technique, row and column totals are then constrained to the original number of Ps and As by the Fratar method.

The third technique produces an "improved" HBW trip table by taking advantage of the knowledge that for some OD pairs, the sample database is known to be an improvement over the trip table generated by the gravity model. Specifically, this is the case where the number of trips in the sample database cell is greater than the corresponding cell in the gravity model trip table. The technique retains the number of trips shown in the sample database where the sample is greater than the gravity model table. It then employs downward, iterative adjustment of the remainder of the trip table, constraining the adjusted trip table to the original number of Ps and As (also by Fratar method).

Given:

- Sample file, TAZMATCH, of employee TAZ and firm TAZ - uses 3rd quarter 1997 data and under-represents businesses with multiple locations within the DSM area and firms and employees with difficult to match addresses; ignores trip chaining and ridesharing
- Original non-directional (asymmetrical) binary HBW trip table, GM90.TRP, from 1990 DSM Tranplan model (643 Zone version), derived from estimates of zone socioeconomic data, (estimates?) of employment at firms in 1990 and gravity model assumptions (trip length distributions and travel time estimates)
- Tranplan Software, version 8 (methodology may also be compatible with later versions)

Find:

- loaded volumes using the synthesized or "improved" HBW trip tables, and compare to ground counts (validation)

- Compare this validation to 1990 MPO model validation (based on comparison of loaded volumes)
- Prepare statistics, plots and maps to show the improvement (if any) in modeling due to manipulation of DWD and other employment data

1) Synthesize new, or improve Home Based Work (HBW) OD tables

1.1) Synthesize O/D table, method #1, expand by a fixed % for all trip pairs, assuming no spatial bias (create and work in a directory called TRANPLAN\FIXED%). Constrain row and column totals to be equal to original (base) row and column totals (Ps and As). Steps are as follows

1.1.1) Determine the number of HBW trips in original, 1990 (643 zone version) Des Moines model ($tsum_{90}$).[1] Files are located in a directory called TRANPLAN\BASE1990. To compute the number of trips, a Tranplan control file (REPORT.IN) is written to report a summary of the binary HBW trip table:

```
$REPORT MATRIX
$FILES
  INPUT FILE = RTABIN, USERID = $GM90.TRP$
$OPTIONS
  PRINT TRIP ENDS
$END TP FUNCTION
```

This function reports that there are 230,391 home based work trip productions in the model used by the MPO. This is to be used as a control total for adjusting up the sample data.

1.1.2) Determine total number of "trips" in sample file TAZMATCH.DBF by multiplying the number of rows in the database by two (assumes one trip in the AM from employee's TAZ to firm's TAZ, then one PM trip in the reverse direction), $tsum_S$

answer: 58,424 x 2 = 116,848

1.1.3) calculate ratio of number of trips in HBW file to number of "trips" in sample, $R_S = tsum_{90}/tsum_S$

230,391/116,848 = 1.9717

1.1.4) obtain and convert the sample file, TAZMATCH.DBF, into proper format for input to Tranplan Module BUILD TRIP TABLE (see MATRIX UTILITIES PAGE 6-3). Call the file S.TXT

---

[1] Note on original development of HBW trip table: The 1990 model used has 536 internal zones (only first 526 have data) and 643 including externals. MPO1990.XLS, a spreadsheet provided by the Des Moines MPO staff in June, 1998, reports 39,574 Retail and 178,882 Other Employment (218,456 total employment) using these figures, total home-based work trip attractions = 0.83 x 218,456 = 181,318. However, these "attractions" are to be adjusted to be equal to the number of trip productions, produced by cross-classification. This cross-classification is accomplished via a program written by WSA called SOLVEP.EXE. The output file from SOLVEP.EXE, called SOLVB90P.OUT, reports 132,952 DUs and 335,780 population in the internal zone area (560 zones - above 527 disaggregated to accommodate future growth.) Running SOLVEP.EXE produces 231,794 trip productions. (These figures are for the revise 668 zone model, which differ slightly from the 643 zone version used in this project).

Procedure: Using ArcView, make a copy of TAZMATCH.DBF as S.DBF into the TRNAPLAN\FIXED% directory and remove all fields except two (employee TAZ and Employer TAZ). Add two columns of ones (for Tranplan) and export into ASCII space-delimited format as S-INPUT.TXT. Then, to get into the format for Tranplan's BUILD TRIP TABLE module (which requires right justified data in a specific columnar format), obtain a copy of the FORTRAN program S.EXE and run on S-INPUT.TXT to produce S.TXT. The code (S.FOR) written for this project is:

```
        OPEN (1, FILE = 's-input.txt', STATUS = 'old')
        OPEN (2, FILE = 's.txt', STATUS = 'unknown')

        do 50,i=1,1000000
        read (1,*,err=200) nfrom,nto,n1,n2
        write (2,100)nfrom,nto,n1,n2,n2
100     format(i5,i5,i3,i3,34X,i5)
50      continue
200     continue
        end
```

Note: The program writes out 5 columns of information: from TAZ, to TAZ, a "1" indicating the "from" trip purpose, a "1" indicating the "to" trip purpose, and a "1" indicating the weight for the data (each line represents one a.m. trip, so the weight is "1").

1.1.5) Import sample file, S.txt, into Tranplan binary format using the BUILD TRIP TABLE Tranplan Module, which creates S.BIN. This table represents a.m. trips if each is a drive-alone trip.

The Tranplan control file written to perform this function is called BUILD.IN and contains the following:

```
$BUILD TRIP TABLE
$FILES
  INPUT FILE = SRVDATA, USERID = $S.TXT$
  OUTPUT FILE = VOLUME, USERID = $S.BIN$
$OPTIONS
  PRINT TRIP ENDS
$PARAMETERS
  NUMBER OF ZONES = 643
$DATA
  TABLE 1 = ALL
$END TP FUNCTION
```

1.1.6) Using the Tranplan module MATRIX TRANSPOSE, create ST.BIN (represents p.m. trips)

The Tranplan control file written to perform this function is called TRANSP.IN and contains the following:

```
$MATRIX TRANSPOSE
$FILES
  INPUT FILE = TRNSPIN, USERID = $S.BIN$
  OUTPUT FILE = TRNSPOT, USERID = $ST.BIN$
$PARAMETERS
  SELECTED TABLES = 1
$END TP FUNCTION
```

1.1.7) Using the Tranplan module MATRIX MANIPULATE, add S.BIN to ST.BIN to create S-OD.BIN (represents all work trips in the sample).

The Tranplan control file written to perform this function is called MANIP.IN and contains the following:

```
$MATRIX MANIPULATE
$FILES
  INPUT FILE = TMAN1, USERID = $S.BIN$
  INPUT FILE = TMAN2, USERID = $ST.BIN$
  OUTPUT FILE = TMAN3, USERID = $S-OD.BIN$
$DATA
  TMAN3,T1 = TMAN1,T1 + TMAN2,T1
$END TP FUNCTION
```

1.1.8) Using the Tranplan module MATRIX UPDATE, multiply S-OD.BIN by the ratio R_S to create OD-CAL.BIN (factors up the sample to the control total.)

The Tranplan control file written to perform this function is called UPDATE.IN and contains the following:

```
$MATRIX UPDATE
$FILES
  INPUT FILE = UPDIN, USERID = $S-OD.BIN$
  OUTPUT FILE = UPDOUT, USERID = $OD-CAL.BIN$
$DATA
  T1,1-643,1-643,* 1.971715
$END TP FUNCTION
```

1.1.9) Verify proper procedure (check against control total):

If proper procedure was followed, the matrix OD-CAL.BIN should have the same number of total trips as the original HBW trip table (230,391). To verify, run the Tranplan module REPORT MATRIX. The control file, called REPORT1.IN is as follows:

```
$REPORT MATRIX
$FILES
  INPUT FILE = RTABIN, USERID = $OD-CAL.BIN$
$OPTIONS
  PRINT TRIP ENDS
$END TP FUNCTION
```

The results of this process indicate a total number of productions of 230,400. Differences are due to rounding off the ratio.

1.1.10) Perform Fratar analysis to factor the synthesized trip table row and column totals to be equal to original HBW row and column totals (Ps and As):

> a) Determine origin and destination totals for synthesized trip table OD-CAL.BIN. Get these totals form the output of Tranplan's REPORT MATRIX module (step 6 above). Place in Lotus spreadsheet OD-CAL.123.
>
>> (note: some manipulation of the output text file is needed to get the data into a manageable format in the spreadsheet. It is

helpful to clean up the Tranplan output file with a text editor before importing into the spreadsheet as a text file. It is also useful to know something about the Parse function of the spreadsheet. Alternatively, commas may be added to the Tranplan output to facilitate import to the spreadsheet).

b) Switch to the TRANPLAN\BASE1990 directory. Obtain directional (symmetrical) origins and destinations totals from the original HBW table. The non-directional trip matrix must first be adjusted to reflect 24-hr directionality (must be symmetrical). Get these totals from the output of Tranplan running the following control file (PRE-FRAT.IN):

```
$MATRIX TRANSPOSE
$FILES
  INPUT FILE = TRNSPIN, USER ID =$GM90.TRP$
  OUTPUT FILE = TRNSPOT, USER ID =$GM90T.TRP$
$PARAMETERS
  SELECTED TABLES = 1
$END TP FUNCTION

$MATRIX MANIPULATE
$FILES
  INPUT FILE = TMAN1, USER ID =$GM90.TRP$
  INPUT FILE = TMAN2, USER ID =$GM90T.TRP$
  OUTPUT FILE = TMAN3, USER ID =$2GM90.TRP$
$DATA
  TMAN3,T1 = TMAN1,T1 + TMAN2,T1
$END TP FUNCTION

$MATRIX UPDATE
$FILES
  INPUT FILE = UPDIN, USER ID =$2GM90.TRP$
  OUTPUT FILE = UPDOUT, USER ID =$GM90X.TRP$
$DATA
  T1,1-643,1-643, * .5
$END TP FUNCTION

$REPORT MATRIX
$FILES
  INPUT FILE = RTABIN, USERID = $GM90X.TRP$
$OPTIONS
  PRINT TRIP ENDS
$END TP FUNCTION
```

Check to see that the total number of trips is still 230,400. It is 230,514 … close enough.

c) Copy the zone origins and destinations from the TRNPLN.OUT Tranplan output file produced when REPORT MATRIX was run.
Switch back to the TRANPLAN\FIXED% directory. Place origin and destination totals into Lotus spreadsheet OD-CAL.123.

> (note: some manipulation of the output text file is needed to get the data into a manageable format in the spreadsheet. It is helpful to clean up the Tranplan output file with a text editor before importing into the spreadsheet as a text file. It is also

useful to know something about the Parse function of the spreadsheet. Alternatively, commas may be added to the Tranplan output to facilitate import to the spreadsheet).

Compute growth factors for each zone = values from GM90X.TRP file divided by values from OD-CAL.BIN file. See sample from the table (C1)and notes below:

Table C1

| | FIXED % Adjusted Sample | | | | Original HBW | | | | Growth Factor F |
Zone	Orig.	Dest.	Diff.	Avg	Orig.	Dest.	Diff.	Avg	*
1	319	324	-5	321.5	277	371	-94	324	1.0078
2	434	439	-5	436.5	314	312	2	313	0.7171
3	150	152	-2	151	206	219	-13	212.5	1.4073
4	641	649	-8	645	1213	1231	-18	1222	1.8946
5	511	516	-5	513.5	592	581	11	586.5	1.1422
6	781	787	-6	784	542	557	-15	549.5	0.7009
7	69	68	1	68.5	134	132	2	133	1.9416
8	580	582	-2	581	396	393	3	394.5	0.6790
9	183	185	-2	184	213	215	-2	214	1.1630
10	312	310	2	311	380	389	-9	384.5	1.2363
11	246	247	-1	246.5	69	58	11	63.5	0.2576
12	532	530	2	531	545	552	-7	548.5	1.0330
13	816	822	-6	819	1198	1192	6	1195	1.4591
14	564	567	-3	565.5	1768	1766	2	1767	3.1247
15	946	948	-2	947	866	875	-9	870.5	0.9192

note: although origin and destination row and column totals should be equal for all zones (a 24 hour model implies that all vehicles that leave a zone will return to that zone during the day sometime), differences are due to accumulation of round-off errors in summing individual trip pair values (e.g., when 3 is multiplied by 0.5 in matrix update, sometimes Tranplan rounds up to 2, sometimes down to 1); therefore, an average of O and D is taken prior to computation of Fratar growth factors. Hence, only origin growth factors (hereafter known simply as growth factor) are needed by Tranplan.

* when sample is zero, factor is taken to be zero. This causes only a minor difference in total trips (~5000 out of ~230,000 work trips)

d) Create a control file for the Tranplan Fratar Model Module, (FRATARF.IN). Incorporate TAZ number and growth factors from OD-CALF.123, following format specified on MODELS -- PAGE 4-4 of the Tranplan manual. (this is in the Fratar Model section of the manual).

Control file looks like:

$FRATAR MODEL
$FILES

```
            INPUT FILE = FRATIN, USER ID = $OD-CAL.BIN$
            OUTPUT FILE = FRATOUT, USER ID = $OD-CAL-F.BIN$
$DATA
FO   1 1   10078
FO   2 1   07171
FO   3 1   14073
FO   4 1   18946
.
.
$END TP FUNCTION
```

      e) run the Tranplan Fratar Model, producing a final, fixed% sythesized trip table OD-CAL-F.bin

1.1.11) again, verify proper procedure using control totals

If proper procedure was followed, the final matrix OD-CAL-F.BIN should have the same number of total trips as the original HBW trip table (230,391). To verify, we ran the Tranplan module REPORT MATRIX. The control file, called REPORT2.IN is as follows:

```
$REPORT MATRIX
$FILES
  INPUT FILE = RTABIN, USERID = $OD-CAL-F.BIN$
$OPTIONS
  PRINT TRIP ENDS
$END TP FUNCTION
```

The results of this process indicate a total number of productions of 225,196. Differences are due primarily to elimination of cells with zero sample volumes.

1.2) Synthesize OD table via method #2, expand trip pairs by a variable %, assuming spatial bias by zip code (introduced by geocode problem, multiple employer, or physical address problem) and no other bias (create and work in a directory called TRANPLAN\VARIABLE). Constrain row and column totals to equal the Ps and As of the original HBW trip table.

      1.2.1) Determine representation rate by zip code, for employees living in each TAZ,

            1.2.1a) start with file ALL.PAT (this is the file of 545,000 employees that work at firms with at least one DSM area location. This file represents about 90% of all employees covered by unemployment insurance - the ones that have a valid SSN on their driver's license. Note: open this file with ArcView, (add theme) but do not turn the layer on (view). Simply open the .dbf file in a window). Assume no spatial bias is introduced by driver's license SSN omission.

            1.2.1b) Determine the total number of employees in each Zip Code:

            Since the file has 9 digit zipcodes and the file is large, with the EZIP field selected, use the ArcView command SUMMARIZE to create a temporary table (SUM1.DBF containing two fields, EZIP and COUNT). Then in the temporary table, create a new field of 5 digit zip codes (EZIP5, character with 5 places). Calculate the 5 digit zip codes (truncate the 9 digit ones, EZIP5 = EZIP.left(5)), kill the EZIP field and with the EZIP5 field selected, use the ArcView command SUMMARIZE. Then, in the SUMMARIZE dialog box, tell ArcView to include a field to sum the values in the COUNT (summary) field

(SUM_COUNT). This produces another temporary table of the number of employees in each of the 5 digit zip codes. The table is saved as SUM2.DBF and has EZIP5, COUNT and SUM_COUNT fields. In SUM2.DBF, check that the frequencies by zip code sum to the original total number of records in ALL.PAT (checked). Kill the field COUNT and rename the field SUM_COUNT to CT_ALL.

1.2.1c) Determine the number of employees that reside in DSM area zip codes in the sample file:

TAZMATCH.DBF (this is the sample file of 58,424 "single-location-employer" employee/employer addresses that were geocoded within the model area), with the EZIP field selected, use the ArcView command SUMMARIZE to create a temporary table (SUM3.DBF containing two fields, EZIP and COUNT). Then in the temporary table, create a new field of 5 digit zip codes (EZIP5, character with 5 places). Calculate the 5 digit zip codes (truncate the 9 digit ones, EZIP5 = EZIP.left(5)), kill the EZIP field and with the EZIP5 field selected, use the ArcView command SUMMARIZE. Then, in the SUMMARIZE dialog box, tell ArcView to include a field to sum the values in the COUNT (summary) field (SUM_COUNT). This produces another temporary table of the number of employees in each of the 5 digit zip codes. The table is saved as SUM4.DBF and has EZIP5, COUNT and SUM_COUNT fields. In SUM4.DBF, check that the frequencies by zip code sum to the original total number of records in TAZMATCH.DBF (checked). Kill the field COUNT and rename the field SUM_COUNT to CT_MATCH.

1.2.1d) Join the table resulting from part b) (SUM2.DBF) to the table from part c) (SUM4.DBF). Add a new field (HZR) to this combined table, and calculate it as the ratio of sample counts (CT_MATCH) to all counts (CT_ALL) for each zip code. Save the resulting join a SUM5.DBF.

1.2.1e) From DNR web site, obtain a zip-code polygon coverage for DSM area (zip.shp)

1.2.1f) Obtain a point coverage of TAZ centroids called TAZ.SHP (the point coverage facilitates overlaying centroid points on zip code polygons to determine zip code for each TAZ (Note: since we are only dealing with TAZ centroids, some TAZs may actually be split amongst Zipcodes. We assume spatial autocorrelation in hit rates by adjacent Zipcodes and assume that a TAZ hit rate would not be significantly affected by this issue.)

1.2.1g) perform a spatial join of ZIP.SHP to TAZ.SHP to determine which Zipcode each TAZ (centroid) falls in. Save the combination as ZIPTAZ.DBF. This table has 526 records (one for each internal TAZ). Kill all fields except ZIP and TAZ to produce a correspondence table and save as ZIP2TAZ.TXT.

1.2.1h) Join the result of step d) (SUM5.DBF) which includes HZR, to ZIP2TAZ.TXT by zipcode, so that the hit rate, HZR is associated with each TAZ. Save that join as HITTAZ_R.DBF. Export the table as a text file to facilitate import to Tranplan (HITTAZ_R.TXT).

1.2.2) Determine representation rate by zip code, for employees that work in each TAZ,

1.2.2a) start with fixed length, ~ delimeted file DES_FIRM.TXT (this is the 4-county file with number of employees and zip code for work locations – note, may be biased as this includes all firms in 1994 from old DES data, also

includes all firms, even those with non-physical addresses -- possible magnitude of error is???). The file includes two fields, ZIP and N_EMP and 13,842 records (firms). There are 251,992 employees (N_EMPs). Import into ArcView and change the ZIP field from integer to character with 5 spaces.

1.2.2b) Determine the number of employees in each Zip Code:

With the ZIP field selected, use the ArcView command SUMMARIZE to create a temporary table, also in the SUMMARIZE dialog box, tell ArcView to include a field to sum the values of the N_EMP field (SUM_N_EMP). This produces another temporary table (SUM6.DBF containing three fields, ZIP, COUNT and SUM_N_EMP). In SUM6.DBF, check that the frequencies by zip code sum to the original total number of records in DES_FIRM.TXT (checked, but note: some of these will be outside the borders of TAZs (in the 4 county area outside the model area)). Kill the field COUNT and rename the field SUM_N_EMP to CNT_F_T.

1.2.2c) Determine the number of employees that work in DSM area zip codes in the sample file:

TAZMATCH.DBF (this is the sample file of 58,424 "single-location-employer" employee/employer addresses that were geocoded within the model area), with the FZIP field selected, use the ArcView command SUMMARIZE to create a temporary table (SUM7.DBF containing two fields, FZIP and COUNT). Then in the temporary table, create a new field of 5 digit zip codes (FZIP5, character with 5 places). Calculate the 5 digit zip codes (truncate the 9 digit ones, FZIP5 = FZIP.left(5)), kill the FZIP field and with the FZIP5 field selected, use the ArcView command SUMMARIZE. Then, in the SUMMARIZE dialog box, tell ArcView to include a field to sum the values in the COUNT (summary) field (SUM_COUNT). This produces another temporary table of the number of employees in each of the 5 digit zip codes. The table is saved as SUM8.DBF and has FZIP5, COUNT and SUM_COUNT fields. In SUM8.DBF, check that the frequencies by zip code sum to the original total number of records in TAZMATCH.DBF (checked). Kill the field COUNT and rename the field SUM_COUNT to CNT_F_S.

1.2.2d) Join the table resulting from part b) (SUM6.DBF) to the table from part c) (SUM8.DBF). Add a new field (HZF) to this combined table, and calculate it as the ratio of sample counts (CNT_F_S) to all counts (CNT_F_T) for each zip code. Save the resulting join as SUM9.DBF

1.2.2e) Join the result of step d) (SUM9.DBF) which includes HZF to ZIP2TAZ.TXT by zipcode, so that the hit rate, HZF is associated with each TAZ. Save that join as HITTAZ_F.DBF. Export the table as a text file to facilitate import to Tranplan (HITTAZ_F.TXT).

1.2.3) Perform Fratar analysis to factor up each OD pair to account for unique employer and geocode bias

a) Working in TRANPLAN\VARIABLE, import the two text files HITTAZ_R.TXT and HITTAZ_F.TXT into a new spreadsheet, OD-CALV.123. Determine origin and destination growth factors, F_i and F_j as:

i) $F_i = 1/HZR_i$

ii) $F_j = 1/HZF_j$

b) obtain a copy of S.BIN, the sample OD table.

c) Create a control file for the Tranplan Fratar Model Module, (FRATARV.IN). Incorporate TAZ number, $F_{i,}$, and F_j from OD-CALV.123, following format specified on MODELS -- PAGE 4-4 of the Tranplan manual. (this is in the Fratar Model section of the manual).

Control file looks like:

```
$FRATAR MODEL
$FILES
     INPUT FILE = FRATIN, USER ID = $S.BIN$
     OUTPUT FILE = FRATOUT, USER ID = $SV.BIN$
$DATA
FO   1 1    276
FO   2 1    276
FO   3 1    276
FO   4 1    276
.
.
FD   1 1    348
FD   2 1    348
FD   3 1    348
FD   4 1    348
.
.
$END TP FUNCTION
```

d) run the Tranplan Fratar Model, producing an expanded trip table SV.bin (for sample-variable adjusted). Note: Tranplan output reports that factoring up all origins produces 193,272 trip origins in the Fratar output table. Factoring up all destinations would have produced 250,225 trip destinations. The Fratar model hold destinations equal to origins, so the resulting Fratared table has 193,272 "trips".

1.2.4) using the Tranplan module MATRIX TRANSPOSE, create SVT.bin

1.2.5) using the Tranplan module MATRIX MANIPULATE, add SV.bin to SVT.bin to create 2-OD-V.bin (There are now 386,534 "trips")

1.2.6) Perform a second Fratar analysis to factor the synthesized trip table row and column totals to be equal to original HBW row and column totals (Ps and As):

a) Determine origin and destination totals for synthesized trip table 2-OD-V.BIN. Get these totals from the output of Tranplan's REPORT MATRIX module. Add them to Lotus spreadsheet OD-CALV.123.

(note: some manipulation of the output text file is needed to get the data into a manageable format in the spreadsheet. It is helpful to clean up the Tranplan output file with a text editor before importing into the spreadsheet as a text file. It is also useful to know something about the Parse function of the spreadsheet. Alternatively, commas may be added to the Tranplan output to facilitate import to the spreadsheet).

b) Copy the original HBW zone origins and destinations from the OD-CAL.123 spreadsheet (in the TRANPLAN\FIXED% directory). Paste them into OD-CALV.123. Compute growth factors for each zone = values from original HBW table divided by values from 2-OD-V.BIN file.

c) Create a control file for the Tranplan Fratar Model Module, (FRATARV2.IN). Incorporate TAZ number and growth factors from OD-CALV.123, following format specified on MODELS -- PAGE 4-4 of the Tranplan manual. (this is in the Fratar Model section of the manual).

File looks like:

```
$FRATAR MODEL
$FILES
     INPUT FILE = FRATIN, USER ID = $2-OD-V.BIN$
     OUTPUT FILE = FRATOUT, USER ID = $OD-CAL-V.BIN$
$DATA
FO   1 1    075
FO   2 1    052
FO   3 1    102
FO   4 1    140
FO   5 1    082
.
.
.
$END TP FUNCTION
```

e) run the Tranplan Fratar Model on 2-OD-V.BIN, producing a final, variable% synthesized trip table OD-CAL-V.BIN

1.2.7) Verify proper procedure using control totals

If proper procedure was followed, the final matrix OD-CAL-V.BIN should have the same number of total trips as the original HBW trip table (230,391). The output of the FRATAR model reports a matrix with 224,958 trips, with differences due to rounding of growth factors.

1.3) create "improved" OD table via method #3. This method replaces cells in the original trip table with sample data where sample data are higher. It then adjusts the remainder of the trip table downward, constraining row and column totals to equal the Ps and As of the original HBW trip table. Create and work in a directory called TRANPLAN\IMPROVE.

1.3.1) First, create a matrix of trips that represent those cells where the sample trip table, S.BIN, are greater than corresponding HBW trip cells in the original, non-directional (asymmetric) gravity model trip table, GM90.TRP.

To do this, use Tranplan's MATRIX MANIPULATE module to subtract the GM90.TRP cells from S.BIN cells. (result is called S-MINUS.BIN). Then, using MATRIX UPDATE, replace negative cells with zero and positive cells with one. (result is called S-GM-01.BIN). Using MATRIX MANIPULATE, multiply the zero-one matrix by the sample matrix to obtain S-GT.BIN (S-GT stands for sample, greater). This matrix has zeros in cells where the sample was less than or equal to the original table. Use the following control file (MANIP.IN):

```
$MATRIX MANIPULATE
$FILES
```

```
  INPUT FILE = TMAN1, USERID = $S.BIN$
  INPUT FILE = TMAN2, USERID = $GM90.TRP$
  OUTPUT FILE = TMAN3, USERID = $S-MINUS.BIN$
$DATA
  TMAN3,T1 = TMAN1,T1 - TMAN2,T1
$END TP FUNCTION

$MATRIX UPDATE
$FILES
  INPUT FILE = UPDIN, USERID = $S-MINUS.BIN$
  OUTPUT FILE = UPDOUT, USERID = $S-GM-01.BIN$
$DATA
  T1,1-643,1-643,R 1,EQ,0
  T1,1-643,1-643,R 1,GT,0
  T1,1-643,1-643,R 0,LT,0
$END TP FUNCTION

$MATRIX MANIPULATE
$FILES
  INPUT FILE = TMAN1, USERID = $S.BIN$
  INPUT FILE = TMAN2, USERID = $GM-S-01.BIN$
  OUTPUT FILE = TMAN3, USERID = $S-GT.BIN$
$DATA
  TMAN3,T1 = TMAN1,T1 * TMAN2,T1
$END TP FUNCTION
```

1.3.2) Next, create a matrix of trips that represent those cells where the HBW trip cells in the gravity model trip table, GM90.TRP are greater than or equal to corresponding cells in the sample trip table, S.BIN.

To do this, use Tranplan's MATRIX MANIPULATE module to subtract the S.BIN cells from GM90.TRP cells. (result is called GM-MINUS.BIN). Then, using MATRIX UPDATE, replace negative cells with zero and positive cells with one. (result is called GM-S-01.BIN). Using MATRIX MANIPULATE, multiply the zero-one matrix by the HBW matrix to obtain GM-GE.BIN (GM-GE stands for GM90, greater than or equal to). This matrix has zeros in cells where the sample was greater than the original table. Use the following control file (MANIP2.IN):

```
$MATRIX MANIPULATE
$FILES
  INPUT FILE = TMAN1, USERID = $S.BIN$
  INPUT FILE = TMAN2, USERID = $GM90.TRP$
  OUTPUT FILE = TMAN3, USERID = $GM-MINUS.BIN$
$DATA
  TMAN3,T1 = TMAN2,T1 - TMAN1,T1
$END TP FUNCTION

$MATRIX UPDATE
$FILES
  INPUT FILE = UPDIN, USERID = $GM-MINUS.BIN$
  OUTPUT FILE = UPDOUT, USERID = $GM-S-01.BIN$
$DATA
  T1,1-643,1-643,R 1,EQ,0
  T1,1-643,1-643,R 1,GT,0
  T1,1-643,1-643,R 0,LT,0
```

$END TP FUNCTION

$MATRIX MANIPULATE
$FILES
  INPUT FILE = TMAN1, USERID = $GM90.TRP$
  INPUT FILE = TMAN2, USERID = $GM-S-01.BIN$
  OUTPUT FILE = TMAN3, USERID = $GM-GE.BIN$
$DATA
  TMAN3,T1 = TMAN1,T1 * TMAN2,T1
$END TP FUNCTION

1.3.3) Using Tranplan's REPORT MATRIX, determine the total number of trips in each of the new matrices and compute the percentage of the original gravity model HBW table represented by each. (Recall that the total number of HBW trips in the original HBW trip table is 230,391.)

The total number of trips in the S-GT matrix is 38,154, representing 16.6% of the gravity model HBW trip table.

The total number of trips in the GM-GE matrix is 223,878 representing 97.2% of the gravity model HBW trip table

1.3.4) Factor down the GM-GE trip table, so that when added to the S-GT matrix, the total will be the same as the original number of trips in the HBW gravity model trip table. The factor is equal to [230,391-SUM(S-GT)]/SUM(GM-GE) = 0.85867. Do this with MATRIX UPDATE to create GM-GE-I.BIN

1.3.5) Use MATRIX MANIPULATE to add S-GT.BIN to GM-GE-I.BIN to comprise a new, improved directional trip table, OD-I.BIN. Note that the original HBW trip table represents all work trips starting at the home and ending at work. This represents twice the actual number of a.m. trips and omits trips typically made during the p.m. The following procedures adjust the trip table to reflect a.m and p.m. directionality.

1.3.6) using the Tranplan module MATRIX TRANSPOSE, create OD-I-T.BIN

1.3.7) using the Tranplan module MATRIX MANIPULATE, add OD-I.BIN to OD-I-T.BIN to create 2-OD-I.BIN (There are now two times the actual number of trips).

1.3.8) Using the Tranplan module MATRIX UPDATE, multiply 2-OD-I.BIN by 0.5 to obtain OD-I.BIN.

1.3.9) Perform Fratar analysis factor up or down improved trip table rows and columns to be equal to original row and column totals (Ps and As)

        a) Create a spreadsheet in TRANPLAN\IMPROVE called IMPROVE.123. Using REPORT MATRIX, determine zone origin and destination totals for improved trip table OD-I2.BIN. Copy them into the IMPROVE.123 spreadsheet.

                (note: some manipulation of the output text file is needed to get the data into a manageable format in the spreadsheet. It is helpful to clean up the Tranplan output file with a text editor before importing into the spreadsheet as a text file. It is also useful to know something about the Parse function of the spreadsheet. Alternatively, commas may be added to the Tranplan output to facilitate import to the spreadsheet).

b) Copy the original HBW zone origins and destinations from the OD-CAL.123 spreadsheet (in the TRANPLAN\FIXED% directory). Paste them into IMPROVE.123. Compute growth factors for each zone = values from original HBW table divided by values from OD-I.BIN file.

c) Create a control file for the Tranplan Fratar Model Module, (FRATARI.IN). Incorporate TAZ number and growth factors from OD-I.123, following format specified on MODELS -- PAGE 4-4 of the Tranplan manual. (this is in the Fratar Model section of the manual).

Control file looks like:
$FRATAR MODEL
$FILES
     INPUT FILE = FRATIN, USER ID = $OD-IX.BIN$
     OUTPUT FILE = FRATOUT, USER ID = $OD-II.BIN$
$DATA
FO   1 1   095
FO   2 1   093
FO   3 1   099
FO   4 1   104
.
.
$END TP FUNCTION

d) run the Tranplan Fratar Model on OD-IX, producing a final, improved trip table OD-CAL-I.bin

1.3.7) verify proper procedure using control totals

If proper procedure was followed, the final matrix OD-I3.BIN should have the same number of total trips as the original HBW trip table (230,391). To verify, we ran the Tranplan module REPORT MATRIX.

The results of this process indicate a total number of productions of 230,617, close enough.

2) Run Model with original and synthesized (or "improved") trip tables, export network with loaded volumes to MapInfo.

2.1) Run DSM Model with original HBW table, export network with loaded volumes to MapInfo

2.1.1) Working in directory TRANPLAN\BASE1990, obtain necessary files and run Tranplan to obtain original 1990base binary loaded network file (NETXLOD.BIN) -- see flowchart for files names and processes

2.1.2) Create and work in a directory called MAPINFO\BASE1990. Copy required files from TRANPLAN\BASE1990 directory:

NETXLOD.BIN (binary loaded network)
GM1990.PA (productions and attractions file)

Obtain and copy the MapBasic executable MODEL.MBX and Fortran program TP_MI.EXE into the same directory (both of these programs were created and are available from CTRE).

2.1.3) Run the Tranplan utility NETCARD to export the binary trip table into a text format. Options for NETCARD are as follows:

> Enter input file name>netxlod.bin
> Enter output file name>netxlod.txt
> Do you wish speeds to be output (rather than time)
> (Note -- speeds will not be rounded)   (Y/N)?y
> Speed Factor> 1
> Loaded volumes will be in the CAPACITY 2 field
> Do you wish CAPACITY 2 to output in the CAPACITY 1 field (Y/N)?y
> Loaded network appears to have been created with BPR iterations
> Do you wish to average the iterations (Y/N)?y
> Enter factor to multiply the loaded volumes, e.g. 1.0>1
> If you wish a specific iteration time (or speed) to be in the TIME2 field
> enter iteration (or zero for no replacement -- 99 for last iteration) >99
> Only one-way format on output (Y/N)?n
> Do you wish header and option records on output (Y/N)?n

2.1.4) to facilitate import into MapInfo, using a text editor such as Wordpad or Textpad, break the NETXLOD.TXT file into two parts, one for node information (X-NODES.TXT) and one for link information (X-LINKS.TXT).

2.1.5) to import loaded network into MapInfo, Run the MapInfo mapbasic program MODEL.MBX. This will provide the user a menu option called TP_MI. From this menu, choose REGISTERING, which will provide the user a DOS prompt. At the DOS prompt, run the program TP_MI.EXE and answer the following:

> Please input the name of the file containing the node data : x-nodes.txt
>
> Please input the name of the file containing the productions and attractions data : gm1990.pa
>
> Please input the name of the file containing the link data : x-links.txt
>
> Please enter the number of centroids in the network: 643
>
> If TRANPLAN coordinates are in 1/100th miles or any fraction of miles, Y would be appropriate
> Would you like to multiply the x and y coordinates by a factor? (Y/N)
> y
>
> please input the factor for the x coordinate
> 52.8
>
> please input the factor for the y coordinate
> 52.8
>
> Enter the appropriate description for the node file.
>   Enter 1 for small coordinate format.
>   Enter 2 for large coordinate format.
> 2

Type EXIT when the program finishes (may take a few minutes)

Now the MapInfo program creates the MapInfo tables. You must choose a coordinate system. Follow the instructions and specify State Plane Coordinate System for Iowa South, NAD27.

The program will now create the MapInfo Tables.

You will need to use the MapInfo function Update Column to add capacity2 and capacity4 fields and store the result in the total_loaded_volume field (to get two way volumes).

Delete all files in the MAPINFO\BASE1990 directory except LINKS.* files. Also keep NETXLOD.BIN, GM1990.PA, MODEL.MBX and TP_MI.EXE (in case recreating the MapInfo LINKS table is required).

2.2) Run DSM Model with Fixed ratio synthesized trip table (method 1). Export network with loaded volumes to MapInfo.

2.2.1) Working in directory TRANPLAN\FIXED%, obtain copies of necessary files from original Tranplan model (from TRANPLAN\BASE1990 directory, see flowchart 2):

RUNX-F.IN trip distribution control file
LOADX-F.IN traffic assignment control file

NET90CMS.BIN binary network
GM90.TRP binary internal trip tables
EXTEXT90.TRP binary external trip table

2.2.2) substitute the trip table produced from the sample (OD-CAL-F.BIN) for the HBW trip table and run the Tranplan model.

To do this, use the Tranplan module MATRIX MANIPULATE , with the following control file, MANIP2.IN, which also reports an error check (after substituting the new HBW table, the total number of trips should remain the same:

```
$MATRIX MANIPULATE
$FILES
   INPUT FILE = TMAN1, USERID = $OD-CAL-F.BIN$
   INPUT FILE = TMAN2, USERID = $GM90.TRP$
   OUTPUT FILE = TMAN3, USERID = $GM90-F.TRP$
$DATA
TMAN3,T1 = TMAN1,T1
TMAN3,T2 = TMAN2,T2
TMAN3,T3 = TMAN2,T3
TMAN3,T4 = TMAN2,T4
TMAN3,T5 = TMAN2,T5
$END TP FUNCTION
$REPORT MATRIX
$FILES
   INPUT FILE = RTABIN, USERID = $GM90.TRP$
$OPTIONS
   PRINT TRIP ENDS
$END TP FUNCTION
$REPORT MATRIX
$FILES
   INPUT FILE = RTABIN, USERID = $GM90-F.TRP$
```

$OPTIONS
  PRINT TRIP ENDS
$END TP FUNCTION

Checking the output file, GM90.TRP has 230,391 HBW trips. GM90-F.TRP has 225,194. Close enough.

2.2.3) Run Tranplan to obtain a new binary loaded network file (NETXLODF.BIN) using the trip table modified by a fixed percent (gm90-f.trp). Must modify RUNX.IN to begin with MATRIX MANILPULATE and modify both RUNX.IN and LOADX.IN to use appropriate filenames as shown in flowchart 2. (new control files are called RUNX-F.IN and LOADX-F.IN)

2.2.4) Create and work in a directory called MAPINFO\FIXED%. Copy required files from TRANPLAN\FIXED% directory:

NETXLODF.BIN (binary loaded network)
GM1990.PA (productions and attractions file)

Obtain and copy the MapBasic executable MODEL.MBX and Fortran program TP_MI.EXE into the same directory (both of these programs were created and are available from CTRE).

2.2.5) Run the Tranplan utility NETCARD to export the binary trip table into a text format. Options for NETCARD are as follows:

> Enter input file name>netxlodf.bin
> Enter output file name>netxlodf.txt
> Do you wish speeds to be output (rather than time)
> (Note -- speeds will not be rounded)   (Y/N)?y
> Speed Factor> 1
> Loaded volumes will be in the CAPACITY 2 field
> Do you wish CAPACITY 2 to output in the CAPACITY 1 field (Y/N)?y
> Loaded network appears to have been created with BPR iterations
> Do you wish to average the iterations (Y/N)?y
> Enter factor to multiply the loaded volumes, e.g. 1.0>1
> If you wish a specific iteration time (or speed) to be in the TIME2 field
> enter iteration (or zero for no replacement -- 99 for last iteration) >99
> Only one-way format on output (Y/N)?n
> Do you wish header and option records on output (Y/N)?n

2.2.6) To facilitate import into MapInfo, using a text editor such as Wordpad or Textpad, break the NETXLODF.TXT file into two parts, one for node information (X-NODES.TXT) and one for link information (X-LINKS.TXT).

2.2.7) To import loaded network into MapInfo, Run the MapInfo MapBasic program MODEL.MBX. This will provide the user a menu option called TP_MI. From this menu, choose REGISTERING, which will provide the user a DOS prompt. At the DOS prompt, run the program TP_MI.EXE and answer the following:

> Please input the name of the file containing the node data : x-nodes.txt
>
> Please input the name of the file containing the productions and attractions data : gm1990.pa
>
> Please input the name of the file containing the link data : x-links.txt

Please enter the number of centroids in the network: 643

If TRANPLAN coordinates are in 1/100th miles or any fraction of miles, Y would be appropriate
Would you like to multiply the x and y coordinates by a factor? (Y/N)
y

please input the factor for the x coordinate
52.8

please input the factor for the y coordinate
52.8

Enter the appropriate description for the node file.
    Enter 1 for small coordinate format.
    Enter 2 for large coordinate format.
2

Type EXIT when the program finishes (may take a few minutes)

Now the MapInfo program creates the MapInfo tables. You must choose a coordinate system. Follow the instructions and specify State Plane Coordinate System for Iowa South, NAD27.

The program will now create the MapInfo Tables.

You will need to use the MapInfo function Update Column to add capacity2 and capacity4 fields and store the result in the total_loaded_volume field (to get two way volumes).

Delete all files in the MAPINFO\FIXED% directory except LINKS.* files. Also keep NETXLODF.BIN, GM1990.PA, MODEL.MBX and TP_MI.EXE (in case recreating the MapInfo LINKS table is required).

2.3) Run DSM Model with variable ratio synthesized trip table (method 2). Export network with loaded volumes to MapInfo.

2.3.1) Working in directory TRANPLAN\VARIABLE, obtain copies of necessary files from original Tranplan model (from TRANPLAN\BASE1990 directory, see flowchart 3):

RUNX-V.IN trip distribution control file
LOADX-V.IN traffic assignment control file

NET90CMS.BIN binary network
GM90.TRP binary internal trip tables
EXTEXT90.TRP binary external trip table

2.3.2) substitute the trip table produced from the sample (OD-CAL-V.BIN) for the HBW trip table and run the Tranplan model.

To do this, use the Tranplan module MATRIX MANIPULATE, with the following control file, MANIP2.IN, which also reports an error check (after substituting the new HBW table, the total number of trips should remain the same:

```
$MATRIX MANIPULATE
$FILES
  INPUT FILE = TMAN1, USERID = $OD-CAL-V.BIN$
  INPUT FILE = TMAN2, USERID = $GM90.TRP$
  OUTPUT FILE = TMAN3, USERID = $GM90-V.TRP$
$DATA
TMAN3,T1 = TMAN1,T1
TMAN3,T2 = TMAN2,T2
TMAN3,T3 = TMAN2,T3
TMAN3,T4 = TMAN2,T4
TMAN3,T5 = TMAN2,T5
$END TP FUNCTION
$REPORT MATRIX
$FILES
  INPUT FILE = RTABIN, USERID = $GM90.TRP$
$OPTIONS
  PRINT TRIP ENDS
$END TP FUNCTION
$REPORT MATRIX
$FILES
  INPUT FILE = RTABIN, USERID = $GM90-V.TRP$
$OPTIONS
  PRINT TRIP ENDS
$END TP FUNCTION
```

Checking the output file, GM90.TRP has 230,391 HBW trips. GM90-V.TRP has 224,924. Close enough.

2.3.3) Run Tranplan to obtain a new binary loaded network file (NETXLODV.BIN) using the trip table modified by a fixed percent (gm90-V.trp). Must modify RUNX.IN to begin with MATRIX MANILPULATE and modify both RUNX.IN and LOADX.IN to use appropriate filenames as shown in flowchart 3. (new control files are called RUNX-V.IN and LOADX-V.IN)

2.3.4) Create and work in a directory called MAPINFO\VARIABLE. Copy required files from TRANPLAN\VARIABLE directory:

NETXLODV.BIN (binary loaded network)
GM1990.PA (productions and attractions file)

Obtain and copy the MapBasic executable MODEL.MBX and Fortran program TP_MI.EXE into the same directory (both of these programs were created and are available from CTRE).

2.3.5) Run the Tranplan utility NETCARD to export the binary trip table into a text format. Options for NETCARD are as follows:

> Enter input file name>netxlodv.bin
> Enter output file name>netxlodv.txt
> Do you wish speeds to be output (rather than time)
> (Note -- speeds will not be rounded)   (Y/N)?y
> Speed Factor> 1
> Loaded volumes will be in the CAPACITY 2 field
> Do you wish CAPACITY 2 to output in the CAPACITY 1 field (Y/N)?y
> Loaded network appears to have been created with BPR iterations
> Do you wish to average the iterations (Y/N)?y

Enter factor to multiply the loaded volumes, e.g. 1.0>1
If you wish a specific iteration time (or speed) to be in the TIME2 field
enter iteration (or zero for no replacement -- 99 for last iteration) >99
Only one-way format on output (Y/N)?n
Do you wish header and option records on output (Y/N)?n

2.3.6) To facilitate import into MapInfo, using a text editor such as Wordpad or
Textpad, break the NETXLODV.TXT file into two parts, one for node information (X-
NODES.TXT) and one for link information (X-LINKS.TXT).

2.3.7) To import loaded network into MapInfo, Run the MapInfo MapBasic program
MODEL.MBX. This will provide the user a menu option called TP_MI. From this
menu, choose REGISTERING, which will provide the user a DOS prompt. At the DOS
prompt, run the program TP_MI.EXE and answer the following:

Please input the name of the file containing the node data : x-nodes.txt

Please input the name of the file containing the productions and attractions data
: gm1990.pa

Please input the name of the file containing the link data : x-links.txt

Please enter the number of centroids in the network: 643

If TRANPLAN coordinates are in 1/100th miles or any fraction of miles, Y
would be appropriate
Would you like to multiply the x and y coordinates by a factor? (Y/N)
y

please input the factor for the x coordinate
52.8

please input the factor for the y coordinate
52.8

Enter the appropriate description for the node file.
    Enter 1 for small coordinate format.
    Enter 2 for large coordinate format.
2

Type EXIT when the program finishes (may take a few minutes)

Now the MapInfo program creates the MapInfo tables. You must choose a coordinate
system. Follow the instructions and specify State Plane Coordinate System for Iowa
South, NAD27.

The program will now create the MapInfo Tables.

You will need to use the MapInfo function Update Column to add capacity2 and
capacity4 fields and store the result in the total_loaded_volume field (to get two way
volumes).

Delete all files in the MAPINFO\VARIABLE directory except LINKS.* files. Also
keep NETXLODV.BIN, GM1990.PA, MODEL.MBX and TP_MI.EXE (in case
recreating the MapInfo LINKS table is required).

2.4) Run DSM Model with "improved" trip table (method 3). Export network with loaded volumes to MapInfo.

2.4.1) Working in directory TRANPLAN\IMPROVE, obtain copies of necessary files from original Tranplan model (from TRANPLAN\BASE1990 directory, see flowchart 4):

RUNX-I.IN trip distribution control file
LOADX-I.IN traffic assignment control file

NET90CMS.BIN binary network
GM90.TRP binary internal trip tables
EXTEXT90.TRP binary external trip table

2.4.2) substitute the trip table produced from the sample (OD-CAL-I.BIN) for the HBW trip table and run the Tranplan model.

To do this, use the Tranplan module MATRIX MANIPULATE , with the following control file, MANIP5.IN, which also reports an error check (after substituting the new HBW table, the total number of trips should remain the same:

```
$MATRIX MANIPULATE
$FILES
   INPUT FILE = TMAN1, USERID = $OD-CAL-I.BIN$
   INPUT FILE = TMAN2, USERID = $GM90.TRP$
   OUTPUT FILE = TMAN3, USERID = $GM90-I.TRP$
$DATA
TMAN3,T1 = TMAN1,T1
TMAN3,T2 = TMAN2,T2
TMAN3,T3 = TMAN2,T3
TMAN3,T4 = TMAN2,T4
TMAN3,T5 = TMAN2,T5
$END TP FUNCTION
$REPORT MATRIX
$FILES
   INPUT FILE = RTABIN, USERID = $GM90.TRP$
$OPTIONS
   PRINT TRIP ENDS
$END TP FUNCTION
$REPORT MATRIX
$FILES
   INPUT FILE = RTABIN, USERID = $GM90-I.TRP$
$OPTIONS
   PRINT TRIP ENDS
$END TP FUNCTION
```

Checking the output file, GM90.TRP has 230,391 HBW trips. GM90-I.TRP has 230,623. Close enough.

2.4.3) Run Tranplan to obtain a new binary loaded network file (NETXLODI.BIN) using the trip table modified by a fixed percent (gm90-I.trp). Must modify RUNX.IN to begin with MATRIX MANILPULATE and modify both RUNX.IN and LOADX.IN to use appropriate filenames as shown in flowchart 4. (new control files are called RUNX-I.IN and LOADX-I.IN)

2.4.4) Create and work in a directory called MapInfo\IMPROV. Copy required files from TRANPLAN\IMPROV directory:

NETXLODI.BIN (binary loaded network)
GM1990.PA (productions and attractions file)

Obtain and copy the MapBasic executable MODEL.MBX and Fortran program TP_MI.EXE into the same directory (both of these programs were created and are available from CTRE).

2.4.5) Run the Tranplan utility NETCARD to export the binary trip table into a text format. Options for NETCARD are as follows:

        Enter input file name>netxlodi.bin
        Enter output file name>netxlodi.txt
        Do you wish speeds to be output (rather than time)
        (Note -- speeds will not be rounded)   (Y/N)?y
        Speed Factor> 1
        Loaded volumes will be in the CAPACITY 2 field
        Do you wish CAPACITY 2 to output in the CAPACITY 1 field (Y/N)?y
        Loaded network appears to have been created with BPR iterations
        Do you wish to average the iterations (Y/N)?y
        Enter factor to multiply the loaded volumes, e.g. 1.0>1
        If you wish a specific iteration time (or speed) to be in the TIME2 field
        enter iteration (or zero for no replacement -- 99 for last iteration) >99
        Only one-way format on output (Y/N)?n
        Do you wish header and option records on output (Y/N)?n

2.4.6) To facilitate import into MapInfo, using a text editor such as Wordpad or Textpad, break the NETXLODI.TXT file into two parts, one for node information (X-NODES.TXT) and one for link information (X-LINKS.TXT).

2.4.7) To import loaded network into MapInfo, Run the MapInfo MapBasic program MODEL.MBX. This will provide the user a menu option called TP_MI. From this menu, choose REGISTERING, which will provide the user a DOS prompt. At the DOS prompt, run the program TP_MI.EXE and answer the following:

        Please input the name of the file containing the node data : x-nodes.txt

        Please input the name of the file containing the productions and attractions data
        : gm1990.pa

        Please input the name of the file containing the link data : x-links.txt

        Please enter the number of centroids in the network: 643

        If TRANPLAN coordinates are in 1/100th miles or any fraction of miles, Y
        would be appropriate
        Would you like to multiply the x and y coordinates by a factor? (Y/N)
        y

        please input the factor for the x coordinate
        52.8

        please input the factor for the y coordinate
        52.8

Enter the appropriate description for the node file.
Enter 1 for small coordinate format.
Enter 2 for large coordinate format.
2

Type EXIT when the program finishes (may take a few minutes)

Now the MapInfo program creates the MapInfo tables. You must choose a coordinate system. Follow the instructions and specify State Plane Coordinate System for Iowa South, NAD27.

The program will now create the MapInfo Tables.

You will need to use the MapInfo function Update Column to add capacity2 and capacity4 fields and store the result in the total_loaded_volume field (to get two way volumes).

Delete all files in the MAPINFO\IMPROVE directory except LINKS.* files. Also keep NETXLODF.BIN, GM1990.PA, MODEL.MBX and TP_MI.EXE (in case recreating the MapInfo LINKS table is required).

3). Display, mapping and statistical assessment of results.

3.1) Compare loaded network volumes using synthesized OD table with volumes using original HBW table and with ground counts. Which model has a higher R-squared?

3.1.1) Create a single MapInfo table that incorporates all travel model outputs (COMPARE.TAB):

1 - Ground Counts
2 - Original, 1990 Calibrated MPO model, based on Gravity model HBW trips
3 - Model based on fixed% synthesized trip table
4 - Model based on variable% synthesized trip table
5 - Model based on "improved" trip table

Create a directory called MAPINFO\COMPARE. Copy the MAPINFO\BASE1990 links table to that directory.

In the links table add five new columns (using the MapInfo TABLE MAINTENANCE option). Create new integer fields for ground counts (GROUND), original DSM MPO model (BASE1990), fixed% synthesized trip table (FIXED), variable% synthesized trip table (VARIABLE), and "improved" trip table (IMPROVE).

To add ground counts, start with file provided by DSM MPO with ground counts for selected links (COUN1990.XLS, found in the TRANPLAN BASE1990 directory). In the spreadsheet, create a column for link ID which will be the lower numbered node followed by a space and the higher numbered node for each link. Create a column for the lower numbered node, and in each cell, put a conditional formula to identify the lower numbered node (if statement). Do the same for another column for the higher numbered node (put this column just after the lower numbered node conditional formula column). Copy and past the two new columns to a text processor such as textpad that is capable of block editing. In TextPad, replace the tab seperating each node number with a space. Copy and paste the new "ID" field back in to the Excel Spreadsheet. Create a new Excel spreadsheet (GROUND.XLS, also in TRANPLAN BASE1990 directory) that contains only the ID and ground count (GROUND) fields.

In the MapInfo table "links" in the MAPINFO\COMPARE directory, use COLUMN UPDATE to update four of the new columns with appropriate columns from the links tables of BASE1990, FIXED%, VARIABLE, and IMPROVE. Each of the four model links tables must be open along with the COMPARE\links table in order to process the column updates.

In MapInfo, to update the GROUND column, open the Excel file (GROUND.XLS) and join links to ground where ID = ID (Recall that ID is concatenated from anode and bnode for each link).

For ease of use, remove all columns except the five new ones and the following: A, B, length, capacity1 (capacity of the link), and ID.

To compute correlation coefficient, export the data (using file-save as) in dbase format to a table called R2.TAB. Import that table in EXCEL. Use TOOLS DATA ANALYSIS to compute correlation coefficients for each model. Results indicate an R^2 of 0.901 for the original, base1990 model.

Both of the synthesized trip tables produce slightly better R^2 (0.905 for the fixed% adjusted and 0.911 for the variable% adjusted.) The "improved" method did not change the results $R^2 = 0.901$.

Because the variable rate model produced the greatest improvement, its resulting validation was graphically compared to that of the base1990 model (see Figures C1 and C2). Note on each figure the curve representing the maximum (TRB) recommended error in a model forecast (depends on variability in ground counts, hence, is monotonically decreasing). An additional graphic points out the improvement of the variable model over the base1990 model, particularly for higher capacity links (those of greatest interest to regional modelers).
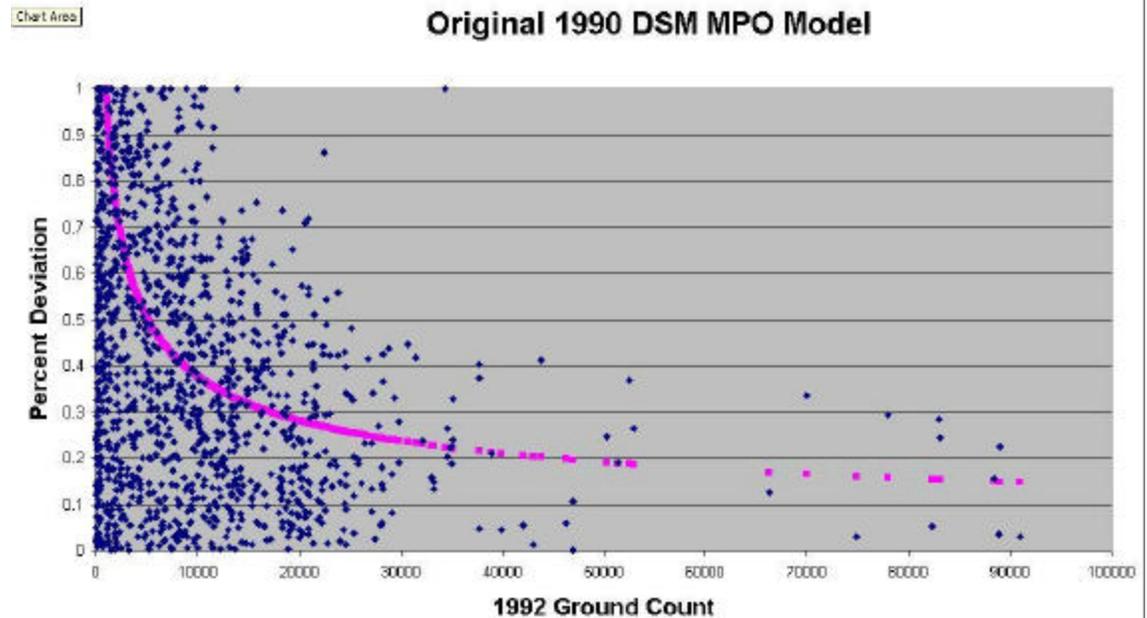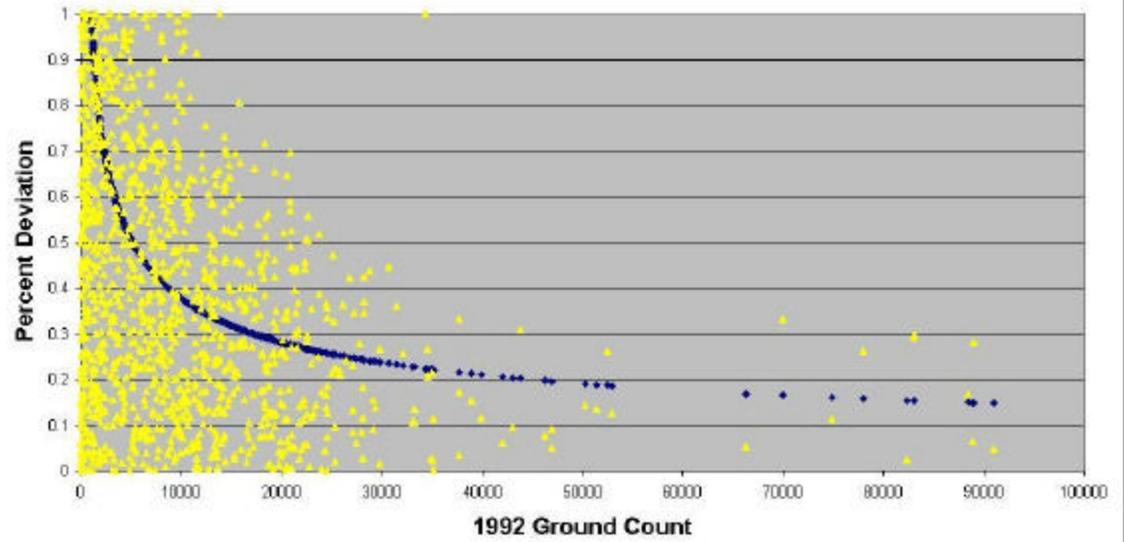

Figure C1: Evaluation of Original 1990 Des Moines MPO Model

67

Variable Synthesized Model Results

Figure C2: Des Moines Model Using Variable Synthesized HBW Trip Table

While the original, base1990 model results in 564 links beyond the TRB recommended
tolerance, the variable model reduces that number to 539. The improvement is
particularly noteworthy for higher capacity links, where the variable model improves
conformance with tolerance by 58% for links over 20,000 ADT, and 63% for links over
30,000 ADT. (These figures are only for those links with ground counts, which was the
majority of links.)